On the Dynamics of HIV and Malaria Infection - Insights From Mathematical Models

by

Bernhard Paul Konrad

Dipl.-Math. Karlsruhe Institute of Technology, 2010

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Mathematics)

The University of British Columbia
(Vancouver)

September 2015

Abstract

We develop and apply mathematical models to obtain insights into the dynamics of human immunodeficiency virus (HIV) and malaria infection. We consider three case studies.

1. The duration of the time between exposure and detectability of HIV infection is difficult to estimate because precise dates of exposure are rarely known. Therefore, the reliability of clinical HIV testing during the first few weeks of infections is unknown, creating anxiety among HIV-exposed individuals and their physicians. We address this knowledge gap by fitting stochastic models of early HIV infection to detailed viral load time-courses, taken shortly after exposure, from 78 plasma donors. Since every plasma donor in our data eventually becomes infected, we condition our model to reflect this bias before fitting to the data. Our model prediction for the mean eclipse period is 8-10 days. We further quantify the reliability of a negative test t days after potential exposure to inform physicians about the value of initial and follow-up testing.

2. The recently launched Get Checked Online (GCO) program aims at increasing the HIV testing rate in the Vancouver men who have sex with men population by facilitating test taking and result delivery. We develop mathematical models and extract parameter values from surveys and interviews to quantify GCO's population-level impact. Our models predict that the epidemic is growing overall, that its severeness is increased by the presence of a high-risk group and that, even at modest effectiveness, GCO might avert 34-66 new infections in the next five years.

3. Metarhizium anisopliae is a naturally occurring fungal pathogen of mosquitoes that has been engineered to act against malaria by effectively blocking onward transmission from the mosquito vector. We develop and analyse two mathematical

models to examine the efficacy of this fungal pathogen. We find that, in many plausible scenarios, the best effects are achieved with a reduced or minimal pathogen virulence, even if the likelihood of resistance to the fungus is negligible. The results depend on the interplay between two main effects: the ability of the fungus to reduce the mosquito population, and the ability of fungus-infected mosquitoes to compete for resources with non-fungus-infected mosquitoes.

Preface

I was the lead investigator for the project in Chapter 3 and corresponding Appendix B, where I was responsible for establishing the study, analyzing the data, developing and analyzing the models and manuscript composition. D. Taylor and G. Ogilvie assembled the data set. J. Conway helped with mathematical analysis. D. Coombs was the supervisory author on this project and helped with manuscript composition.

I was the lead investigator for the project in Chapter 4, where I was responsible for analyzing the data, developing and analyzing the models and manuscript composition. The study was established by M. Gilbert and D. Coombs. M. Gilbert organized the collection of survey data for use in this study. W. Michelow helped analyze the data. M. Gilbert, W. Michelow and D. Coombs helped develop the models.

A version of Chapter 5 and corresponding Appendix C have been published in the Malaria Journal [66]. I was the lead investigator, responsible for all major areas of establishing the study, developing and analyzing the models and manuscript composition. M. Lindstrom, J. Zhu and A. Gumpinger assisted in developing and analyzing the models. D. Coombs was the supervisory author on this project and helped with manuscript composition.

Table of Contents

List of Tables

List of Figures

Glossary

ART antiretroviral therapy

BCCDC British Columbia Centre for Disease Control

MSM men who have sex with men, a high-risk group for HIV infection.

IDU injecting drug user, a high-risk group for HIV infection.

GCO Get Checked Online

STI sexually transmitted infection

HIV human immunodeficiency virus

SIV simian immunodeficiency viruses, closely related to HIV the natural host of SIV are non-human primates. Unlike HIV, SIV infections in their natural hosts are usually non-pathogenic.

ODE ordinary differential equation

PDE partial differential equation

HIM health Initiative for Men

IRS indoor residual spraying

DDT dichlorodiphenyltrichlo-roethane

PDF Portable Document Format

BC British Columbia

AIDS acquired immunodeficiency syndrome

CD4 cluster of differentiation 4, a glycoprotein found on the surface of immune cells that HIV requires to infect a cell

Acknowledgments

This thesis would not have been possible without the support of many great people throughout the years.

First and foremost I want to thank my family for supporting my wish to seek the best education and work with the most exciting group, even when that meant that I would be on the other side of the planet for several years.

Thank you Isabell Graf for being on my side and supporting me, especially during the difficult and most busy times. Without your help and inspiration this journey would not have been possible.

I want to thank my supervisor Daniel Coombs for giving me the opportunity to work in his research team and for his support, encouragement, and guidance throughout. I remember countless inspiring conversations full of ideas and energy, always with a healthy portion of humor. It makes me proud to have such a smart and driven *Doktorvater*, and I hope to can repay some of your trust with the work in this thesis.

Thank you to my supervisory committee Leah Keshet, Rachel Kuske, Brian Wetton and Ruy Ribeiro for your valuable input and advice.

My experience at UBC would not have been nearly as pleasant without many friends that I was fortunate to meet and work with for several years. I'm excited to follow your next adventures: Carmen Bruni, Christina Koch, Michael Lindstrom, Iain Moyles, William Thompson and, of course, Alejandra Donají Herrera Reyes!

I am grateful for financial support from UBC, the International Graduate Training Centre by the Pacific Institute for the Mathematical Sciences, and the Canadian Institute of Health Research.

Chapter 1

Introduction

1.1 Mathematical modeling of infectious diseases

The main goal of mathematical biology is to obtain insights into biological systems that lead to deeper understanding and useful predictions. To do so requires describing the biological system in a mathematical manner, using well-defined assumptions and simplifications, analyzing the resulting mathematical problem, and interpreting its results in the biological context. This approach is most powerful in systems that are difficult to study experimentally, in particular when facing technical limitations or ethical considerations. In such situations mathematical models can be used to surface otherwise hidden dynamics and predict the impact of various alterations on the study system or variables of interest. Therefore it is natural and useful to apply mathematical models to study infectious diseases, which can broadly be classified into two categories.

Within-host models study the dynamics within a single host or individual, such as the interaction of pathogens with the immune system or the signaling pathways responsible for the function of immune cells. Typically the system is difficult to measure directly, but some general interaction principles that govern the dynamics are known or proposed. By translating these general principles into well-defined rules, mathematical models are able to reveal hidden dynamics, increase the understanding of the underlying biological system, and ultimately provide suggestions for new ways to fight the infection. One highlight of within-host infectious disease

modeling is a story from the early days of studying the human immunodeficiency virus (HIV). Two teams of experimentalists and theoreticians had formed: George Shaw with Martin Novak and David Ho with Alan Perelson. These teams published two historic papers [52, 111] in 1995 that finally refuted a false understanding of HIV dynamics and ultimately led to new era in fighting the infection. At that time it was known that the viral load and number of T helper cells (and more generally cluster of differentiation 4 (CD4) cells, the main target of HIV) of an HIV-infected individual was roughly constant for many years until eventual disease progression, which was believed to be indicative of slow and static within-host dynamics. In the studies patients who received the recently developed first generation of antiretroviral therapy (ART) were analyzed. The time-series of the viral load of the patients was measured and showed an exponential decline with rate of about $0.6\ \text{day}^{-1}$, but it was unclear how this result should be interpreted and what it implied. So a simple two-dimensional system of ordinary differential equations (ODES) was set-up to describe the hidden dynamics under ART:

$$\begin{aligned} \frac{dT^*}{dt} &= -\delta T^*, \\ \frac{dV}{dt} &= pT^* - cV, \end{aligned} \tag{1.1}$$

where infected target cells T^* die at rate δ and produce virions at rate p, while virions are cleared at rate c. Being a linear system the solution of the change of viral load is quickly found to be

$$V(t) = V_0 \frac{ce^{-\delta t} - \delta e^{-ct}}{c-\delta} \approx V_0 e^{-\delta t}. \tag{1.2}$$

The approximation is valid because the viral clearance rate was known to be much faster than the death rate of infected cells, $c \gg \delta$. This shows that the observed exponential decay rate corresponds to the death rate of infected target cells, which are thus found to have an average half-life of $\ln(2)/0.6 \approx 1.2$ days only! This game-changing realization, that millions of an infected person's CD4 cells are killed and replaced every day, finally refuted the theory that disease progression is slow because "not much is happening". In particular, those results imply that HIV is very

susceptible to interference in the viral life cycle and thus inspired further search for new ART agents. When several anti-HIV drugs were combined to avoid the evolution of resistance, a new era in the fight against HIV infection had begun. In 1996, for the first time, an HIV diagnosis was no longer a death-sentence. This forms a remarkable example of how a simple mathematical model was able to help improve the lives of millions of people.

The second broad category of infectious-disease models are between-host models, where the focus is on the dynamics of how an infection spreads in a population. These models are often developed with the aim of measuring the contribution of certain transmission channels, predicting the impact of health interventions, and ultimately suggesting optimal resource allocation. In the ideal case such models can contribute to policy decisions and facilitate a fact-based balancing of competing public-health priorities. One of the first such mathematical models that the author of this thesis encountered was *Predicting the epidemiological impact of antiretroviral allocation strategies in KwaZulu-Natal: the effect of the urban–rural divide* by David Wilson, James Kahn and Sally Blower [115], which also inspired the author's first publication [65]. The pure model's ethically questionable prediction is that focusing all resources in the cities should have the optimal public-health outcome, which is an important fact that mathematical modeling can make the political leaders aware of when they design the best allocation strategies. Another classic example of mathematical modeling is the very first malaria model by Sir Donald Ross [95]. The pioneer of identifying the mechanism by which malaria spreads developed and used epidemiological models to quantify and extend his laboratory findings. His models were also used to estimate the cost-effectiveness of reducing the mosquito population as a counter-measure to prevent new human malaria cases. Ross received the 1902 Nobel Prize in Medicine for his discoveries in the transmission of malaria, but considered his mathematical models as his most important scientific contribution [58].

In this thesis we use mathematical models to reveal hidden within-host HIV dynamics in the first days after exposure, quantify the population-level impact and cost-effectiveness of facilitating HIV testing in Vancouver, and reveal how to optimally engineer a recently developed anti-malarial agent. In order to understand the background and motivation of these models we first give a more detailed introduc-

tion into the biology of HIV and malaria.

1.2 The biology and modeling of the human immunodeficiency virus

HIV causes the acquired immunodeficiency syndrome (AIDS), which is responsible for an estimated 39 million deaths since its official discovery in 1983. Despite enormous progress since the 1980's, we still do not have a potent response to this worldwide epidemic. The number of annual AIDS related deaths is still in the millions, no cure (neither clinical, nor functional) has been found, and despite enormous effort no broadly effective vaccine has been developed to date. Major reasons for these shortcomings include the immense diversity of HIV due to its high mutation rate, the many ways HIV can be transmitted, e.g., sexually, via contaminated blood (transfusions, needle sharing, occupational) or from mother to child (during pregnancy, birth, or even breastfeeding), the lack of a good animal model, social stigma and false beliefs especially in some countries of the developing world, and the severe resource limitations in the countries that are most affected.

The main targets of HIV are helper T-cells. Ironically and tragically, these cells play an important role in regulating the adaptive immune response, which is responsible for fighting foreign antigens. Infected T-cells die at a heightened rate and, in an untreated individual, their loss can not be compensated for. Hence the number of T-cells drops over the course of infection until the adaptive immune system is so compromised that opportunistic infections can take hold to cause disease and eventual death.

In order to stop the disease progression we have to understand the viral lifecycle. A HIV particle enters a target cell by binding to the CD4 receptor, as well as a chemokine reception, most commonly CCR5 or CXCR4. The binding initiates fusion with the cell membrane so that the genetic material of the virus can enter the host cell. Next, reverse transcriptase transcribes the viral RNA in DNA which is then transported to the cell's nucleus and integrated in the host DNA. The discovery of reverse-transcriptase was a paradigm shift that resulted in a Nobel prize – up to this point the central dogma of molecular biology was that information flows from DNA to RNA to protein, but not in the reverse direction. This process is also

very error-prone and the reason for HIV's high mutation rate that allows the virus to escape the host's immune response and quickly develop drug resistance. The integrated viral DNA gets transcribed into new viral RNA that leaves the nucleus to be assembled to new viral particles. These new virions are soon released by budding or bursting and go on to find new host cells to infect.

In this thesis we study the early *in vivo* viral dynamics of the first few days following HIV exposure (Chapter 2 and Chapter 3). In most cases early HIV infection has very mild or no symptoms, so that many people are not aware of their infection. In the case when someone reports a recent possible exposure to the virus it can be difficult to confirm or exclude infection when the virus population may not have yet grown to a detectable level. This uncertainty is a source of anxiety for clinicians and patients alike [103], and catching infections early can be beneficial for the individual as well as the population in general:

- Evidence from nonhuman primates indicates that HIV reservoirs may be established very early after exposure [29], and that the viral set point, which is established a few weeks after exposure, is an important factor in predicting time to progression to AIDS [78]. Hence an early intervention that failed to prevent infection may still have a positive long-term effect if the seeding of the HIV reservoir had been disrupted or if the viral set point was lowered.

- The importance of early HIV infection on a population level is highlighted by the fact that the proportion of new infections attributed to individuals in the early stage of infection is disproportionately high, even if the exact numerical estimates vary, see Figure 1.1. One reason for this is that recently infected individuals have a large initial viral load, which increases the likelihood of transmitting the infection to a susceptible partner.

Due to the low probability that a single exposure leads to infection, and since the number of virions and cells initially involved is small, the early phase is very difficult to study clinically. Therefore, this is an area where mathematical models can make major contributions to gaining new insights that would be difficult or impossible to obtain otherwise. In fact, the relatively low number of virions and cells may allow approaches that would be numerically unfeasible in more advanced

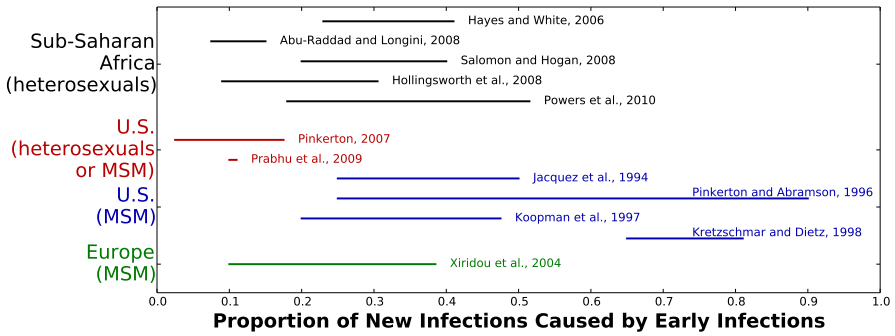


Figure 1.1: Importance of acute HIV infection in new transmissions. Modified from [20].

stages of the infection. The first project of this thesis is motivated by two practical questions from clinicians at HIV testing clinics, *"What is the expected length of time from exposure until the virus grows to a detectable level?"* and *"When should a confirmatory test be scheduled to catch a real infection as early as possible and simultaneously exclude infection if negative?"*.

Having studied the within host viral dynamics around an HIV test we investigate in another project the impact of testing on the disease prevalence in the population. When infected individuals are aware of their infection they can adjust their risk behaviour to reduce the likelihood to transmit the infection onwards. Therefore, increasing the testing rate can be a viable strategy to decrease the disease incidence in a population. This is particularly true in risk groups like the Vancouver men who have sex with men (MSM) community, where prevalence is high and awareness of infection is sufficiently low that many transmission events occur every year. The recently implemented Get Checked Online (GCO) program facilitates HIV testing by providing printable lab forms and test results online. We develop mathematical models to predict the population-level impact of GCO, which we parameterize with survey data from the at-risk population in question.

1.3 Old and new ways to combat and model malaria infection

Malaria is an infectious disease which is caused by the *Plasmodium* parasite, which affects the red blood cells and other tissues of its host. There are many species of malaria, but among those infecting humans the most common are *Plasmodium falciparum* and *Plasmodium vivax* [94].

According to the WHO [114], malaria is endemic in over 100 countries and 40% of the world's population live in malaria-endemic areas. In many developing countries the disease is the leading cause of illness and death. Despite a slightly decreasing trend, there are still more than 225 million annual malaria cases, causing almost 800,000 deaths, which corresponds to 2.23% of all deaths worldwide. 90% of the malaria-caused deaths occur in sub-Saharan Africa, with young children and pregnant women among the groups that are most affected.

The dominant mode of parasite transmission between humans is facilitated by mosquitoes of the *Anopheles* genus. Most mosquito genera cannot actually transmit malaria between humans and only female mosquitoes bite, as they require blood meals for the development of their eggs. The life-cycle of the malaria parasite is surprisingly complex: Upon biting a human, an infected mosquito passes on the parasite as sporozoites through its saliva, where they travel through the bloodstream to the liver. After an incubation period that can range from several days to several months the parasite reaches the merozoite stage where it begins to invade and rupture red blood cells as it reproduces. Once it has begun to infect red blood cells, the merozoite stage progresses to the gametocyte stage where the parasite develops into male and female gametocytes, which can be transmitted back to mosquitoes. When an infected female mosquito bites a human in this stage, the female gametes are fertilized and develop into oocysts on the wall of the mosquito's gut. Over approximately ten days the oocyst ruptures and immature sporozoites leave the gut and enter the salivary glands of the mosquito where they mature and can then be transmitted to a susceptible human again.

The first mathematical models for malaria transmission were two-dimensional compartment models of the infected human and mosquito population. The very first model was developed by Ross in 1911 [94], which was later extended by

Macdonald [73] and Anderson and May [6], among others. More recently Chitnis [18, 19] developed a series of SEIR malaria models where many population and disease dynamic effects are considered in great detail. Others have adjusted and extended these mathematical models to include compartments which distinguish between infected and infectious humans and mosquitoes and to reflect the fact that humans can become (temporarily) immune or treated, investigate the impact of indoor residual spraying include spatial heterogeneities and allow for time-dependent parameters to account for environmental factors such as rainfall and humidity [4, 64, 99, 101].

The most straightforward way to fight human malaria infection is to remove mosquitoes. Since mosquitoes often remain indoors for some time after a blood meal a very promising strategy is indoor residual spraying (IRS) of pesticides (see, e.g., [101] for a simple mathematical model), which had great success in several locations around the world and was even able to eradicate malaria from certain areas. For example, the *National Malaria Eradication Program* was started in 1947 and eradicated malaria in the United States by 1951, and similar stories can be told from regions in India, Sri Lanka and other countries. More historical information can be found in [47], while [114] is a good starting point for more recent developments.

However, pesticides can be toxic for humans if taken in large quantities or when they contaminate agricultural land. For this reason the popular pesticide dichlorodiphenyltrichlo-roethane (DDT) has been banned in most countries by the *Stockholm Convention 2004*. Another disadvantage of pesticides is that mosquitoes can quickly develop resistance, sometimes even within a timescale of months to a few years (see e.g. [6, 10, 56, 109, 114]). Largely due to these difficulties no new pesticides have been approved for malaria control since the 1970's [114].

Our modeling work is inspired by a recent Science paper [31], that proposed a new way of fighting the disease. The authors engineered a *Metarhizium anisopliae* fungus strain that can neutralize malarial sporozoites in the mosquito vector, preventing onward transmission to humans. In laboratory experiments their engineered fungus was able to dramatically reduce the sporozoite counts in the salivary glands of mosquitoes, making malaria-infected mosquitoes less infectious to humans. This new pathogen is particularly appealing because it does not increase the

mosquito's death rate until later life-stages when most reproduction has happened, so we do not expect a rapid development of fungus-resistance in the mosquito population. Further, since this fungus can be engineered in the laboratory we may be able to modify its induced death rate to maximize the intervention's impact on the human malaria prevalence.

Increasing the virulence is likely to induce fungus-resistance, but even if this effect is neglected the optimal virulence is not obvious: On the one hand, the offspring of fungus-infected mosquitoes are potential vectors for transmitting malaria, so increasing the virulence to reduce the total size of the mosquito population could be beneficial. On the other hand, it's the number of malaria-infected mosquitoes that drives the human epidemic, not the total number of mosquitoes. Since fungus-infected and malaria-infected mosquitoes compete for the same resources it may be more beneficial to lower the virulence of the fungus so that the fungus-infected mosquitoes survive longer to compete with susceptible and malaria-infected mosquitoes.

To quantify these effects and trade-offs we built a mathematical model that includes humans and mosquitoes and larvae, and investigate how the prediction of the human malaria prevalence changes when fungus parameters are varied. Hence our model can be used to guide the future development of this engineered pathogen, and ultimately help developing new ways to fight malaria infection.

1.4 How this thesis is structured

In this thesis we present three case studies where mathematical models reveal new insights into the dynamics of HIV and malaria infection. Over the last couple of years we extended well-known results about continuous-time branching processes to model and efficiently analyze the dynamics of the viral load in early HIV infection. We give a coherent analysis of these methods in Chapter 2. One of the key clinical questions in the research of early HIV infection is the optimal timing of early testing, in particular, for how many days after exposure the vial load remains too small to be detectable. Our modeling work in Chapter 3 uses novel mathematical models and viral-load time series data to quantify the length of this, so-called, *eclipse phase*, and also provides quantitative recommendations for tim-

ing of follow-up tests. In the following Chapter 4 we model the population-level impact of Get Checked Online, a program that recently launched in Vancouver and aims to reduce HIV incidence by simplifying how HIV tests are scheduled and how test results are delivered. In Chapter 5 we investigate the effectiveness of a recently proposed anti-malaria fungus and how to engineer its characteristics to optimize the impact in reducing the human population-level malaria prevalence. Finally, we conclude with a discussion of future research directions of our work and mathematical biology in general.

This thesis has three appendices: In Appendix A we analyze variants of the within-host HIV models presented in Chapter 2 that include infected cells, infectious virions and non-infectious virions. Appendix B provides additional details on how the time series were pre-processed in Chapter 3, and also shows the model predictions for each of the different patients. Appendix C gives the detailed proofs of mathematical statements made in Chapter 5 and explains how the model parameters were obtained from the data.

Chapter 2

Mathematical background of branching processes

This chapter introduces the theory of continuous-time branching processes, with a focus on their application to early HIV infection. We are going to derive important properties and formulations which we will apply to a case study in Chapter 3. In contrast to classical textbooks on this topic [39, 46, 59, 60] we do not aim to formulate our findings in the most general way. Instead we follow the practical needs of the application and research questions that we are going to answer with this set of mathematical tools. Our approach allows us to avoid the technical formulations seen elsewhere that are aimed at a statistical and stochastical audience, and hopefully makes the techniques of this chapter easily accessible to the broader mathematical modeling community. Having this community in mind we will derive an implementation-ready formulation of all the main results that can readily be implemented in a numerical computing environment like MATLAB or PYTHON. Even though none of the individual mathematical results presented here are new in a strict sense, they have never been combined, interpreted and applied in this way to be suitable for within-host viral dynamics modeling and to ultimately help answer import clinical questions in this area of research today.

2.1 Introduction

Over the previous years the Coombs HIV group (Alejandra Herrera, Bernhard Konrad, Daniel Coombs and Jessica Conway) has worked on translating theoretical knowledge about branching processes into a more practical form that is ready to implement in a numerical computing environment like MATLAB or PYTHON. We also extended some known results and combined them in new ways to suit our application to within-host HIV modeling. The first application of these techniques was the modeling of HIV dynamics of patients on treatment, see [21]. The key steps used for this foundation are summarized in Section 2.4 and Section 2.5, and further extensions of this method has allowed us to study several HIV-related research questions [22, 50, 51]. One major benefit of the tools developed here is that the popular Gillespie algorithm, which is computationally very time consuming, can be avoided entirely for branching processes. Unfortunately, though, the methods of this chapter only apply to processes that satisfy the branching property.

We start by analyzing the birth-death process in great detail. The key ideas and mathematical justifications for more complex multi-type branching processes are very similar, but unlike these more complicated models the single-type branching process admits an analytical solution which is convenient to verify our more general techniques and results. We first define our notation and state the assumptions that we make for our analysis. Then we derive the Kolmogorov differential equations, also called the master equations, for the probability of each possible state of the system at any time t. As key tool in the analysis of branching processes we then introduce the probability generating function and derive its important properties. Rewriting the Kolmogorov differential equations in terms of the probability generating function we obtain a system of partial differential equations (PDES) and a ODE, respectively. Here is where the branching property is used to simplify the infinite system of ODES to a finite system of ODES, which implies that the probability generating function can readily be approximated by your favorite numerical ODE solver. Using the Cauchy-Euler integral formula we then find an integral expression for the desired corresponding probability distribution.

With this theoretical underpinning we then turn our attention to research questions in early HIV infection: calculating the probability that the size of the virus

population is above a detection threshold, and calculating the time from exposure to detectability of infection. In the latter case we see that conditioning the model on non-extinction is critical for estimating the length of this time period, and so we derive the mathematical framework to handle this variant. We then repeat the analysis for multi-type branching processes to show how the techniques are to be adjusted to handle more sophisticated models.

2.2 The birth-death model

The birth-death model is a classical example of a branching process and introduced in many textbooks, e.g. [39, 60, 67]. In this process the random variable N is integer-valued and events, or reactions, occur at random times drawn from exponential distributions. We define the birth and death rates as b and d, respectively. Then the system can be represented schematically as

$$\begin{aligned} &\text{individual is born} && N \xrightarrow{b} N+1, \\ &\text{individual dies} && N \xrightarrow{d} N-1. \end{aligned}$$

The random variable N counts the number of individuals in the population at time t. We define the probability of having n individuals at time $t \geq \tau$, given an initial population of n_0 individuals at time τ:

$$P(n, n_0, t, \tau) := \mathbb{P}(N(t) = n \,|\, N(\tau) = n_0).$$

Our first assumption is that the system behaves like a stationary Markov chain. That is, the future is independent of the past, given the present (Markov property)and $P(n, n_0, t, \tau)$ only depends on the difference $t - \tau$, not the absolute time: $P(n, n_0, t, \tau) = P(n, n_0, t - \tau)$ (stationary).

The second assumption is that the probability of an event in a population of a single individual in the time span of length h is, to first order, proportional to the reaction rate and the duration h:

$$\begin{aligned} P(2, 1, h) &= bh + o(h) \\ P(0, 1, h) &= dh + o(h) \end{aligned}$$

$$P(1,1,h) = 1-(b+d)h+o(h)$$

The probability of more than one event is assumed to be $o(h)$. Here $o(h)$ denotes functions f for which $\lim_{h\to 0}\frac{f(h)}{h} = 0$. When we further assume that individuals act independently we can generalize this assumption to a population of size n. For example, a net gain of one individual in a time interval of length h occurs exactly when one individual gives birth and no other event occurs. All other ways to achieve a net-gain of one individual are of order $o(h)$. Therefore

$$\begin{aligned} P(n+1,n,h) &= n(bh+o(h))(1-(b+d)h+o(h))^{n-1}+o(h) \\ &= nbh+o(h), \end{aligned}$$

and similarly, for a net loss of one individual,

$$\begin{aligned} P(n-1,n,h) &= n(dh+o(h))(1-(b+d)h+o(h))^{n-1}+o(h) \\ &= ndh+o(h). \end{aligned}$$

Note also that the two assumptions together imply that the waiting time to the next event is exponentially distributed. Indeed, let T_1 be the time at which the first new individual is born, and define the waiting time $W(t) = \mathbb{P}(T_1 \geq t)$. By definition $W(0) = P(T_1 \geq 0) = 1$ and using the assumptions we find

$$\begin{aligned} W(t+h) &= \mathbb{P}(T_1 \geq t+h) = \mathbb{P}(T_1 \geq t)\mathbb{P}(T_1 \geq t+h \,|\, T_1 \geq t) \\ &= \mathbb{P}(T_1 \geq t)\mathbb{P}(T_1 \geq h \,|\, T_1 \geq 0) \\ &= W(t)W(h) \\ &= W(t)(1-bnh+o(h)). \end{aligned}$$

Subtracting $W(t)$ on both sides and then dividing by h and letting h tend to zero this implies $W'(t) = -bnW(t)$. Solving this differential equation we find that $W(t) = e^{-bnt}$, and hence the waiting time for the birth event is exponentially distributed. The death event is handled analogously. Therefore, the waiting time until the next event is also exponentially distributed, as the minimum of exponentially distributed random variables.

2.3 Gillespie's Algorithm - a powerful but costly stochastic simulation algorithm

Developed by Doob [27] as early as 1945, Gillespie's algorithm became popular in 1977 when Gillespie published what became a landmark paper for stochastic simulations [37]. The algorithm realizes an exact solution path of the evolution of the population size over time. It works equally well for branching processes and more complicated Markov processes where the waiting time for reactions is exponentially distributed, and is a very popular algorithm in many different areas of mathematical biology and beyond. Unlike most comparable procedures, every reaction is explicitly simulated and no small time steps are approximated by larger time steps of fixed length. Gillespie himself was motivated by chemical reactions of spatially uniformly mixed species in a fixed volume. Therefore a possible reaction is sometimes referred to as *reaction channel*.

The algorithm works as follows. At time t calculate the rate of each possible reaction p_j, for $j = 1, \ldots, n$. Since the waiting time for each reaction is exponentially distributed with parameter p_j, the time until the next reaction is also exponentially with parameter $p_0 = \sum_{j=1}^{n} p_j$, as the minimum of exponentially distributed random variables. Given this rate the time until the next event is obtained by drawing a random number r_1 from an exponential distribution with parameter p_0, that is, the time of the next event is $t + r_1$. The likelihood that the next event is reaction j is proportional to its relative probability p_j/p_0, so draw a second random number that is distributed according to these probabilities. In practice this is achieved by drawing a uniformly distributed second random number r_2 and updating the population size according to the k-th reaction, where k is the integer that satisfies $\sum_{j=1}^{k-1} p_j/p_0 < r_2 \leq \sum_{j=1}^{k} p_j/p_0$. Repeat the procedure at $t = t + r_1$ until a desired end-time or population size is reached.

Since each simulation is just one possible realization of the underlying stochastic process, usually many thousand or millions of simulations have to be run to obtain meaningful summary statistics. This becomes particularly time-consuming when reactions occur very frequently. This is a common problem when the population size, and hence the reaction rates become very large. Another problem with the basic version of Gillespie's algorithm is that many simulations have to be run

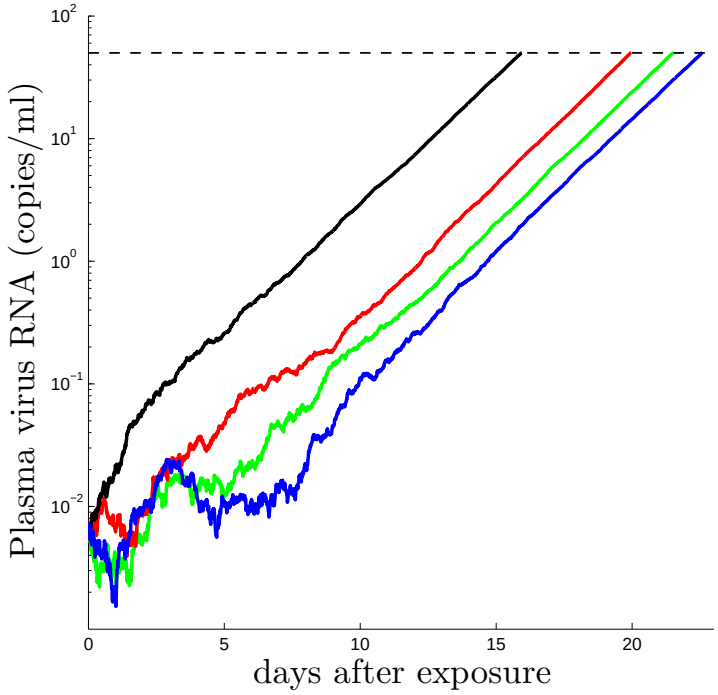


Figure 2.1: Sample paths of the stochastic birth-death process, started with a population size of 100 and ran until a desired threshold is reached. Birth rate $b = 23.9$ day^{-1}, death rate $d = 23$ day^{-1} and threshold 375,000. This threshold corresponds to the HIV detection limit of 50 RNA copies/ml, at 2 RNA copies per virion in a human with 15L of extracellular fluid.

to observe rare events. Solutions to these problems, such as a variety of "leaping" methods as well as importance sampling, have been suggested and remain an active area of research [9, 16, 26, 38, 55, 68, 72, 79, 120]. Figure 2.1 shows the result of four Gillespie simulations of the birth-death process. In the remainder of this chapter we avoid Gillespie's algorithm altogether and instead derive a mathematical method to obtain the probability distribution more directly.

2.4 Derivation of the master equations

Our goal is to calculate the probability $P(n,n_0,t)$ for all times t and population sizes n. Note that it suffices to consider a fixed initial population size n_0, that is, $P(n,n_0,0) = \mathbb{1}_{n,n_0}$, where $\mathbb{1}_{a,b}$ is the indicator function which is 1 when $a = b$ and 0 otherwise. Indeed, if the initial condition is distributed according to a probability distribution $(\rho_{n_0})_{n_0\geq 0}$ then $\mathbb{P}(N(t) = n \,|\, N(0) = \rho_{n_0}) = \sum_{n_0=0}^{\infty} P(n,n_0,t)\rho_{n_0}$. The first step towards solving for $P(n,n_0,t)$ is to consider a time interval $(0,t+h)$ and split it into a long step $(0,t)$ and a short step $(t,t+h)$. Then, for the total population size to change from n_0 to n in $(0,t+h)$, it could have been at any population size k at time t and then transitioned from k to n in the remaining time. This leads to the Chapman-Kolmogorov equation

$$\begin{aligned} P(n,n_0,t+h) &= \sum_{k=0}^{\infty} P(k,n_0,t)P(n,k,h) \\ &= P(n-1,n_0,t)P(n,n-1,h) + P(n+1,n_0,t)P(n,n+1,h) \\ &\quad + P(n,n_0,t)P(n,n,h) + o(h) \\ &= b(n-1)hP(n-1,n_0,t) + d(n+1)hP(n+1,n_0,t) \\ &\quad + (1-(b+d)nh)P(n,n_0,t) + o(h). \end{aligned}$$

Next, we subtract $P(n,n_0,t)$ from both sides, divide the resulting equation by h and let $h \to 0$, keeping in mind that $\lim_{h\to 0} o(h)/h = 0$, to obtain

$$\begin{aligned} \frac{d}{dt}P(n,n_0,t) &= \lim_{h\to 0} \frac{P(n,n_0,t+h) - P(n,n_0,t)}{h} \\ &= b(n-1)P(n-1,n_0,t) + d(n+1)P(n+1,n_0,t) \\ &\quad - (b+d)nP(n,n_0,t), \\ P(n,n_0,0) &= \mathbb{1}_{n,n_0}. \end{aligned} \tag{2.1}$$

This equation is a forward master equation, or Kolmogorov forward equation. It can nicely be interpreted as a balance equation: *[rate of change of state n] = [sum of reactions to get into state n] - [sum of reactions to leave state n].*

In the above we could have also chosen to split the interval $(0,t+h)$ into a small step $(0,h)$ followed by a larger step $(h,t+h)$. In order for the population

size to change from n_0 at time 0 to n at time $t+h$ it could have been at any size k at time h, which leads to the Chapman-Kolmogorov equation

$$\begin{aligned} P(n,n_0,t+h) &= \sum_{k=0}^{\infty} P(k,n_0,h)P(n,k,t) \\ &= P(n_0+1,n_0,h)P(n,n_0+1,t) + P(n_0-1,n_0,h)P(n,n_0-1,t) \\ &\quad + P(n_0,n_0,h)P(n,n_0,t) + o(h) \\ &= bn_0hP(n,n_0+1,t) + dn_0hP(n,n_0-1,t) \\ &\quad + (1-(b+d)n_0h)P(n,n_0,t) + o(h). \end{aligned}$$

We again subtract $P(n,n_0,t)$ from both sides, divide the resulting equation by h and let $h \to 0$, keeping in mind that $\lim_{h\to 0} o(h)/h = 0$, to obtain

$$\begin{aligned} \frac{d}{dt}P(n,n_0,t) &= bn_0P(n,n_0+1,t) + dn_0P(n,n_0-1,t) - (b+d)n_0P(n,n_0,t), \\ P(n,n_0,t) &= \mathbb{1}_{n,n_0}. \end{aligned} \tag{2.2}$$

This equation is called the backward master equation, or Kolmogorov backward equation.

The two sets of master equations are equivalent systems of infinitely many coupled ODES. In very special cases, e.g. in the pure birth or pure death process, these equations can be solved recursively for P. The key to finding P for more sophisticated models is the probability generating function, which we introduce in the next section.

2.5 The probability generating function

We gain a lot of insight into our model when we can solve Kolmogorov's forward or backward differential equation. Unfortunately, a straightforward approach is not possible as both formulations describe a system of infinitely many, highly interconnected linear ODES, and we generally do not have a natural first equation we can solve to proceed inductively. Instead we solve all equations at once using the probability generating function, a power-series expansion with the probabilities as its coefficients.

2.5.1 Definition and basic properties

Let $\overline{\mathbb{D}} = \{z \in \mathbb{C} : |z| \leq 1\}$ and define the multi-variable probability generating function

$$G : \mathbb{N}_0 \times [0,\infty) \times \overline{\mathbb{D}} \to \overline{\mathbb{D}}, \qquad (n_0,t,z) \mapsto \sum_{n=0}^{\infty} P(n,n_0,t) z^n. \tag{2.3}$$

If we again let N be the random variable of the number of individuals at time t in a population that started with n_0 individuals at time $t = 0$ we can write the probability generating function as $G(n_0,t,z) = E[z^N]$. This observation is particularly useful for proving mathematical properties of the probability generating function. Further note that the right hand side is well defined and converges absolutely by the triangle inequality since $\sum_{n=0}^{\infty} P(n,n_0,t) = 1$ holds for all $t \in [0,\infty)$. In particular, we have that $G(n_0,t,1) = 1$. The probability generating function has a lot of useful properties:

1. First and foremost, knowing the probability generating function is equivalent to knowing the probabilities P, in which we are interested.[1] To obtain P from G we make use of the identity

$$\frac{\partial^k}{\partial z^k} z^n = \begin{cases} \frac{n!}{(n-k)!} z^{n-k}, & \text{if } k \leq n, \\ 0, & \text{if } k > n, \end{cases}$$

   so that when $z = 0$ is plugged in we have

$$\left.\frac{\partial^k z^n}{\partial z^k}\right|_{z=0} = n! \mathbb{1}_{n,k}.$$

   Therefore, we can obtain $P(n,n_0,t)$ by n-time differentiation of the definition $G(n_0,t,z) = \sum_{n=0}^{\infty} P(n,n_0,t) z^n$:

$$P(n,n_0,t) = \frac{1}{n!} \left.\frac{\partial^n G(n_0,t,z)}{\partial z^n}\right|_{z=0}. \tag{2.4}$$

2. Setting $n = 0$ in the identity above, we obtain the likelihood $P(0,n_0,t)$ that

[1] This is the main reason why G is sometimes called the "wonderful function" [67].

the population is extinct at time t as $G(n_0,t,0)$.

3. We obtain the mean value of the population size by evaluating

$$E[N] = \sum_{n=0}^{\infty} nP(n,n_0,t) = \sum_{n=0}^{\infty} nP(n,n_0,t)1^{n-1} = \left.\frac{\partial G(n_0,t,z)}{\partial z}\right|_{z=1}. \quad (2.5)$$

4. In fact, we can calculate any higher order moments of N by taking the corresponding number of derivatives. For example, taking two derivatives yields

$$\left.\frac{\partial^2 G(n_0,t,z)}{\partial z^2}\right|_{z=1} = \sum_{n=0}^{\infty} (n^2-n)P(n,n_0,t) = E[N^2] - E[N],$$

which implies that the variance of N is given by

$$\begin{aligned}\mathrm{Var}(N) &= E[N^2] - (E[N])^2 = (E[N^2] - E[N]) + E[N] - (E[N])^2 \\ &= \left.\left(\frac{\partial^2 G(n_0,t,z)}{\partial z^2} + \frac{\partial G(n_0,t,z)}{\partial z} - \left(\frac{\partial G(n_0,t,z)}{\partial z}\right)^2\right)\right|_{z=1}.\end{aligned}$$

2.5.2 The differential equation formulations

Just as there are two master equations, there are also two differential equations that the probability generating function satisfies. The forward master equation leads to a PDE, while the backward master equation leads to a system of infinitely many ODES. Starting with the forward master equation (2.1)

$$\begin{aligned}\frac{d}{dt}P(n,n_0,t) = {} & b(n-1)P(n-1,n_0,t) + d(n+1)P(n+1,n_0,t) \\ & - (b+d)nP(n,n_0,t)\end{aligned}$$

multiply both sides by z^n and sum over all $n \in \mathbb{N}_0$ to obtain

$$\begin{aligned}\sum_{n=0}^{\infty} \frac{d}{dt}P(n,n_0,t)z^n = \sum_{n=0}^{\infty} \big(& b(n-1)P(n-1,n_0,t)z^n + d(n+1)P(n+1,n_0,t)z^n \\ & - (b+d)nP(n,n_0,t)z^n\big).\end{aligned}$$

We now rewrite this term-by-term, making use of the fact that the probability of a negative population size is zero, $P(n,n_0,t)=0$ for all $n<0$.

1. $\sum_{n=0}^{\infty}\frac{d}{dt}P(n,n_0,t)z^n=\partial G/\partial t$,
2. $\sum_{n=0}^{\infty}b(n-1)P(n-1,n_0,t)z^n=b\sum_{k=0}^{\infty}kP(k,n_0,t)z^{k+1}=bz^2\partial G/\partial z$,
3. $\sum_{n=0}^{\infty}d(n+1)P(n+1,n_0,t)z^n=d\sum_{k=0}^{\infty}kP(k,n_0,t)z^{k-1}=d\partial G/\partial z$,
4. $\sum_{n=0}^{\infty}(b+d)nP(n,n_0,t)z^n=(b+d)z\sum_{k=0}^{\infty}kP(k,n_0,t)z^{k-1}=(b+d)z\partial G/\partial z$.

Combining the results above we obtain the PDE formulation for the probability generating function

$$\begin{aligned}\frac{\partial G(n_0,t,z)}{\partial t}&=\left(bz^2+d-(b+d)z\right)\frac{\partial G(n_0,t,z)}{\partial z},\\ G(n_0,0,z)&=z^{n_0}.\end{aligned}\tag{2.6}$$

Applying the same procedure to the backward master equation (2.2) instead we find

$$\begin{aligned}\sum_{n=0}^{\infty}\frac{d}{dt}P(n,n_0,t)z^n=\sum_{n=0}^{\infty}\big(&bn_0P(n,n_0+1,t)z^n+dn_0P(n,n_0-1,t)z^n\\&-(b+d)n_0P(n,n_0,t)z^n\big).\end{aligned}$$

Notice that, unlike for the forward equations, there is no need to shift the indices, so we obtain directly

$$\begin{aligned}\frac{\partial G(n_0,t,z)}{\partial t}&=bn_0G(n_0+1,t,z)+dn_0G(n_0-1,t,z)-(b+d)n_0G(n_0,t,z),\\ G(n_0,0,z)&=z^{n_0}.\end{aligned}\tag{2.7}$$

This is an infinite system of ODES. The crucial step and a key asset of branching processes is that we can reduce this infinite system to a finite system of ODES by using the branching property: Individual branches act independently. Let N_1, N_2 be two independent birth-death processes that both start with a single individual.

Then, using the independence in the third step, we can calculate directly:

$$\begin{aligned}
E[z^{N_1+N_2}] &= \sum_{n=0}^{\infty} \mathbb{P}(N_1+N_2=n)z^n \\
&= \sum_{n=0}^{\infty}\sum_{k=0}^{n} \mathbb{P}(N_1=k, N_2=n-k)z^n \\
&= \sum_{n=0}^{\infty}\sum_{k=0}^{n} \mathbb{P}(N_1=k)\mathbb{P}(N_2=n-k)z^n \\
&= \sum_{n=0}^{\infty}\sum_{k=0}^{\infty} \mathbb{P}(N_1=k)z^k\mathbb{P}(N_2=n-k)z^{n-k} \\
&= \sum_{k=0}^{\infty} \mathbb{P}(N_1=k)z^k \sum_{n=0}^{\infty} \mathbb{P}(N_2=n-k)z^{n-k} \\
&= \left(\sum_{k=0}^{\infty} \mathbb{P}(N_1=k)z^k\right)\left(\sum_{\ell=0}^{\infty} \mathbb{P}(N_2=\ell)z^{\ell}\right) \\
&= E[z^{N_1}]E[z^{N_2}].
\end{aligned}$$

Generalizing this to a population of size n_0 we obtain that

$$G(n_0,t,z) = (G(1,t,z))^{n_0} \tag{2.8}$$

for all $n_0 \in \mathbb{N}_0$. It therefore suffices to solve (2.7) for $G(1,t,z)$ only:

$$\begin{aligned}
&\frac{\partial G(1,t,z)}{\partial t} = bG^2(1,t,z) + d - (b+d)G(1,t,z), \\
&G(1,0,z) = z.
\end{aligned} \tag{2.9}$$

2.5.3 Calculating the probability distribution from the probability density function

We now have two differential equations for the probability density function, the PDE (2.6) and the ODE (2.9). The PDE route is more general, because it can always be derived for more general stochastic processes. However, PDES are much more difficult to solve, even numerically. To derive the ODE we used the extra assumption of the branching property, and obtained a standard ODE whose solution can

easily be approximated numerically. In fact, in the special case of the birth-death process either equation can be solved analytically: The PDE by using the method of characteristics, the ODE is a Riccati differential equation. Either way, we obtain the analytical solution of the probability generating function for the birth-death process as

$$G(n_0,t,z) = \left(\frac{\sigma(t) - 1 + (\gamma - \sigma(t))z}{\gamma\sigma(t) - 1 + \gamma z(1-\sigma(t))} \right)^{n_0}, \tag{2.10}$$

where $\gamma = b/d \neq 1$ and $\sigma(t) = e^{(b-d)t}$, see e.g., [39].

Now that we know the probability generating function it is time to get back the probability distribution of the population size. Recall from (2.4) that the probability distribution can be obtained by evaluating the corresponding derivative of G at $z = 0$. Again, in this special case we can obtain an analytical expression for $P(n,n_0,t)$:

$$P(n,n_0,t) = \sum_{j=0}^{\min(n_0,n)} \binom{n_0}{j} \binom{n_0+n-j-1}{n_0-1} \alpha^{n_0-j} \beta^{n-j} (1-\alpha-\beta)^j,$$

where $\alpha = \frac{\sigma(t)-1}{\gamma\sigma(t)-1}$ and $\beta = \gamma\alpha$, again see e.g., [39].

It is great to have an analytic expression for $P(n,n_0,t)$, but the large sum and huge factorial factors make the above formula unsuitable for numerical algorithms, unless n and n_0 are fairly small. Further, such an analytic solution is only possible if the analytic solution of the generation function G is available, which is not the case in general. However, a numerical approximation of the corresponding generating function of any branching processes is always available by numerically solving the corresponding ODE. Hence we set the analytical solutions for G and $P(n,n_0,t)$ aside for now, and derive a computationally efficient method to find the probability distribution that works for more general branching processes.

An accurate numeric approximation of the solution to the ODE formulation (2.9) is easy to obtain, so we focus on how to translate that to the probability distribution. We could just use (2.4) straight out and start taking derivatives. However, numerical differentiation is unstable, especially when thousands or even millions of derivatives are required. Instead, numerical integration is stable, so we remember Cauchy's integral formula from complex analysis that links differentiation with

integration:

$$\left.\frac{d^n f}{dx^n}\right|_{x=0} = \frac{n!}{2\pi i}\oint_C \frac{f(z)}{(z-a)^{n+1}}\,dz.$$

We choose $f(z) = f_{n_0,t}(z) = G(n_0,t,z)$, which is a holomorphic function inside the complex unit circle and continuous on its closure. Since we eventually want to evaluate the derivatives at $a = 0$ we let C be the circle with radius 1 around the origin in the complex plane and obtain

$$\begin{aligned} P(n,n_0,t) &= \frac{1}{n!}\left.\frac{\partial^n G(n_0,t,z)}{\partial z^n}\right|_{z=0} \\ &= \frac{1}{n!}\left(\frac{n!}{2\pi i}\oint_C \frac{G(n_0,t,z)}{z^{n+1}}dz\right) \\ &= \frac{1}{2\pi}\int_0^{2\pi} G(n_0,t,e^{i\varphi})e^{-in\varphi}d\varphi, \end{aligned}$$

where we substituted $z = e^{i\varphi}$. Next we split the integral at π and substitute $\alpha = 2\pi - \theta$ in the second summand. Finally, $G(1,t,\bar{z}) = \sum_n P(n,n_0,t)\bar{z}^n = \overline{G(1,t,z)}$, $\overline{e^z} = e^{-z}$, and $2\,\mathrm{Re}(z) = z + \bar{z}$ leads to

$$P(n,n_0,t) = \frac{1}{\pi}\,\mathrm{Re}\left(\int_0^{\pi} G(n_0,t,e^{i\varphi})e^{-in\varphi}d\varphi\right). \tag{2.11}$$

This is the single most important formula of this section, because it gives us a numerically effective formula to obtain the probabilities of interest, without having to rely on Gillespie simulations. In Figure 2.2 we see that the probability as calculated using (2.11) matches the aggregated results from Gillespie simulation very well, but is computationally much cheaper. The "trick" of using Cauchy's integral formula to avoid numerical differentiation became popular in the late 60's, but to our knowledge it was first employed on the generating function of a process that models HIV dynamics by Conway and Coombs [21].

2.6 Mean behaviour of the stochastic process

In a branching process the average of the random variable is the solution to the corresponding deterministic ODE. The typical approach is to consider the ODE first,

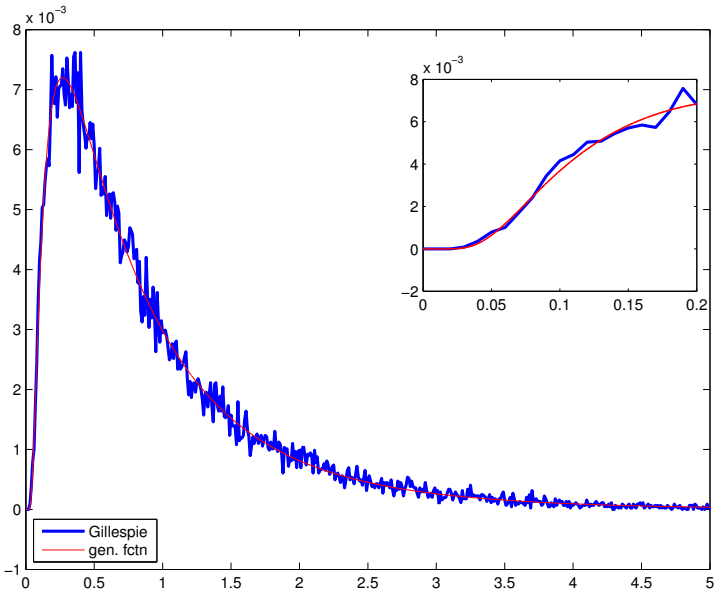


Figure 2.2: The probability $P(150, 100, t)$ of population size $N(t) = 150$ at time t for a population that starts at $N(0) = 100$ individuals, with death rate $d = 23$ day^{-1} and birth rate $b = 23.9$ day^{-1}. The blue line is the aggregate from Gillespie simulations, the red line is obtained by integrating the probability generating function as described in Equation (2.11).

and then translate the process into its stochastic variant. However, with the probability generating function we can also go the other way. This allows a verification that the stochastic variant matches the process that the ODE describes, and, equally importantly, the derivation of an expression for the deterministic mean behaviour in cases where this is not immediately obvious.

To begin with, we verify that the mean behaviour of the birth-death process indeed corresponds to the deterministic ODE. We fix n_0 and t, and recall from (2.5)

$$E[N] = \sum_n nP(n, n_0, t) \left. \frac{\partial G(n_0, t, z)}{\partial z} \right|_{z=1}.$$

From the analytical solution (2.10) of the generating function of the birth-death process we easily verify that indeed $E[N] = n_0 e^{(b-d)t}$, that is, $E[N]$ solves the ODE

$E[\dot{N}] = (b-d)E[N]$, $E[N](0) = n_0$.

In fact, even if the analytical solution of the probability were not available we could still derive the ODE of the mean of the stochastic process. Set $n_0 = 1$, let $f(t) = E[N_1(t)] = \frac{\partial G(1,t,z)}{\partial z}\Big|_{z=1}$ and take the time-derivative to obtain

$$\begin{aligned}
\frac{df}{dt} &= \frac{d}{dt}\left(\frac{\partial G(1,t,z)}{\partial z}\Big|_{z=1}\right) \\
&= \frac{\partial}{\partial z}\frac{d}{dt}G(1,t,z)\Big|_{z=1} \\
&= \frac{\partial}{\partial z}\left(bG^2(1,t,z) + d - (b+d)G(1,t,z)\right)\Big|_{z=1} \\
&= 2b\,\frac{\partial G(1,t,z)}{\partial z}\Big|_{z=1} - (b+d)\,\frac{\partial G(1,t,z)}{\partial z}\Big|_{z=1} \\
&= (b-d)f(t)
\end{aligned}$$

with initial condition $f(0) = \frac{\partial G(1,0,z)}{\partial z}\Big|_{z=1} = 1$. The solution to this differential equation is given by $f(t) = E[N_1(t)] = e^{(b-d)t}$. When we extend this result to an initial population size of n_0 we find that $E[N(t)] = \frac{\partial G(n_0,t,z)}{\partial z}\Big|_{z=1} = n_0 e^{(b-d)t}$, which agrees with the solution we found earlier.

2.7 Calculating cumulative probabilities

We can also quickly calculate the probability that the total population size is above or below a certain threshold. Having the detection limit of an HIV test in mind we call this threshold dL and want to know the probability of being above dL at time t. This random variable tells us when the HIV test of an infected individual would be positive, and is therefore of great importance for the clinical questions on the length of the eclipse phase that we will discuss in Chapter 3.

The straightforward way to calculate the probability that the detection limit is below the threshold is to sum all probabilities of being at any value n, for $n = 0, 1, \ldots, \mathrm{dL} - 1$. For large values of dL, which is the case for the detection limit dL, calculating this cumulative probability involves hundreds of thousands of summands. Instead, a slight algebraic manipulation allows us to replace this large

sum of integrals with a single integral:

$$\begin{aligned}\mathbb{P}(N(t) < \mathrm{dL} \,|\, N(0) = n_0) &= \sum_{n=0}^{\mathrm{dL}-1} P(n, n_0, t) \\ &= \sum_{n=0}^{\mathrm{dL}-1} \left(\frac{1}{\pi} \mathrm{Re} \int_0^{\pi} G(n_0, t, e^{i\varphi}) e^{-in\varphi} d\varphi \right) \\ &= \frac{1}{\pi} \mathrm{Re} \int_0^{\pi} G(n_0, t, e^{i\varphi}) \sum_{n=0}^{\mathrm{dL}-1} (e^{-i\varphi})^n d\varphi \\ &= \frac{1}{\pi} \mathrm{Re} \int_0^{\pi} G(n_0, t, e^{i\varphi}) \frac{1 - e^{-i\mathrm{dL}\varphi}}{1 - e^{-i\varphi}} d\varphi.\end{aligned}$$

Conveniently, a slight adjustment of the integrand suffices to turn the calculation from an individual probability to the cumulative probability. On a technical note, if the numerical integrator evaluates the above integral at the left endpoint $\varphi = 0$, each summand in the geometric series is unity, and hence at this one time point the integrand needs to be replaced by $G(n_0, t, 1)\mathrm{dL}$.

2.8 The probability of extinction and the basic reproduction number R_0

In order to calculate the probability of extinction, again consider the process that starts with a single individual. Let $q_s = P(0, 1, s)$ and $q = \lim_{s \to \infty} q_s$ be the probability that the population is extinct at time s and eventually, respectively. Note that q is well-defined because q_s is increasing in s and bounded by 1. Next, let t be the time when the initial individual dies. The corresponding waiting time $W(\tau)$ $= \mathbb{P}(t \geq \tau)$ is exponentially distributed with parameter d, the death rate. It follows that

$$\begin{aligned} q_s &= P(0, 1, s) \\ &= \mathbb{P}(\text{extinction at } s \,|\, t > s) \times \mathbb{P}(t > s) + \mathbb{P}(\text{extinction at } s \,|\, t \leq s) \times \mathbb{P}(t \leq s) \\ &= 0 \times e^{-ds} + \sum_{n=0}^{\infty} P(n, 1, t) q_{s-t}^n (1 - e^{-ds}). \end{aligned} \tag{2.12}$$

The last line requires justification. If the initial individual did not die by time s, then the population could not have gone extinct. Otherwise, we condition on the population size at time t, and require that each of the identical and independent offspring branching processes goes extinct in the remaining time span $s-t$. Letting s go to infinity yields

$$q = \lim_{s\to\infty} q_s = \lim_{s\to\infty} \sum_{n=0}^{\infty} P(n,1,t) q_{s-t}^n (1-e^{-ds}) = \sum_{n=0}^{\infty} P(n,1,t) q^n = G(1,t,q). \tag{2.13}$$

We now want to show that q is the *smallest* positive number x that satisfies $G(1,t,x) = x$. To this end, define the helper function $h(x) = G(1,t,x) - x$, which has the following properties

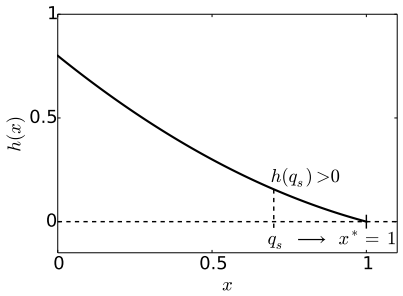
1. $h(0) = P(0,1,t) \geq 0$, $h(1) = 0$,
2. $h'(0) \leq 0$, $h'(1) = E[N] - 1$,
3. h is convex, $h''(x) \geq 0$.

So $x = 1$ is always a root of h and since h is convex it has at most one additional root in $[0,1)$, see Figure 2.3. In fact, this additional root exists if and only if $h'(1) = E[N] - 1 > 0$ which corresponds to the case when the population is expected to grow on average.

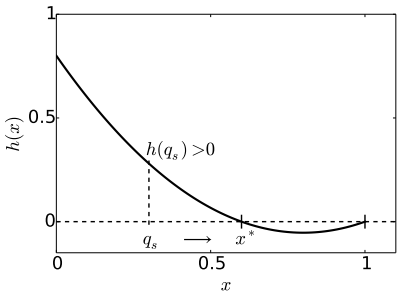
If we assume $E[N] \leq 1$, then the only root of h in $[0,1]$ is 1, so since $q \in [0,1]$ and $h(q) = 0$ we have $q = 1$. This case corresponds to guaranteed extinction. Alternative, assume $E[N] > 1$, so that the population grows on average. Then h has an additional root in $[0,1]$, which we denote by $x^* < 1$. Since $q \geq 0$ is also root of h we obtain $q \geq x^*$. From the convexity of h it follows that $h(x) \geq 0$ for $0 < x \leq x^*$ and $h(x) < 0$ for $x^* < x < 1$. Since q_s is increasing $q_{s-t} \leq q_s$, so that (2.12) yields

$$q_s = \sum_{n=0}^{\infty} P(n,1,t) q_{s-t}^n (1-e^{-ds}) \leq \sum_{n=0}^{\infty} P(n,1,t) q_s^n (1-e^{-ds}) = G(1,t,q_s),$$

or, equivalently, $h(q_s) \geq 0$. Therefore $q_s \leq x^*$ for all $s \geq 0$, and by continuity of q_s we obtain $q = \lim_{s\to\infty} q_s \leq x^*$. Therefore $q = x^*$ is the indeed that smallest root of



(a) The helper function $h(x) = G(1,t,x) - x$ has only a single root $x^* = 1$ in $[0,1]$ if $h'(1) = E[N] - 1 \leq 0$. In this case $q = 1$ and extinction is guaranteed.



(b) The helper function $h(x) = G(1,t,x) - x$ has two roots in $[0,1]$ if $h'(1) = E[N] - 1 > 0$. In this case $q < 1$ and the population may survive indefinitely.

Figure 2.3: Whether the population necessarily ultimately goes extinction depends on whether the average growth rate $E[N]$ is larger that 1.

h, as claimed. In summary, if the population grows on average we have $q < 1$, that is, extinction is not guaranteed and the population may survive indefinitely.

Taking the time-derivative on both sides of (2.13) we see that the probability of extinction corresponds to the steady states of the ODE (2.9) of the probability generating function. In our case this yields

$$0 = bq^2 + d - (b+d)q \quad \Longleftrightarrow \quad q = 1 \text{ or } q = d/b. \tag{2.14}$$

When the death rate is not exceeded by the birth rate, $d/b \geq 1$, we have $q = 1$ and extinction is guaranteed. Otherwise $q = d/b < 1$ and there is a positive probability of $1 - d/b$ that the population survives indefinitely. Further note that, since each individuals act independently due to the branching property, the probability that a population of size n individuals goes extinct equals q^n, the n-th power of the probability that a population of a single individual goes extinct.

The basic reproduction number R_0 is defined as the expected number of descendants of a single individual. In the birth-death process each individual gives birth at rate b over their expected life-span of $1/d$, hence $R_0 = b/d$. For a more systematic derivation investigate the stability of the trivial steady state of the cor-

responding ODE. Here $\dot{N} = (b-d)N$, which has the trivial steady state $N = 0$ and is stable if and only if $R_0 = b/d < 1$.

2.9 Conditional on non-extinction

In a stochastic process where the "extinct state" is absorbing there is always a positive probability that the population goes extinct, regardless of the basic reproduction number R_0. If $R_0 \leq 1$, however, the population almost surely goes extinct eventually. In the example of HIV, most exposures do not lead to infection, that is, the viral population goes extinct. However, extinction is not guaranteed and hence $1 < R_0 = b/d$ and the average growth rate $b-d$ is positive. However, here the *infection bias* becomes critical: only cases where exposure did in fact lead to infection are clinically observed! This means that we have to adjust our stochastic model to reflect this bias when we are fitting it to clinical data.

In Gillespie simulations it is easy to bias the calculations accordingly – simply disregard all realizations that went extinct eventually. The effect of modifying $R_0 = b/d$ (and hence the likelihood of extinction $q = d/b$) while keeping the average slope $b-d$ constant and simultaneously conditioning on non-extinction is highlighted in Figure 2.4, where we use Gillespie simulations with different parameters to estimate the time until a fixed population size is reached.

To better understand the effect of conditioning on non-extinction we derive a method to calculate the mean of the conditioned process without having to fall back to extensive Gillespie simulations. We are interested in the probability that the population reaches size n at time t under the additional condition that it does not go extinct later:

$$P(N(t) = n \,|\, N(\infty) \neq 0, N(0) = n_0).$$

Note that $N(\infty) = 0$ if and only if $N(s) = 0$ for some $s \geq 0$ (the state space is discrete). Also note thtamer conditioning on an event that has probability zero is not a valid operation. Hence, none of the calculations below make sense if $R_0 < 1$ and extinction is guaranteed.

Since the event we condition on is in the future, but the Markov property works for events in the past, we use *Bayes theorem* $P(A|B,C) = \frac{P(B|A,C)P(A|C)}{P(B|C)}$ to rewrite

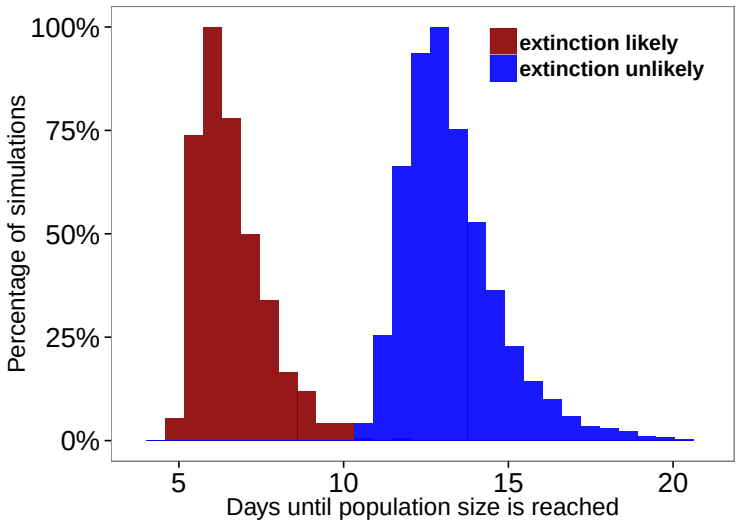


Figure 2.4: Distributions of the time until a population size of 375,000 is reached, obtained from Gillespie simulations of the birth-death process with $b - d = 0.9$ day^{-1}. Paths that did not reach the desired population size were discarded, that is, the processes are conditioned on non-extinction. Different death rates $d = 9 \times 10^4$ day^{-1} (red) and $d = 23$ day^{-1} (blue) result in a different likelihood of extinction $q = d/b$. The larger the likelihood of extinction, the faster the conditioned process reaches the desired population size.

the above as

$$\begin{aligned} &P(N(t) = n \,|\, N(\infty) \neq 0, N(0) = n_0) \\ &\qquad = P(N(t) = n \,|\, N(0) = n_0) \frac{P(N(\infty) \neq 0 \,|\, N(t) = n, N(0) = n_0)}{P(N(\infty) \neq 0 \,|\, N(0) = n_0)}. \end{aligned}$$

We recognize the first term as our original, unconditioned probability from earlier. For this we know how to calculate the probability generating function. The numerator can be simplified to $P(N(\infty) \neq 0 | N(t) = n)$, due to the Markov property of P that history can be neglected if the present state is known, and then further to $P(N(\infty) \neq 0 | N(0) = n)$, due to the time invariance of P that only the relative time difference between the two events matters, not their absolute time. Finally, the complementary event that a population of size n does not go extinct is very easy to

calculate.

$$P(N(\infty) \neq 0 \,|\, N(0) = n) = 1 - P(N(\infty) = 0 \,|\, N(0) = n) = 1 - q^n,$$

where q^n and q are the probabilities that a population of size n and 1 go extinct, respectively. Plugging this into our equation above we obtain

$$\begin{aligned} P(N(t) = n \,|\, N(\infty) \neq 0, N(0) = n_0) &= P(N(t) = n \,|\, N(0) = n_0) \frac{1 - q^n}{1 - q^{n_0}} \\ &= \frac{1}{1 - q^{n_0}} \left(P(n, n_0, t) - P(n, n_0, t) q^n \right) \end{aligned}$$

and therefore, the generating function $\tilde{G}$ of the process that is conditioned on not going extinct is

$$\begin{aligned} \tilde{G}(n_0, t, z) &= \sum_n \left(\frac{1}{1 - q^{n_0}} \left(P(n, n_0, t) - P(n, n_0, t) q^n \right) \right) z^n \\ &= \frac{1}{1 - q^{n_0}} \sum_n \left(P(n, n_0, t) z^n - P(n, n_0, t) (qz)^n \right) \\ &= \frac{G(n_0, t, z) - G(n_0, t, qz)}{1 - q^{n_0}}. \end{aligned}$$

Hence, we can directly obtain $\tilde{G}$ once we know G, the generating function of the unconditioned process. Finally, all we need to do to find the conditional probability distribution $\tilde{P}$ is to replace G with $\tilde{G}$ in the integral equation (2.11) for P.

2.10 Mean behaviour and cumulative probabilities for the conditioned process

It is also possible to extend the calculations for the mean behaviour and cumulative probabilities to the process that is conditioned on non-extinction.

2.10.1 Mean behaviour

We are interested in the mean behaviour of the birth-death process that is conditioned on not going extinct. A corresponding ODE is not obvious, so we let $\tilde{N}$ be the random variable that counts the number of individuals in the birth-death pro-

cess conditioned on non-extinction. As before, we first use the analytical solution of the probability generating function to obtain

$$\begin{aligned} E[\tilde{N}] &= \left.\frac{\partial \tilde{G}(n_0,t,z)}{\partial z}\right|_{z=1} = \frac{1}{1-q^{n_0}} \left.\frac{\partial (G(n_0,t,z) - G(n_0,t,qz))}{\partial z}\right|_{z=1} \\ &= \frac{1}{1-q^{n_0}} \left(E[N] - \left.\frac{\partial G(n_0,t,qz))}{\partial z}\right|_{z=1} \right) \\ &= \frac{1}{1-q^{n_0}} \left(E[N] - n_0 q^{n_0} e^{-(b-d)t} \right) \\ &= \frac{n_0 (e^{(b-d)t} - q^{n_0} e^{-(b-d)t})}{1-q^{n_0}}. \end{aligned}$$

Note that in the limit $q \to 0$, where non-extinction of the unconditioned process becomes guaranteed, the means of the conditioned and unconditioned processes coincide, $E[\tilde{N}] \to E[N]$ ($q \to 0$). Further note that

$$E[\tilde{N}] \to \frac{E[N]}{1-q^{n_0}} \qquad (t \to \infty),$$

that is, the growth rates of the unconditioned and conditioned process eventually converge. In Figure 2.5 we see that in cases where extinction is very likely a naive extrapolation of the growth curve can lead to large error in the extrapolation of the time of exposure. This is because those paths that do not go extinct typically were able to initially grow much faster than average. Therefore, these paths do not need as much time to reach the detection level as one might have estimated from an unconstrained model fit. In short: this *infection bias* leads to an earlier estimate for the time of exposure, and hence to a shorter estimate of the length of the eclipse phase.

As with the unconditioned process it is valuable to re-derive a ODE for $E[\tilde{N}]$ that does not use the analytical solution of the probability generating function, so that the method can be extended to processes where an analytical solution is not available. Starting with the equation for $E[\tilde{N}]$ above we see that we need to calculate $f(t) = \left.\frac{\partial G(n_0,t,qz)}{\partial z}\right|_{z=1}$. The calculation is simplified if we first substitute x

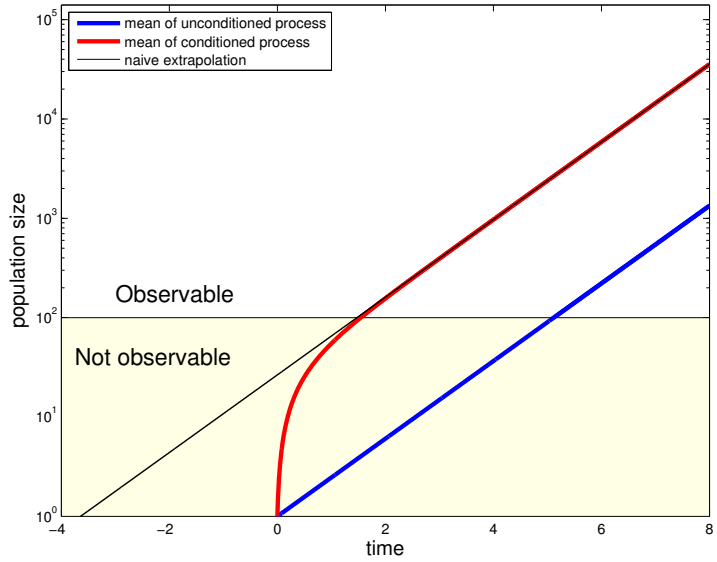


Figure 2.5: Mean of the unconditional $E[N]$ (in blue) and conditional $E[\tilde{N}]$ (in red) process, in the case where extinction is very likely, $q \approx 96\%$. We see a dramatic difference between the means, which can lead to drastically wrong estimates of the initial time of the process (e.g., time of HIV exposure) when a naive fit (black line) is extrapolated to $E[\tilde{N}] = 1$. Parameters: $b = 23.9$ day^{-1}, $d = 23$ day^{-1}, $n_0 = 1$.

$= qz$ to obtain

$$f(t) = \left.\frac{\partial G(n_0,t,qz)}{\partial z}\right|_{z=1} = q \left.\frac{\partial G(n_0,t,x)}{\partial x}\right|_{x=q}.$$

Next, we take the time derivative on both sides and use the PDE formulation (2.6).

$$\begin{aligned}\frac{d}{dt}f(t) &= q\frac{d}{dt}\frac{\partial}{\partial x}G(n_0,t,x)\Big|_{x=q} \\ &= q\,\frac{\partial}{\partial x}\frac{d}{dt}G(n_0,t,x)\Big|_{x=q} \\ &= q\,\frac{\partial}{\partial x}\left(\left(bx^2+d-(b+d)x\right)\frac{\partial}{\partial x}G(n_0,t,x)\right)\Big|_{x=q}\end{aligned}$$

$$= q\,(2bq-(b+d))\frac{\partial}{\partial x}G(n_0,t,x)\Big|_{x=q} + q\,(bq^2+d-(b+d)q)\frac{\partial^2}{\partial x^2}G(n_0,t,x)\Big|_{x=q}.$$

Note that in the second summand we see the same term as on the right hand side of the PDE (2.6), with $z=q$. Since the extinction probability q is such that the left hand side $\frac{\partial}{\partial t}G(n_0,t,q)=0$. Therefore the second summand evaluates to zero. Plugging in $q=d/b$ we obtain

$$\frac{d}{dt}f(t) = -(b-d)f(t), \qquad f(0) = \frac{\partial}{\partial z}G(n_0,0,qz)\Big|_{z=1} = \frac{\partial}{\partial z}(qz)^{n_0}\Big|_{z=1} = n_0 q^{n_0}.$$

The solution to this differential equation is $f(t)=n_0q^{n_0}e^{-(b-d)t}$, and matches the result obtained from the analytical solution above.

Note that we could also use Cauchy's integral formula to evaluate the derivate $f(t)=\frac{\partial G(n_0,t,qz)}{\partial z}\Big|_{z=1}$. Since extinction is not guaranteed we have $q<1$ and thus qz is in the interior of the unit circle where $G(n_0,t,\cdot)$ is analytic. Hence, C can again be the path along the boundary of the unit circle and we could calculate the partial derivative of $G(n_0,t,qz)$ with respect to z at $z=1$ as

$$\begin{aligned}\frac{\partial G(n_0,t,qz)}{\partial z}\Big|_{z=1} &= \frac{\partial G(n_0,t,qz)}{\partial (qz)}\frac{\partial (qz)}{\partial z}\Big|_{z=1}\\ &= q\frac{1}{2\pi i}\int_C \frac{G(n_0,t,w)}{(w-q)^2}dw\\ &= \frac{q}{2\pi}\int_0^{2\pi}\frac{G(n_0,t,e^{i\varphi})e^{i\varphi}}{(e^{i\varphi}-q)^2}d\varphi.\end{aligned}$$

2.10.2 Cumulative probabilities

Finally, we might also be interested in calculating

$$P(N(t)<\mathrm{dL}\,|\,N(\infty)\neq 0, N(0)=n_0),$$

the probability that the population size is below the threshold dL, conditioned on non-extinction of the population. This is straightforward when we use the proba-

bility generating function of the conditioned process:

$$\begin{aligned}
&P(N(t) < \mathrm{dL} \,|\, N(\infty) \neq 0, N(0) = n_0) \\
&\quad = \sum_{n=0}^{\mathrm{dL}-1} P(N(t) = n \,|\, N(\infty) \neq 0\,, N(0) = n_0) \\
&\quad = \sum_{n=0}^{\mathrm{dL}-1} \frac{1}{\pi} \mathrm{Re}\left(\int_0^{\pi} \tilde{G}(n_0, t, e^{i\varphi}) e^{-in\varphi} d\varphi \right) \\
&\quad = \frac{1}{\pi} \mathrm{Re}\left(\int_0^{\pi} \frac{G(n_0, t, e^{i\varphi}) - G(n_0, t, qe^{i\varphi})}{1 - q^{n_0}} \frac{1 - e^{-i\mathrm{dL}\varphi}}{1 - e^{-i\varphi}} d\varphi \right).
\end{aligned}$$

2.11 The T^*-V model

The techniques developed in the sections above are not exclusive to the birth-death process. Indeed, they can be extended to processes with more than one type, so-called *multi-type* branching processes. The steps required to obtain the corresponding results are very similar to the above, so instead of showing every detail we highlight how the techniques are generalized. The model of this section is targeted towards HIV infection and will as such consider HIV particles V and T-lymphocytes as target cells T. HIV infects target cells to turn them into virus-producing, productively infected cells, which we denote with a superscript star.

The T^*-V model is one of the simplest multi-type branching processes that describes the basic mechanism of HIV infection: Virions infect target cells at rate kT and are cleared at rate c, while infected target cells die at death rate δ and produce virions with production rate p. The model can be summarized in ODE form as

$$\begin{aligned}
\dot{T}^* &= kTV - \delta T^*, \\
\dot{V} &= pT^* - (c + kT)V,
\end{aligned} \tag{2.15}$$

with $T^*(0) = n_0$ and $V(0) = v_0$. We are particularly interested in models of early HIV infection, where the virus population size is relatively small and only a few target cells are infected. Because of this our approximation of a constant pool of uninfected target cells T is valid, and, unlike other classical HIV models, we do not have to add a separate equation for T. This assumption implies that (2.15) is linear,

and so the underlying stochastic process is a branching process. Another contrast to classical HIV models is the additional $-kTV$ term to explicitly account for the one virion lost when a new cell is infected. This is done because even a single virion matters when the total number of virions is small.

For the branching process, let n be the number of infected cells, v be the total number of virions, and n_0, v_0 be the initial conditions, respectively. We want to find the probability

$$P(n,v,n_0,v_0,t) = \mathbb{P}((T^*,V)(t) = (n,v) \,|\, (T^*,V)(0) = (n_0,v_0)).$$

2.11.1 The master equations

As before we make the assumption that the process is stationary, and that the probabilities of all the possible events in a time span of length h are, to first order, proportional to their rate and the duration of h:

$$\begin{aligned}
P(n,v,n+1,v,h) &= \delta(n+1)h + o(h),\\
P(n,v,n,v-1,h) &= pnh + o(h),\\
P(n,v,n-1,v+1,h) &= kT(v+1)h + o(h),\\
P(n,v,n,v+1,h) &= c(v+1)h + o(h),
\end{aligned}$$

while the probability of more than one event is taken to be $o(h)$.

From here we use the Chapman-Kolmogorov equations similarly to the birth-death model to calculate

$$\begin{aligned}
P(n,v,n_0,v_0,t+h) &= P(n+1,v,n_0,v_0,t)P(n,v,n+1,v,h)\\
&\quad + P(n,v-1,n_0,v_0,t)P(n,v,n,v-1,h)\\
&\quad + P(n-1,v+1,n_0,v_0,t)P(n,v,n-1,v+1,h)\\
&\quad + P(n,v+1,n_0,v_0,t)P(v,n,n,v+1,h)\\
&\quad + P(n,v,n_0,v_0,t)P(n,v,n,v,h) + o(h)\\
&= \delta(n+1)hP(n+1,v,n_0,v_0,t) + pnhP(n,v-1,n_0,v_0,t)\\
&\quad + kT(v+1)hP(n-1,v+1,n,v,t)
\end{aligned}$$

$$
\begin{aligned}
&+c(v+1)hP(n,v+1,n_0,v_0,t)\\
&+(1-bh)P(n,v,n_0,v_0,t)+o(h),
\end{aligned}
$$

which leads to the forward master equation:

$$
\begin{aligned}
\frac{d}{dt}P(n,v,n_0,v_0,t) &= \delta((n+1)P(n+1,v,n_0,v_0,t)-nP(n,v,n_0,v_0,t))\\
&\quad +p(nP(n,v-1,n_0,v_0,t)-nP(n,v,n_0,v_0,t))\\
&\quad +kT((v+1)P(n-1,v+1,n_0,v_0,t)-vP(n,v,n_0,v_0,t))\\
&\quad +c((v+1)P(n,v+1,n_0,v_0,t)-vP(n,v,n_0,v_0,t)),\\
P(n,v,n_0,v_0,0) &= \mathbb{1}_{n,n_0}\mathbb{1}_{v,v_0}.
\end{aligned}
\tag{2.16}
$$

Using the opposite time-step in the Chapman-Kolmogorov equations leads us to

$$
\begin{aligned}
P(n,v,n_0,v_0,t+h) &= P(n_0-1,v_0,n_0,v_0,h)P(n,v,n_0-1,v_0,t)\\
&\quad +P(n_0,v_0+1,n_0,v_0,h)P(n,v,n_0,v_0+1,t)\\
&\quad +P(n_0+1,v_0-1,n_0,v_0,h)P(n,v,n_0+1,v_0-1,t)\\
&\quad +P(n_0,v_0-1,n_0,v_0,h)P(n,v,n_0,v_0-1,t)\\
&\quad +P(n_0,v_0,n_0,v_0,h)P(n,v,n_0,v_0,t)+o(h)\\
&= \delta n_0hP(n,v,n_0-1,v_0,t)+pn_0hP(n,v,n_0,v_0+1,t)\\
&\quad +kTv_0hP(n,v,n_0+1,v_0-1,t)\\
&\quad +cv_0hP(n,v,n_0,v_0,t)+(1-bh)P(n,v,n_0,v_0,t)+o(h),
\end{aligned}
$$

from which we obtain the backward master equation as

$$
\begin{aligned}
\frac{d}{dt}P(n,v,n_0,v_0,t) &= \delta n_0(P(n,v,n_0-1,v_0,t)-P(n,v,n_0,v_0,t))\\
&\quad +pn_0(P(n,v,n_0,v_0+1,t)-P(n,v,n_0,v_0,t))\\
&\quad +kTv_0(P(n,v,n_0+1,v_0,t)-P(n,v,n_0,v_0,t))\\
&\quad +cv_0(P(n,v,n_0,v_0-1,t)-P(n,v,n_0,v_0,t)),\\
P(n,v,n_0,v_0,0) &= \mathbb{1}_{n,n_0}\mathbb{1}_{v,v_0}.
\end{aligned}
\tag{2.17}
$$

2.11.2 The probability generating function

The probability generating function is now a function of two complex variables instead of one:

$$G : \mathbb{N}_0^2 \times [0,\infty) \times \overline{\mathbb{D}}^2 \to \overline{\mathbb{D}}, \qquad (n_0, v_0, t, z_1, z_2) \mapsto \sum_{n=0}^{\infty} \sum_{v=0}^{\infty} P(n, v, n_0, v_0, t) z_1^n z_2^v. \tag{2.18}$$

As before, all the information about the branching process is obtainable from the probability generating function.

1. To calculate the marginal probability distributions, let $T_n^*(t)$ be the probability of having n infected T-cells at time t, i.e. $T_n^*(t) := \sum_{v=0}^{\infty} P(n, v, n_0, v_0, t)$. Then we obtain

$$T_n^*(t) = \frac{1}{n!} \frac{\partial^n G(n_0, v_0, t, z_1, 1)}{\partial z_1^n}\bigg|_{z_1=0}.$$

   Similarly, letting $V_v(t)$ be the probability of having v virus particles at time t, i.e. $V_v(t) := \sum_{n=0}^{\infty} P(n, v, n_0, v_0, t)$, we obtain that

$$V_v(t) = \frac{1}{v!} \frac{\partial^v G(n_0, v_0, t, 1, z_2)}{\partial z_2^v}\bigg|_{z_2=0}.$$

2. In particular, the probability that all infected cells are extinct at time t is given by $\sum_{v=0}^{\infty} P(0, v, n_0, v_0, t) = T_0^*(t) = G(n_0, v_0, t, 0, 1)$, and the probability of the virus to be extinct at time t is given by $\sum_{n=0}^{\infty} P(n, 0, n_0, v_0, t) = V_0(t)$ $= G(n_0, v_0, t, 1, 0)$.

3. The mean value of the number of infected cell and virus particles is calculated as

$$E[T^*] = \sum_{n=0}^{\infty} n T_n^*(t) = \frac{\partial G(n_0, v_0, t, z_1, 1)}{\partial z_1}\bigg|_{z_1=1},$$
$$E[V] = \sum_{v=0}^{\infty} v V_v(t) = \frac{\partial G(n_0, v_0, t, 1, z_2)}{\partial z_2}\bigg|_{z_2=1},$$

respectively.

4. The variance and all higher moments are calculated analogously to a single-type branching process.

Next, we derive the PDE and ODE formulation of the probability generating function. As before, we start with the forward master equation (2.16), multiply both sides by $z_1^n z_2^v$ and sum over all n and v to find

$$\begin{aligned}\sum_{n=0}^{\infty}\sum_{v=0}^{\infty}\frac{d}{dt}P(n,v,n_0,v_0,t)z_1^n z_2^v = \sum_{n=0}^{\infty}\sum_{v=0}^{\infty} z_1^n z_2^v \Big[& \delta\big((n+1)P(n+1,v,n_0,v_0,t) \\ & - nP(n,v,n_0,v_0,t)\big) \\ & + p\left(nP(n,v-1,n_0,v_0,t) - np(n,v,n_0,v_0,t)\right) \\ & + kT\big((v+1)P(n-1,v+1,n_0,v_0,t) \\ & - vP(n,v,n_0,v_0,t)\big) \\ & + c\big((v+1)P(n,v+1,n_0,v_0,t) \\ & - vP(n,v,n_0,v_0,t)\big)\Big].\end{aligned}$$

We now rewrite this term-by-term, making use of the fact that the probability vanishes whenever one of the species has a negative number of individuals.

1. $\sum_{n=0}^{\infty}\sum_{v=0}^{\infty}\frac{d}{dt}P(n,v,n_0,v_0,t)z_1^n z_2^v = \partial G/\partial t.$
2. $\delta\sum_{n=0}^{\infty}\sum_{v=0}^{\infty}(n+1)P(n+1,v,n_0,v_0,t)z_1^n z_2^v$
   $= \delta\sum_{n=0}^{\infty}\sum_{v=0}^{\infty} nP(n,v,n_0,v_0,t)z_1^{n-1}z_2^v = \delta(\partial G/\partial z_1).$
3. $\delta\sum_{n=0}^{\infty}\sum_{v=0}^{\infty} nP(n,v,n_0,v_0,t)z_1^n z_2^v = \delta z_1(\partial G/\partial z_1).$
4. $p\sum_{n=0}^{\infty}\sum_{v=0}^{\infty} nP(n,v-1,n_0,v_0,t)z_1^n z_2^v = pz_2\sum_{n=0}^{\infty}\sum_{v=0}^{\infty} nP(n,v,n_0,v_0,t)z_1^n z_2^v$
   $= pz_1z_2(\partial G/\partial z_1).$
5. $p\sum_{n=0}^{\infty}\sum_{v=0}^{\infty} nP(n,v,n_0,v_0,t)z_1^n z_2^v = pz_1(\partial G/\partial z_1).$
6. $kT\sum_{n=0}^{\infty}\sum_{v=0}^{\infty}(v+1)P(n-1,v+1,n_0,v_0,t)z_1^n z_2^v$
   $= kT\sum_{n=0}^{\infty}\sum_{v=0}^{\infty} vP(n,v,n_0,v_0,t)z_1^{n+1}z_2^{v-1} = kT\lambda z_1(\partial G/\partial z_2).$
7. $kT\sum_{n=0}^{\infty}\sum_{v=0}^{\infty} vP(n,v,n_0,v_0,t)z_1^n z_2^v = kTz_2(\partial G/\partial z_2).$

8. $c\sum_{n=0}^{\infty}\sum_{v=0}^{\infty}(v+1)P(n,v+1,n_0,v_0,t)z_1^n z_2^v$
$= c\sum_{n=0}^{\infty}\sum_{v=0}^{\infty} vP(n,v,n_0,v_0,t)z_1^n z_2^{v-1} = c(\partial G/\partial z_2)$.

9. $c\sum_{n=0}^{\infty}\sum_{v=0}^{\infty} vP(n,v,n_0,v_0,t)z_1^n z_2^v = cz_2(\partial G/\partial z_2)$.

We thus arrive at the following PDE for the probability generating function of the T^*-V model:

$$\begin{aligned}\frac{\partial G(n_0,v_0,t,z_1,z_2)}{\partial t} &= [\delta(1-z_1)+pz_1(z_2-1)]\frac{\partial G(n_0,v_0,t,z_1,z_2)}{\partial z_1}\\ &\quad + [kT(z_1-z_2)+c(1-z_2)]\frac{\partial G(n_0,v_0,t,z_1,z_2)}{\partial z_2},\\ G(n_0,v_0,0,z_1,z_2) &= z_1^{n_0}z_2^{v_0}.\end{aligned} \tag{2.19}$$

Multiplying both sides of the backward master equation (2.17) with $z_1^n z_2^v$ and summing over all n and v we find

$$\begin{aligned}\frac{\partial G(n_0,v_0,t,z_1,z_2)}{\partial t} &= \delta n_0\big(G(n_0-1,v_0,t,z_1,z_2)-G(n_0,v_0,t,z_1,z_2)\big)\\ &\quad + pn_0\big(G(n_0,v_0+1,t,z_1,z_2)-G(n_0,v_0,t,z_1,z_2)\big)\\ &\quad + kTv_0\big(G(n_0+1,v_0-1,t,z_1,z_2)-G(n_0,v_0,t,z_1,z_2)\big)\\ &\quad + cv_0\big(G(n_0,v_0-1,t,z_1,z_2)-G(n_0,v_0,t,z_1,z_2)\big),\\ G(n_0,v_0,0,z_1,z_2) &= z_1^{n_0}z_2^{v_0}.\end{aligned} \tag{2.20}$$

The branching property in the case of this multi-type process is

$$G(n_0,v_0,t,z_1,z_2) = \left(G(1,0,t,z_1,z_2)^{n_0}\right)\left(G(0,1,t,z_1,z_2)^{v_0}\right).$$

We will abbreviate $G_1(t) := G(1,0,t,z_1,z_2)$ and $G_2(t) := G(0,1,t,z_1,z_2)$. Using the branching property we can reduce the system of infinitely many ODES (2.20)

to the following two-dimensional system of ODES

$$\begin{aligned}
\frac{\partial G_1(t)}{\partial t} &= \delta(1-G_1(t)) + p(G_1(t)G_2(t) - G_1(t)),\\
\frac{\partial G_2(t)}{\partial t} &= kT(G_1(t) - G_2(t)) + c(1-G_2(t)),\\
G_1(0) &= z_1,\\
G_2(0) &= z_2.
\end{aligned} \tag{2.21}$$

To obtain the marginal probability distributions, e.g., of the number of virus particles $V_v(t)$ we use the very same steps as in the single-type process, namely Cauchy's integral formula with simple positively oriented closed circle with radius one around the origin, and the substitution $\alpha = 2\pi - \theta$, to find

$$V_v(t) = \frac{1}{\pi}\,\mathrm{Re}\left(\int_0^{\pi} G(n_0, v_0, t, 1, e^{i\theta})e^{-iv\theta}\,d\theta\right). \tag{2.22}$$

2.11.3 Calculating joint probability

So far we have restricted ourselves to marginal probabilities. This is not required, we can also calculate joint probabilities. Using the definition of the probability generating function,

$$\begin{aligned}
P(n,v,n_0,v_0,t) &= \frac{1}{v!}\,\frac{\partial^v}{\partial z_2^v}\left(\frac{1}{n!}\,\frac{\partial^n}{\partial z_1^n} G(n_0,v_0,t,z_1,z_2)\Big|_{z_1=0}\right)\Bigg|_{z_2=0}\\
&= \frac{1}{v!}\,\frac{\partial^v}{\partial z_2^v}\left(\frac{1}{2\pi i}\int_C \frac{G(n_0,v_0,t,z_1,z_2)}{z_1^{n+1}}dz_1\right)\Bigg|_{z_2=0}\\
&= \frac{1}{2\pi v!}\,\frac{\partial^v}{\partial z_2^v}\left(\int_0^{2\pi} G(n_0,v_0,t,e^{i\varphi},z_2)e^{-in\varphi}d\varphi\right)\Bigg|_{z_2=0}\\
&= \frac{1}{(2\pi)^2 i}\int_C \frac{\int_0^{2\pi} G(n_0,v_0,t,e^{i\varphi},z_2)e^{-in\varphi}d\varphi}{z_2^{v+1}}dz_2\\
&= \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi} G(n_0,v_0,t,e^{i\varphi},e^{i\psi})e^{-in\varphi}e^{-iv\psi}d\varphi d\psi.
\end{aligned}$$

We can simplify this to reduce the interval we need to integrate over. The price is that there will be two slightly different integrands. To begin with, write

$$\begin{aligned}P(n,v,n_0,v_0,t) &= \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi} G(n_0,v_0,t,e^{i\varphi},e^{i\psi})e^{-in\varphi}e^{-iv\psi}d\varphi d\psi \\ &= \frac{1}{(2\pi)^2}\Big(\underbrace{\int_0^{\pi}\int_0^{\pi} f d\varphi d\psi}_{=A} + \underbrace{\int_{\pi}^{2\pi}\int_0^{\pi} f d\varphi d\psi}_{=B} \\ &\qquad + \underbrace{\int_0^{\pi}\int_{\pi}^{2\pi} f d\varphi d\psi}_{=C} + \underbrace{\int_{\pi}^{2\pi}\int_{\pi}^{2\pi} f d\varphi d\psi}_{=D}\Big).\end{aligned}$$

Substituting we find that $D = \overline{A}$ and $C = \overline{B}$, the complex conjugate. Indeed, starting with D we substitute $\varphi' = 2\pi - \varphi$ and $\psi' = 2\pi - \psi$. Then

$$\begin{aligned}D &= \int_{\pi}^{0}\int_{\pi}^{0} G(n_0,v_0,t,e^{-i\varphi'},e^{-i\psi'})e^{in\varphi'}e^{iv\psi'}(-d\varphi')(-d\psi') \\ &= \int_0^{\pi}\int_0^{\pi} \overline{G(n_0,v_0,t,e^{i\varphi'},e^{i\psi'})e^{-in\varphi'}e^{-iv\psi'}}\, d\varphi' d\psi' \\ &= \overline{A},\end{aligned}$$

where we use the fact that $P(n,v,n_0,v_0,t)$ is real and hence

$$\begin{aligned}G(n_0,v_0,t,z_1,z_2) &= \sum_n\sum_v P(n,v,n_0,v_0,t)z_1^n z_2^v \\ &= \overline{\sum_n\sum_v P(n,v,n_0,v_0,t)\overline{z_1^n z_2^v}} \\ &= \overline{G(n_0,v_0,t,\overline{z_1},\overline{z_2})}.\end{aligned}$$

Accordingly one can show that $C = \overline{B}$, since

$$\begin{aligned}G(n_0,v_0,t,z_1,\overline{z_2}) &= \sum_n\sum_v P(n,v,n_0,v_0,t)z_1^n\overline{z_2^v} \\ &= \overline{\sum_n\sum_v P(n,v,n_0,v_0,t)\overline{z_1^n}z_2^v} \\ &= \overline{G(n_0,v_0,t,\overline{z_1},z_2)}.\end{aligned}$$

Consolidating these results we find that

$$P(n,v,n_0,v_0,t) = \frac{1}{(2\pi)^2}(A+\overline{A}+B+\overline{B}) = \frac{1}{2(\pi)^2}\,\mathrm{Re}(A+B).$$

Finally, substituting $\phi' = 2\pi - \psi$ in B we obtain that the joint probability is given by

$$\begin{aligned} P(n,v,n_0,v_0,t) = \frac{1}{2\pi^2}\,\mathrm{Re}\Big[\int_0^\pi \int_0^\pi \Big\{ & G(n_0,v_0,t,e^{i\varphi},e^{i\psi})e^{-in\varphi}e^{-iv\psi} \\ & + G(n_0,v_0,t,e^{i\varphi},e^{-i\psi})e^{-in\varphi}e^{iv\psi}d\varphi d\psi \Big\}\Big]. \end{aligned} \tag{2.23}$$

2.11.4 Mean behaviour of the stochastic process

In order to calculate the mean number of virions we recall

$$\begin{aligned} E[V] &= \sum_v v\mathbb{P}(V=v) = \sum_v v\left(\sum_n P(n,v,n_0,v_0,t)\right) \\ &= \left.\frac{\partial G(n_0,v_0,t,z_1,z_2)}{\partial z_2}\right|_{(z_1,z_2)=(1,1)}. \end{aligned}$$

We will derive an ODE for $\frac{\partial G(n_0,v_0,t,z_1,z_2)}{\partial z_2}$. For this, we take a time-derivative of the expression, use the PDE formulation (2.19) and then swap the partial derivatives, to find

$$\begin{aligned} \frac{\partial}{\partial t}\frac{\partial G(n_0,v_0,t,z_1,z_2)}{\partial z_2} &= \frac{\partial}{\partial z_2}\frac{\partial G(n_0,v_0,t,z_1,z_2)}{\partial t} \\ &= \frac{\partial}{\partial z_2}\Big[(\delta(1-z_1)+pz_1(z_2-1))\frac{\partial}{\partial z_1}G \\ &\qquad + (kT(z_1-z_2)+c(1-z_2))\frac{\partial}{\partial z_2}G\Big] \\ &= pz_1\frac{\partial G}{\partial z_1} + (\delta(1-z_1)+pz_1(z_2-1))\frac{\partial^2 G}{\partial z_2 z_1} \\ &\quad - (c+kT)\frac{\partial G}{\partial z_2} + (kT(z_1-z_2)+c(1-z_2))\frac{\partial^2 G}{\partial z_2^2}. \end{aligned}$$

Evaluating at $(z_1, z_2) = (1,1)$ we obtain

$$\frac{\partial}{\partial t}\left.\frac{\partial G(n_0,v_0,t,z_1,z_2)}{\partial z_2}\right|_{(z_1,z_2)=(1,1)} = p\left.\frac{\partial G(n_0,v_0,t,z_1,z_2)}{\partial z_1}\right|_{(z_1,z_2)=(1,1)} - (c+kT)\left.\frac{\partial G(n_0,v_0,t,z_1,z_2)}{\partial z_2}\right|_{(z_1,z_2)=(1,1)},$$

and similarly

$$\frac{\partial}{\partial t}\left.\frac{\partial G(n_0,v_0,t,z_1,z_2)}{\partial z_1}\right|_{(z_1,z_2)=(1,1)} = kT\left.\frac{\partial G(n_0,v_0,t,z_1,z_2)}{\partial z_1}\right|_{(z_1,z_2)=(1,1)} - \delta\left.\frac{\partial G(n_0,v_0,t,z_1,z_2)}{\partial z_2}\right|_{(z_1,z_2)=(1,1)}.$$

Hence, the time-dependent functions $f_1(t) = \left.\frac{\partial G(n_0,v_0,t,z_1,z_2)}{\partial z_1}\right|_{(z_1,z_2)=(1,1)}$ and $f_2(t) = \left.\frac{\partial G(n_0,v_0,t,z_1,z_2)}{\partial z_2}\right|_{(z_1,z_2)=(1,1)}$, which is what we want to calculate, indeed satisfy the same ODE (2.15) that we started with:

$$\begin{aligned} \dot{f}_1 &= kT f_2 - \delta f_1, \\ \dot{f}_2 &= p f_1 - (c+kT) f_2, \end{aligned} \tag{2.24}$$

with initial conditions $f_1(0) = \sum_n \sum_v nP(n,v,n_0,v_0,0) = n_0$ and $f_2(0) = v_0$. Solving this ODE gives $f_2(t) = E[V(t)]$.

2.11.5 Extinction probability, R_0 and r

We obtain the probability of extinction by finding the steady states of (2.21), which are

$$q_1 = 1, \quad q_2 = 1, \quad \text{or} \quad q_1 = \frac{\delta}{p}\frac{c+kT}{kT}, \quad q_2 = \frac{\delta}{p} + \frac{c}{c+kT}.$$

The interpretation is that q_1 is the probability that the process goes extinct if started with a single cell, while q_2 is the probability that the process goes extinct if started with a single virion.

The basic reproduction number R_0 for this model is the expected number of

new infected cells that are descendants of a single infected cell. Since virions infect target cells at rate kT and have an average life-span on $1/(c+kT)$, while target cells produce virions at rate p over their average life span $1/\delta$ we obtain

$$R_0 = \frac{pkT}{\delta(c+kT)}. \tag{2.25}$$

Alternatively, R_0 can be obtained by finding the criterion for the trivial steady state of the original linear ODE to be stable, which leads to the same result.

Further, we get the initial growth rate r as the largest real part of the eigenvalues of the matrix given by

$$\begin{bmatrix} \dot{T}^* \\ \dot{V} \end{bmatrix} = \begin{bmatrix} -\delta & kT \\ p & -c-kT \end{bmatrix} \begin{bmatrix} T^* \\ V \end{bmatrix}$$

with characteristic polynomial

$$\lambda^2 + (\delta + c + kT)\lambda + \delta(c+kT) - pkT.$$

For $R_0 > 1$ this polynomial has a negative and a positive root, hence the initial growth rate is given by

$$r = \frac{1}{2}\left[-\delta - c - kT + \sqrt{(\delta - c - kT)^2 + 4pkT}\right]. \tag{2.26}$$

2.11.6 Conditioning on non-extinction

We calculate the probability of observing a viral load of $V(t) = v$, conditioned on the process not going extinct. First, calculate the probability of the state $(T^*, V)(t)$ $= (n, v)$, then sum over all n later. We abbreviate T^*V_t for $(T^*, V)(t)$ and the initial conditions $T^*V_0 = (n_0, v_0)$ with IC. Using Bayes' rule

$$\begin{aligned} &P(T^*V_t = (n,v) \,|\, T^*V_\infty \neq \mathbf{0}, \text{IC}) \\ &\qquad = P(T^*V_t = (n,v) \,|\, \text{IC}) \frac{P(T^*V_\infty \neq \mathbf{0} \,|\, T^*V_t = (n,v), \text{IC})}{P(T^*V_\infty \neq \mathbf{0} \,|\, \text{IC})} \end{aligned}$$

$$= P(T^*V_t = (n,v)\,|\,\mathrm{IC})\frac{1-P(T^*V_\infty = \mathbf{0}\,|\,T^*V_t = (n,v))}{1-P(T^*V_\infty = \mathbf{0}\,|\,\mathrm{IC})}$$
$$= P(T^*V_t = (n,v)\,|\,\mathrm{IC})\frac{1-(q_1)^n(q_2)^v}{1-q_1^{n_0}q_2^{v_0}}.$$

Now, we multiply with $z_1^n z_2^v$ and sum over all n, v to obtain the probability generating function $\tilde{G}$ of the conditioned process,

$$\begin{aligned}\tilde{G}(n_0,v_0,t,z_1,z_2) &:= \sum_n\sum_v \left(P(T^*V_t = (n,v)\,|\,\mathrm{IC})\frac{1-(q_1)^n(q_2)^v}{1-q_1^{n_0}q_2^{v_0}}\right) z_1^n z_2^v \\ &= \frac{1}{1-q_1^{n_0}q_2^{v_0}}\Big(\sum_n\sum_v P(T^*V_t = (n,v)\,|\,\mathrm{IC}) z_1^n z_2^v \\ &\qquad\qquad - \sum_n\sum_v P(T^*V_t = (n,v)\,|\,\mathrm{IC})(q_1 z_1)^n (q_2 z_2)^v\Big) \\ &= \frac{G(n_0,v_0,t,z_1,z_2) - G(n_0,v_0,t,q_1 z_1, q_2 z_2)}{1-q_1^{n_0}q_2^{v_0}}.\end{aligned}$$

As in the non-conditioned process, setting $z_1 = 1$ gives the probability generating function of the marginal probability distribution for the number of virions, i.e.

$$\begin{aligned}\tilde{G}(n_0,v_0,t,1,z_2) &= \frac{G(n_0,v_0,t,1,z_2) - G(n_0,v_0,t,q_1,q_2 z_2)}{1-q_1^{n_0}q_2^{v_0}} \\ &= \sum_n\sum_v P(T^*V_t = (n,v)\,|\,T^*V_\infty \neq \mathbf{0},\mathrm{IC}) z_2^v,\end{aligned}$$

so that, as before,

$$\begin{aligned}&\sum_n P(T^*V_t = (n,v)\,|\,T^*V_\infty \neq \mathbf{0},\mathrm{IC}) \\ &\quad= \frac{1}{1-q_1^{n_0}q_2^{v_0}}\frac{1}{v!}\left.\frac{\partial^v\left(G(n_0,v_0,t,1,z_2) - G(n_0,v_0,t,q_1,q_2 z_2)\right)}{\partial z_2^v}\right|_{z_2=0} \\ &\quad= \frac{1}{1-q_1^{n_0}q_2^{v_0}}\frac{1}{v!}\left(\frac{v!}{2\pi i}\int_C \frac{G(n_0,v_0,t,1,z_2) - G(n_0,v_0,t,q_1,q_2 z_2)}{z_2^{v+1}} dz_2\right) \\ &\quad= \frac{1}{1-q_1^{n_0}q_2^{v_0}}\frac{1}{2\pi}\int_0^{2\pi}\left(G(n_0,v_0,t,1,e^{i\varphi}) - G(n_0,v_0,t,q_1,q_2 e^{i\varphi})\right)e^{-iv\varphi}d\varphi \\ &\quad= \frac{1}{2\pi}\int_0^{2\pi}\tilde{G}(n_0,v_0,t,1,e^{i\varphi})e^{-iv\varphi}\,d\varphi\end{aligned}$$

$$= \frac{1}{\pi} \operatorname{Re}\left(\int_0^{\pi} \tilde{G}(n_0, v_0, t, 1, e^{i\varphi}) e^{-iv\varphi}\, d\varphi \right).$$

We used the same steps as for the one-state model, and in particular the fact that $G(n_0, v_0, t, \overline{z_1}, \overline{z_2}) = \overline{G(n_0, v_0, t, z_1, z_2)}$.

2.11.7 Mean behaviour and cumulative probabilities for the conditioned T^*-V model

Now on to the mean of the conditioned process. Let $\tilde{V}$ be the random variable of the number of virus particles of the conditioned process.

$$\begin{aligned}
E[\tilde{V}(t)] &= \sum_v v \mathbb{P}(\tilde{V} = v) = \sum_v v \left(\sum_n P(T^*V_t = (n,v) \,|\, T^*V_\infty \neq \mathbf{0}, \mathrm{IC}) \right) \\
&= \sum_n \sum_v P(T^*V_t = (n,v) \,|\, T^*V_\infty \neq \mathbf{0}, \mathrm{IC}) 1^n v 1^{v-1} \\
&= \left. \frac{\partial \tilde{G}(n_0, v_0, t, z_1, z_2)}{\partial z_2} \right|_{(z_1,z_2)=(1,1)} \\
&= \frac{1}{1 - q_1^{n_0} q_2^{v_0}} \left[\left. \frac{\partial G(n_0, v_0, t, z_1, z_2)}{\partial z_2} \right|_{(z_1,z_2)=(1,1)} \right. \\
&\qquad \left. - \left. \frac{\partial G(n_0, v_0, t, q_1 z_1, q_2 z_2)}{\partial z_2} \right|_{(z_1,z_2)=(1,1)} \right].
\end{aligned}$$

We recognize the former term as the mean of the unconditioned process, which we can calculate by solving the ODE (2.15). For the second term we have two possibilities.

Integral formulation

We use $q_1 < 1$, $q_2 < 1$ so that the point we evaluate the probability generating function inside the unit circle, where G is analytic.

$$\begin{aligned}
\left. \frac{\partial G(n_0, v_0, t, q_1 z_1, q_2 z_2)}{\partial z_2} \right|_{(z_1,z_2)=(1,1)} &= \left. \frac{\partial G(n_0, v_0, t, q_1, q_2 z_2)}{\partial q_2 z_2} \frac{\partial (q_2 z_2)}{\partial z_2} \right|_{q_2 z_2 = q_2} \\
&= q_2 \frac{1}{2\pi i} \int_C \frac{G(n_0, v_0, t, q_1, w)}{(w - q_2)^2}\, dw
\end{aligned}$$

$$= \frac{q_2}{2\pi} \int_0^{2\pi} G(n_0, v_0, t, q_1, e^{i\varphi}) \frac{e^{i\varphi}}{(e^{i\varphi} - q_2)^2} \, d\varphi.$$

ODE formulation

As in the unconditioned process, we can derive an ODE formulation for $g_1(t) = \frac{\partial G(n_0,v_0,t,q_1z_1,q_2z_2)}{\partial z_1}\Big|_{(z_1,z_2)=(1,1)}$ and $g_2(t) = \frac{\partial G(n_0,v_0,t,q_1z_1,q_2z_2)}{\partial z_2}\Big|_{(z_1,z_2)=(1,1)}$ by using the forward PDE (2.19) for $G(n_0, v_0, t, x_1, x_2)$. Note that

$$\begin{aligned}
g_1(t) &= \frac{\partial G(n_0, v_0, t, q_1 z_1, q_2 z_2)}{\partial z_1}\bigg|_{(z_1,z_2)=(1,1)} \\
&= \frac{\partial (q_1 z_1)}{\partial z_1} \frac{\partial G(n_0, v_0, t, q_1 z_1, q_2 z_2)}{\partial (q_1 z_1)}\bigg|_{(q_1 z_1, q_2 z_2)=(q_1,q_2)} \\
&= q_1 \, \frac{\partial G(n_0, v_0, t, x_1, x_2)}{\partial x_1}\bigg|_{(x_1,x_2)=(q_1,q_2)} \\
g_2(t) &= q_2 \, \frac{\partial G(n_0, v_0, t, x_1, x_2)}{\partial x_2}\bigg|_{(x_1,x_2)=(q_1,q_2)}.
\end{aligned}$$

Taking the time derivative of the above yields

$$\begin{aligned}
\dot{g}_1(t) &= q_1 \frac{\partial}{\partial t} \, \frac{\partial G(n_0, v_0, t, x_1, x_2)}{\partial x_1}\bigg|_{(x_1,x_2)=(q_1,q_2)} \\
&= q_1 \frac{\partial}{\partial x_1} \Bigg[(\delta(1-x_1) + p x_1 (x_2 - 1)) \frac{\partial G(n_0, v_0, t, x_1, x_2)}{\partial x_1} \\
&\qquad + (kT(x_1 - x_2) + c(1 - x_2)) \frac{\partial G(n_0, v_0, t, x_1, x_2)}{\partial x_2} \Bigg] \Bigg|_{(x_1,x_2)=(q_1,q_2)} \\
&= q_1 \Bigg[(-\delta + p(x_2 - 1)) \frac{\partial G(n_0, v_0, t, x_1, x_2)}{\partial x_1} + 0 \\
&\qquad + kT \frac{\partial G(n_0, v_0, t, x_1, x_2)}{\partial x_2} + 0 \Bigg] \Bigg|_{(x_1,x_2)=(q_1,q_2)} \\
&= (p(q_2 - 1) - \delta) g_1(t) + \frac{kT q_1}{q_2} g_2(t) \\
&= kT \frac{q_1}{q_2} g_2(t) - (\delta - p(q_2 - 1)) g_1(t)
\end{aligned}$$

$$
\begin{aligned}
\dot{g}_2(t) &= q_2 \frac{\partial}{\partial t} \left. \frac{\partial G(n_0, v_0, t, x_1, x_2)}{\partial x_2} \right|_{(x_1,x_2)=(q_1,q_2)} \\
&= q_2 \frac{\partial}{\partial x_2} \left[(\delta(1-x_1) + p x_1 (x_2 - 1)) \frac{\partial G(n_0, v_0, t, x_1, x_2)}{\partial x_1} \right. \\
&\qquad \left. + (kT(x_1 - x_2) + c(1-x_2)) \frac{\partial G(n_0, v_0, t, x_1, x_2)}{\partial x_2} \right] \Bigg|_{(x_1,x_2)=(q_1,q_2)} \\
&= q_2 \left[p x_1 \frac{\partial G(n_0, v_0, t, x_1, x_2)}{\partial x_1} + 0 \right. \\
&\qquad \left. - (kT + c) \frac{\partial G(n_0, v_0, t, x_1, x_2)}{\partial x_2} + 0 \right] \Bigg|_{(x_1,x_2)=(q_1,q_2)} \\
&= p q_2 g_1(t) - (c + kT) g_2(t).
\end{aligned}
$$

The second-order partial derivatives vanish because the factor multiplying them simplifies to zero because q_1 and q_2 are such that $\frac{\partial}{\partial t} G(n_0, v_0, t, q_1, q_2) = 0$. Next, let's calculate the initial condition.

$$
\begin{aligned}
g_1(0) &= q_1 \left. \frac{\partial G(n_0, v_0, 0, x_1, x_2)}{\partial x_1} \right|_{(x_1,x_2)=(q_1,q_2)} \\
&= q_1 \left. \frac{\partial}{\partial x_1} \sum_n \sum_v P(n, v, n_0, v_0, 0) x_1^n x_2^v \right|_{(x_1,x_2)=(q_1,q_2)} \\
&= q_1 \left. \frac{\partial}{\partial x_1} \sum_n \sum_v \mathbb{1}_{(n,v),(n_0,v_0)} x_1^n x_2^v \right|_{(x_1,x_2)=(q_1,q_2)} \\
&= q_1 \left. \frac{\partial}{\partial x_1} x_1^{n_0} x_2^{v_0} \right|_{(x_1,x_2)=(q_1,q_2)} \\
&= q_1 n_0 (q_1)^{n_0 - 1} q_2^{v_0} = n_0 q_1^{n_0} q_2^{v_0},
\end{aligned}
$$

and correspondingly,

$$
\begin{aligned}
g_2(0) &= q_2 \left. \frac{\partial G(n_0, v_0, 0, x_1, x_2)}{\partial x_2} \right|_{(x_1,x_2)=(q_1,q_2)} \\
&= q_2 \left. \frac{\partial}{\partial x_2} \sum_n \sum_v P(n, v, n_0, v_0, 0) x_1^n x_2^v \right|_{(x_1,x_2)=(q_1,q_2)}
\end{aligned}
$$

$$
\begin{aligned}
&= q_2 \left.\frac{\partial}{\partial x_2} \sum_n \sum_v \mathbb{1}_{(n,v),(n_0,v_0)} x_1^n x_2^v \right|_{(x_1,x_2)=(q_1,q_2)} \\
&= q_2 \left.\frac{\partial}{\partial x_2} x_1^{n_0} x_2^{v_0} \right|_{(x_1,x_2)=(q_1,q_2)} \\
&= q_2 v_0 q_1^{n_0} (q_2)^{v_0-1} = v_0 q_1^{n_0} q_2^{v_0}.
\end{aligned}
$$

Hence in summary, in order to get $E[\tilde{V}(t)]$ we have to first find $f_2(t)$ by solving (2.24), then we calculate $g_2(t)$ by solving the ODE

$$
\begin{aligned}
\dot{g}_1(t) &= kT\frac{q_1}{q_2} g_2(t) - (\delta - p(q_2-1)) g_1(t), \quad && g_1(0) = n_0 q_1^{n_0} q_2^{v_0}, \\
\dot{g}_2(t) &= p q_2 g_1(t) - (c+kT) g_2(t), && g_2(0) = v_0 q_1^{n_0} q_2^{v_0},
\end{aligned}
\tag{2.27}
$$

(note: $\delta - p(q_2-1) = \delta/q_1$) and plug the results in the equations

$$
E[\tilde{V}(t)] = \frac{f_2(t) - g_2(t)}{1 - q_1^{n_0} q_2^{v_0}}.
$$

Since this ODE is linear it is possible to derive an analytical expression for $E[\tilde{V}(t)]$. Finally, note that in the limit $q_1 \to 1$, $q_2 \to 1$ where extinction becomes a certainty, the ODES (2.15) and (2.27) coincide.

The equations for the cumulative probabilities are exactly the same as in the birth-death process, with the probability generating function corresponding to the conditioned or unconditioned T^*-V model, respectively.

2.12 The T^*-V-W model with infectious and non-infectious virions

The process by which infected cells produce new HIV virions is very error-prone. Indeed, only a fraction Q of virions produced are infectious, where Q is estimated to be on the order of $Q \approx 10^{-2} - 10^{-4}$ [63, 75, 96]. Non-infectious virions, denoted as W, are not able to infect new target cells and hence are harmless, and since a free virion has a very high clearance rate the non-infectious virus compartment is often ignored or modelled by adjusting the virus production rate p. However, the presence of non-infectious virions is recorded in HIV RNA tests, which are unable

to evaluate the functionality of the virus in the assay. The model can be written in ODE form as

$$\begin{aligned} \dot{T}^* &= kTV - \delta T^*, \\ \dot{V} &= pQT^* - (c+kT)V, \\ \dot{W} &= p(1-Q)T^* - cW, \end{aligned} \tag{2.28}$$

with $T^*(0) = n_0$, $V(0) = v_0$, and $W(0) = w_0$.

Due to the uncertainty about the parameters Q and p we decided to not include a non-infectious virus compartment in our models in Chapter 3. Nevertheless, the T^*-V-W model is worth investigating because it offers new challenges on how to calculate statistics about the total number of virions $V+W$,

$$P(n,v,w,n_0,v_0,w_0,t) = \mathbb{P}((T^*,V,W)(t) = (n,v,w) \,|\, (T^*,V,W)(0) = (n_0,v_0,w_0)).$$

Most of the other mathematical derivations for the T^*-V-W model are similar to the T-V model and all details and results are shown in Appendix A.

2.13 The T_1^*-T_2^*-V model

When an HIV virion enters a target cell several integration and replication steps need to happen before this newly infected cell starts to produce virus particles. This intra-cellular delay is featured in another variant of the classic virus dynamics model, the T_1^*-T_2^*-V model. Here T_1^*-cells do not yet actively produce new virions while T_2^*-cells do. Hence this model introduces a single new parameter, the length of the intra-cellular time delay $1/s$ and can be written in ODE form as

$$\begin{aligned} \dot{T}_1^* &= kTV - sT_1^*, \\ \dot{T}_2^* &= sT_1^* - \delta T_2^* \\ \dot{V} &= pT_2^* - (c+kT)V, \end{aligned}$$

with $T_1^*(0) = m_0$, $T_2^*(0) = n_0$, and $V(0) = v_0$. In this variant each cell that gets infected will also eventually become productive, which is a justified approximation because a cell's lifespan is much larger than the intra-cellular delay. The T_1^*-T_2^*-V model is our model of choice for calculating the length of the eclipse phase in

Chapter 3, because it includes key features of the timing of how the virus population grows or goes extinct, while keeping a small number of parameters. The mathematical derivations for this model are similar to the other models discussed in this chapter so far, and the results are shown in Chapter 3.

2.14 Conclusion and future work

In this chapter we studied single- and multiple-type continuous-time Markov chain branching processes and derived results for several HIV models. Gillespie's algorithm is a simple and widely-used stochastic simulation algorithm to analyze these type of processes. While being straightforward to implement, it is also computationally expensive, especially when analyzing rare events, very fast reaction times, and potentially large populations. Here we derive semi-analytical approaches as alternatives to these extensive computer simulations, and apply them to several HIV models. The mathematical background for our derivations have mostly been known to the statistic, stochastic and numerical community, and fragments can be found in many textbooks. However, we are the first to systematically consolidate this knowledge so that it can readily be applied to models of latent and early HIV infection, and more. We have also identified and briefly analyzed the *infection bias*. This is a critical effect that needs to be considered when estimating the exposure time, especially when infection is unlikely yet the only available data is from cases where infection did occur.

We believe that insights from our analysis of the infection bias could also be useful when estimating the time of the outbreak of epidemics: Only diseases that infect a certain number of people can be observed. If such a diseases had a large probability of going extinct early, but was in fact able to spread, our analysis may be susceptible to the infection bias. In particular, a naive back-extrapolation to estimate the time when the outbreak started from observed incidence would predict a date too far in the past if the infection bias is ignored. We are not aware of a study that has taken this effect in account.

Finally, we reiterate that the techniques presented here only work under the branching assumption that each lineage acts independently. While Gillespie's algorithm can still be used, the methods developed here will not be applicable oth-

erwise. Having said that, our techniques can be generalized in other ways, for example constant reaction rates can be replaced by time-dependent reaction rates, see [22].

Chapter 3

On the duration of the period between exposure to HIV and detectable infection

The content of this chapter and the corresponding Appendix B stem from a collaboration with the British Columbia Centre for Disease Control. I am the lead author of this joint work with Darlene Taylor, Jessica Conway, Gina Ogilvie and Daniel Coombs.

Abstract: HIV infection cannot be detected immediately after exposure because plasma viral loads are too small initially. The duration of this phase of infection (the "eclipse period") is difficult to estimate because precise dates of exposure are rarely known. Therefore, the reliability of clinical HIV testing during the first few weeks of infections is unknown, creating anxiety among HIV-exposed individuals and their physicians. We address this by fitting stochastic models of early HIV infection to detailed viral load records for 78 plasma donors, taken during the period of exposure and infection. We first show that the classic birth-death model does not satisfactorily describe early infection. We therefore apply a stochastic model that includes infected cells and virions separately. Since every plasma donor in our data eventually becomes infected, we must condition the model to reflect this bias, before fitting to the data. Applying our best estimates of unknown parameter

values, we estimate the mean eclipse period to be 8-10 days. We further estimate the reliability of a negative test t days after potential exposure. These results can be used to inform patients and provide a baseline for understanding the value and timing of initial and follow-up testing.

3.1 Motivation and background

Rapid and reliable diagnosis of HIV infection is an important factor in optimizing patient care after possible exposure to the virus. However, diagnosis during the first few weeks after exposure is challenging because initially the virus is present at only very low levels – the inoculumn size is difficult to estimate but certainly several orders of magnitude lower than the detection limit – and immune responses take time to develop to a detectable level [1, 8, 41, 61, 62, 71].

The length of time between exposure and reliable detection of infection (also called the *eclipse period* or *window period*) depends on the virus population dynamics and the laboratory test that is applied. Clinically available HIV testing technologies detect viral RNA, p24 antigen (a viral protein that makes up most of the viral core), HIV antibodies (ELISA) and antibody/antigen in combination (Western Blot), and are known to have different window periods.

The relative times at which tests become positive were estimated by Fiebig et al. in a definitive 2003 paper [32], see Figure 3.1. In that work, Fiebig stage I is defined as the time period when only a clinical viral RNA test (with a threshold of 50 copies/ml of viral RNA) is positive. In Fiebig stage II a p24 antigen test would also be positive, and in Fiebig stage III sensitive HIV antibody tests are also positive. The individual durations of these first three Fiebig stages were estimated to be 5.0, 5.3, and 3.2 days, respectively. However, systematic estimates of the length of time between risky exposure and a viral RNA test becoming positive are still lacking [32, 103]. The length of the eclipse period is difficult to estimate because HIV transmission is rare [108] and precise exposure dates are usually not known. Typically the eclipse period is estimated to last for around 10-14 days: for instance the WHO has recently suggested that "the window period [for the most sensitive HIV antibody tests] is about three weeks" [113]. This implies an estimate of around 11 days for the length of the eclipse period, because according to Fiebig

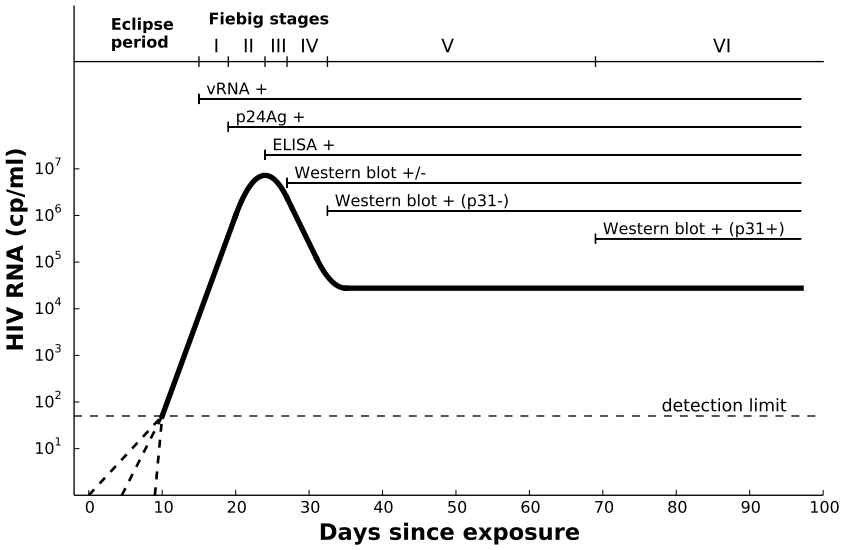


Figure 3.1: Schematics of typical dynamics of plasma virus RNA in a recently infected patient, and approximate time points when different HIV tests become positive, i.e. can detect infection. *vRNA* looks for virions directly, *p24Ag* looks for the p24 antigen, a viral protein. *ELISA* tests for antibodies, and *Western blot* combines several of the above. The dynamics below the detection limit are unknown, the unobserved growth rate could be smaller, the same, or larger than the observed growth rate, as indicated by the dashed lines. Modified from [32].

et al. a sensitive antibody test becomes positive about 10 days after an RNA test does [32].

Extrapolating results from animal models of infection, such as the rhesus macaque, must be done cautiously. Infection after a single sexual or occupational exposures in humans is highly unlikely (estimated at 4 per 10,000 events for penile-vaginal intercourse and 138 per 10,000 events for receptive anal intercourse) [69, 85], but high-dose simian immunodeficiency viruses (SIV) challenges (up to 10^3 – 10^5 more than the median required infective does) are commonly used to obtain infection, see e.g. [5, 81]. Even though mortality, the humoral immune response and the peak plasma viral load do not seem to differ in animals between high- and low-dose challenges, the timing of detectable infection may vary [76]. Further-

more, if a series of low-dose challenges are made, it can be difficult to know which challenge eventually led to the sustained infection [62, 92].

As a consequence, a knowledge gap concerning the overall lengths of the window periods for different tests remains. This lack of knowledge leads to considerable uncertainty and anxiety for patients and clinicians alike [54, 57, 70, 82, 116, 117]. The specificity (true negative rate) of an HIV test taken within the eclipse phase is also unknown, and hence it is unclear when to schedule further tests to balance the competing goals of early detection and reliable exclusion of infection.

In this work we extract time series of HIV RNA test results from seroconversion panels (a group of serial bleeds from an individual during early infection when the first antibody response is mounted) and seek to quantify the length of the eclipse period and the false negative rate of an HIV test as a function of time since exposure. From this data we first obtain the empirical maximal initial viral growth rate, similar to [93]. Then we explore and select stochastic models of virus dynamics from exposure to detectability of infection via RNA testing, by comparing fits to parameter ranges from the literature. Our data is intrinsically biased in that it only includes measurements from plasma donors that were eventually confirmed as infected. We therefore derive corrections to the basic models in order to adjust for this bias. The best model is then used to calculate, within the confines of our assumptions, the most likely time of exposure, and hence the length of the eclipse period. Finally, we quantify the specificity of an HIV RNA test as a function of time since exposure and put forward recommendations for the timing of follow-up tests.

Throughout this work we focus on commonly used tests that measure viral RNA, with 50 RNA copies per milliliter of plasma as the lower limit of detection, and we also assume that no antiretroviral treatment has been used shortly before or after the exposure and time of testing.

3.2 Extracting the maximal viral growth rates from seroconversion data

In order to answer our research questions on the early viral dynamics of HIV infection we use data from seroconversion panels that have been used and described

by others, in [93] (51 panels, no duplicates), and ourselves, in [103] (427 panels, including duplicates that are repeated in several tables). Seroconversion panels provide the earliest possible viral load measurements in newly HIV infected individuals. This data was obtained from regular plasma donors (plasma donations on average every five days) in the late 1990's and early 2000's, before RNA tests were routinely available. Plasma from regular donors is typically stored for several days before it is used to allow for screening and removal if any disease is discovered in the donor later. In our dataset each seroconversion panel contains the results of 2 to 26 HIV RNA tests on distinct days from a minimum of 4 to a maximum of 194 days since the first plasma sample was obtained from the plasma donor. Because the data was taken before sensitive HIV RNA tests were available, it was, most likely, an HIV antibody test that became positive in each of our plasma donor's last sample, which implied that all previously stored plasma donations from this newly infected individual were removed from circulation and frozen for future use in research. These frozen plasma samples are very rare and extremely valuable, and are for example used to benchmark the sensitivity of new HIV tests. All plasma samples in our dataset were later tested with new HIV RNA tests, similar to those commonly used to detect new infection in HIV clinics today. These new tests can detect and quantify the virus in the donor's plasma in samples where infection could not be detected by the less advanced HIV tests at the time of donation. Four typical examples of the seroconversion time series are shown in Figure 3.2.

After carefully cleaning the data to remove and combine duplicate time series, see Appendix B, we obtained, for each distinct time series, the maximal viral growth rate as the maximum slope between any two data points during the ramp-up phase, defined here as all time points before either the viral load drops by more than 75%, or the dataset ends. We also classify each dataset into one of the two categories: Class A are those datasets where the maximum slope is measured between two definite measurements, that is, measurements that are not below the detection limit (50 copies of RNA per milliliter of plasma) or above the maximum detectable (varying with the test). Datasets where the maximum slope involved a data point that was below the detection limit or above the detection limit were classified as class B. For this class, the actual maximal rate of increase may therefore be greater than what we inferred. To see if this potential underestimate is significant we com-

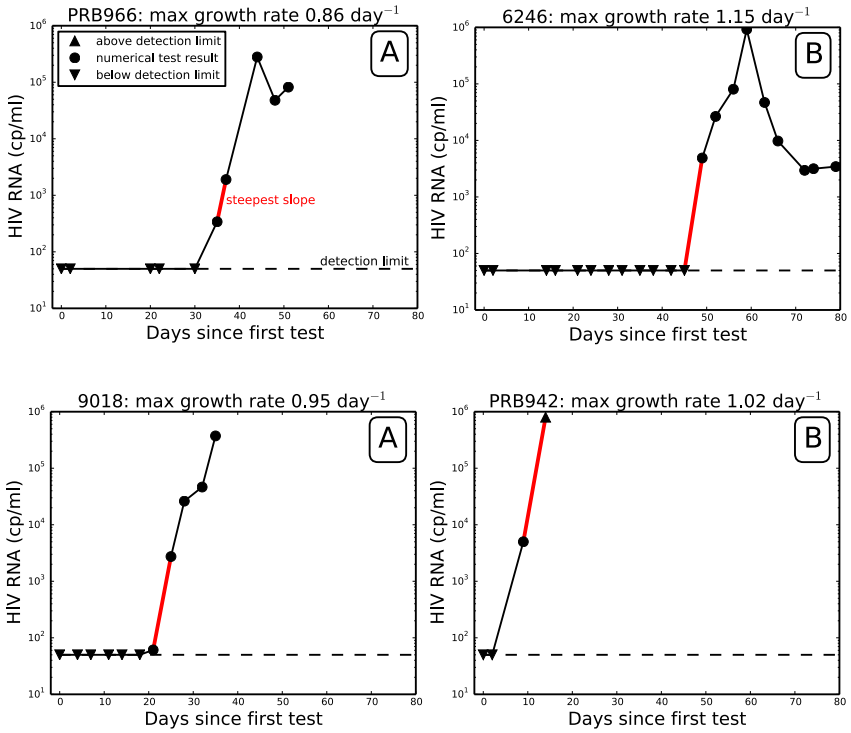


Figure 3.2: Examples of plasma virus RNA of recently infected patients, from the seroconversion panels obtained in [93] and [103]. Values below the detection limit of 50 copies/ml are reported as 50 copies/ml. The exposure dates for these patients are unknown and is to be estimated from the characteristic initial exponential growth rate and mathematical models in this work.

pared this set of 27 maximal rates of increase with the remaining 51 maximal rates (Figure 3.3), and found no statistical difference ($p = 0.25$ with Student's t-test). Based on this result we chose to treat all 78 slopes the same way for the remainder of this study. See Appendix B for the list of all time series and their classification.

Now that we have extracted the maximal viral growth rate for each of the suitable seroconversion panels we will carefully choose the most suitable mathematical model to obtain estimates for the length of the eclipse phase from these growth rates.

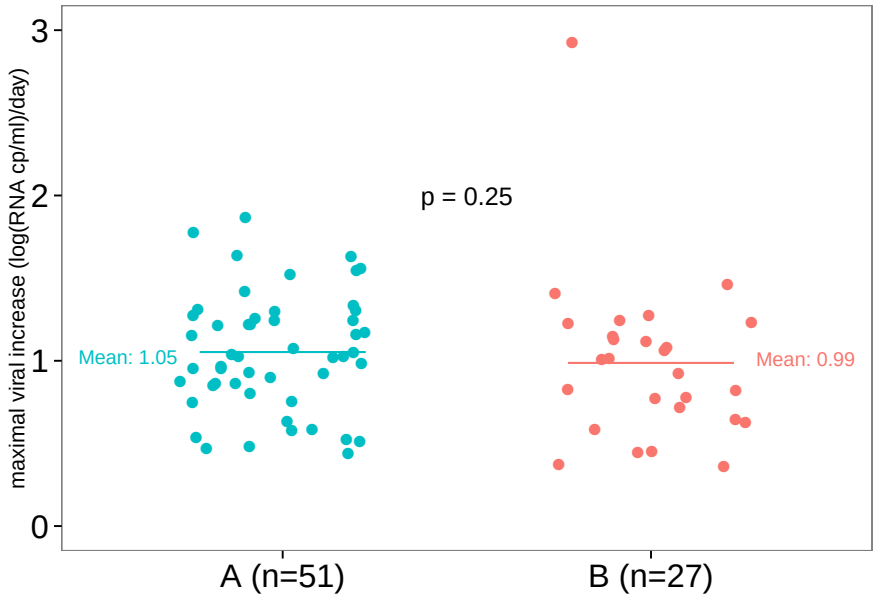


Figure 3.3: Maximal observed rate of increase of viral RNA concentration, as calculated for 51 class A seroconversion panels (mean: 1.05 log(RNA copies)/ml/day) and 27 class B panels (mean: 0.99 log(RNA copies)/ml/day). The difference between the two categories is not significant. If the outlier in group B is removed, the difference is borderline significant ($p = 0.048$).

3.3 The simple birth-death model

In the previous section we extracted the maximal initial growth rate of the virus population after it reaches a detectable level. This is the foundation of quantitative research into early HIV infection, and has been done by others before [93], albeit with only a subset of our data. The novel part of this chapter is how carefully selected mathematical models turn these observations into estimates of the length of the eclipse phase, and the clinical consequences of these estimates.

The within-host events leading to infection after HIV exposure are not fully known. Hence, instead of building a mathematical model with many unknown parameters which covers all possible scenarios in detail, we use the opposite approach and aim for the simplest model, with a minimal number of parameters, that

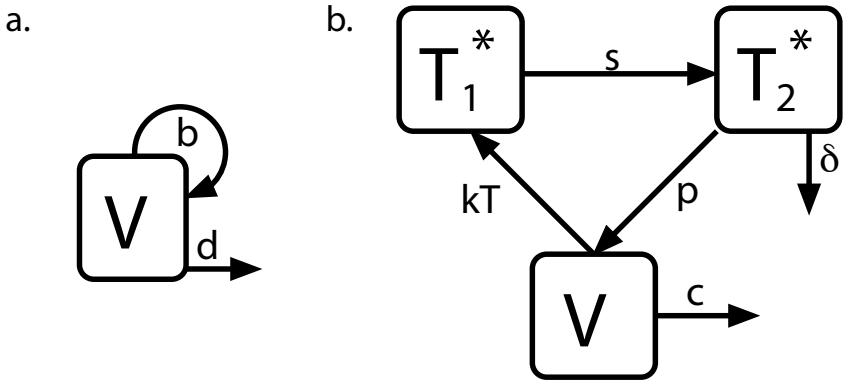


Figure 3.4: Model schematics. a. Minimal stochastic birth-death model. Virions reproduce at rate b and cleared at rate d. b. Preferred three-state model. Virions infect cells at rate kT and are cleared at rate c. At rate s infected cells become productively infected, producing new virions at rate p and dying at rate δ.

explains the observed data and fits our limited knowledge of early HIV infection.

The minimal requirement for a reasonable model of early HIV infection is that the model exhibits exponential growth that fits the innate growth rate of viral increase observed in the HIV RNA data. We also argue that the model has to exhibit a reasonable per-exposure probability of infection, in the range of 0.1%-5% [40, 85, 110], and allow for multiple founder strains, as observed in 19%-60% of cases, depending on the type of exposure [1, 8, 41, 61, 62, 71]. Estimating the size of the inoculum is extremely challenging, and it potentially depends heavily on the (unknown) type of exposure and the (unknown) viral load of the donor. *In vitro* experiments [85] suggest that 10-15,000 is a reasonable range for the number of virions that might breach the mucosal barriers and potentially initiate infection.

The simplest mathematical model that allows for exponential growth and viral extinction is the classic stochastic birth-death process for a single species (Figure 3.4a). The number of virions V increases via discrete stochastic reproduction events that occur at rate bV, and decreases via death events that occur at a rate dV. Starting with an initial population size of $V(0) = v_0$, the average number of virions is $E[V(t)] = v_0 e^{(b-d)t}$, corresponding to the solution of the differential equation

$dV/dt = bV - dV$. Later in this subsection will see that the simple birth-death model is not consistent with our knowledge of early HIV infection, and hence we will ultimately use a more sophisticated model for the data and research questions at hand. However, the exponential growth model of the expected value of the birth-death process allows a naive, yet natural approach to estimating the length of the window period, and hence deserves exploration:

3.3.1 The simple birth-death model does not adequately describe early infection

Following the simple birth-death process and setting the length of the eclipse phase τ to be the time it takes for the expected number of virions to grow from v_0 to detectable levels v_{det}, we quickly calculate

$$v_{\text{det}} = v_0 e^{(b-d)\tau} \quad \Longrightarrow \quad \tau = \frac{\ln(v_{\text{det}}/v_0)}{b-d}. \tag{3.1}$$

The slope $b-d$ corresponds to the observed viral growth rates from Figure 3.3, the detection limit is 50 RNA copies/ml, which corresponds to 50 [RNA/ml] $\times$ 15,000 [ml/human] / 2 [RNA/virion] = 375,000 [virions]. Varying the inoculum size v_0 as the most difficult-to-estimate parameter between large (400 virions), medium (75 virions) and small (10 virions) we obtain estimates of 7.8 days (4.2-15.2 days 95% confidence interval), 9.7 days (5.2-18.9 days) and 12.0 days (6.4-23.4 days), see Figure 3.5.

When we look at the stochastic birth-death model in more detail, however, we notice that it is not consistent with key HIV parameters. Indeed, using the expected value of the viral load as above ignores the likelihood that exposure leads to infection $\text{P}_{\text{ext}} = (d/b)^{v_0}$, only the slope of the growth rate $b-d$ is taken into account. However, a good early infection model should meet the following parameter restrictions to be consistent with the literature: (i) $\text{P}_{\text{ext}} \geq 0.95$, to match the per-exposure probability of infection, which is difficult to measure in an unbiased way, but most estimates suggest this is the least low bound, see e.g. [85, 108]; (ii) $10 \leq v_0 \leq 15,000$, to allow for multiple founder strains ([1, 8, 41, 61, 62, 71]) and match in-vitro estimates of the largest possible inoculum in vaginal transmission [14]; (iii) We expect the viral death rate d to be reasonably close to the measured

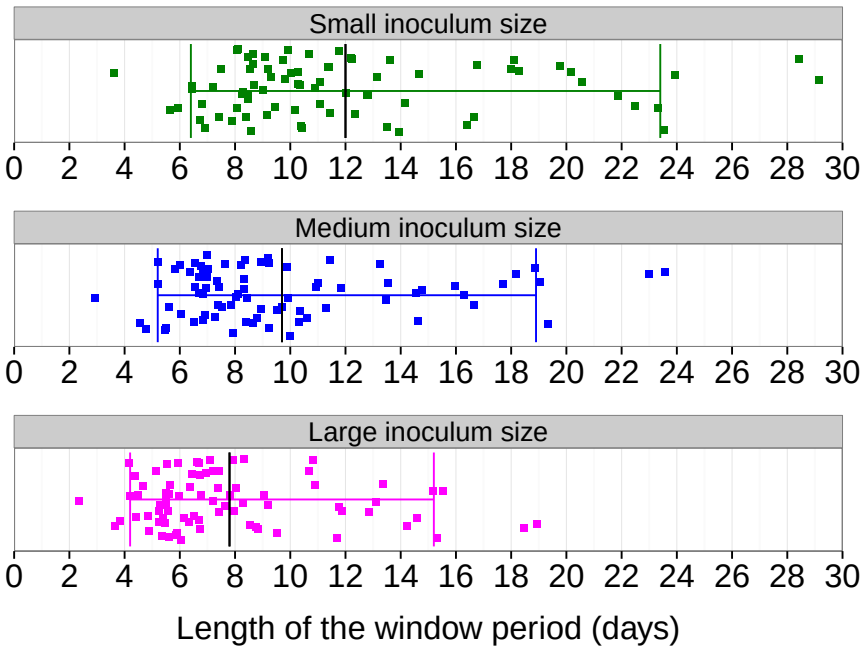


Figure 3.5: The estimated length of the eclipse phase for all 78 plasma donors according to the inoculumn size. Each square represents the estimate for one donor. The vertical bars indicate the mean and 95% confidence interval. Here we use the naive log-linear model (3.1), which is not able to capture key HIV characteristics, see Section 3.3.1. Hence this plot is an illustration-of-concept only. For the final results refer to Figure 3.8.

viral clearance rate of 23 day^{-1} [91, 119].

The birth-death model is sufficiently simple that the death rate can be expressed in simple form as

$$d = \frac{r{\mathrm{P}_{\mathrm{ext}}}^{1/v_0}}{1-{\mathrm{P}_{\mathrm{ext}}}^{1/v_0}},$$

where $r = b - d$ is the initial growth rate from the data. This expression for d increases with $\mathrm{P}_{\mathrm{ext}}$ and v_0 virions, yet even for the assumed lower bounds of $\mathrm{P}_{\mathrm{ext}} = 0.95$ and $v_0 = 10$ we obtain $d \simeq 2.0 \times 10^2$ virion^{-1} day^{-1}, which is already one order of magnitude larger than the literature value $d \simeq 23$ virion^{-1} day^{-1}. For the more reasonable estimate $\mathrm{P}_{\mathrm{ext}} = 0.98$ and $v_0 = 75$ the resulting death rate is more than 150 times larger than the literature value. As such, it is impossible for this

model to reflect our knowledge of early infection and therefore we moved on to a more complex model which explicitly includes productively infected cells as well as virions.

3.4 Including infected cells and virions: the T_1^*-T_2^*-V model

To improve on the inadequate model from Section 3.3 we introduce a new model which features infected and productively infected target cells T_1^* and T_2^*, respectively, as well as free virus V, see Figure 3.4b. In this three-state model, productively infected cells T_2^* produce new virions at rate p, and die at rate δ. Virions are cleared at rate c and infect new target cells T_1^* at rate kT. We assume that the depletion of target cells T can be neglected in the early dynamics we are considering, which allows us to use a constant infection rate kT. The rate at which infected target cells T_1^* become productively infected, T_2^*, is denoted by s. This delay is included to capture the experimentally-observed period between cell infection and the release of infectious virions. We also assume that the background death rate of non-productively infected cells is negligible in the early stages of infection.

To account for the stochastic nature of early HIV infection we study the system as a branching process for the probability $P_{m,n,v;m_0,n_0,v_0}(t)$ that $(T_1^*(t), T_2^*(t), V(t)) = (m,n,v)$ given the initial condition $(T_1^*(0), T_2^*(0), V(0)) = (m_0, n_0, v_0)$. The average number of virions as a function of time can then be calculated from the underlying system of linear differential equations

$$\begin{aligned} \frac{dT_1^*}{dt} &= kTV - sT_1^* \\ \frac{dT_2^*}{dt} &= sT_1^* - \delta T_2^* \\ \frac{dV}{dt} &= pT_2^* - (c+kT)V. \end{aligned} \tag{3.2}$$

This is a linear model and so we can calculate the average rate of viral increase r as the real part of the dominant eigenvalue of the differential equation system. The

eigenvalues are the roots of the characteristic polynomial

$$\lambda^3 + (s+\delta+c+kT)\lambda^2 + (s(\delta+c+kT)+\delta(c+kT))\lambda + s(\delta(c+kT)-pkT), \tag{3.3}$$

which can easily be computed numerically.

A standard calculation, see Chapter 2, gives the probability of viral extinction (and hence no ongoing infection),

$$\mathrm{P}_{\mathrm{ext}} = q_1^{m_0} q_2^{n_0} q_3^{v_0}, \tag{3.4}$$

where $q_1 = q_2 = \min\left\{1, \frac{\delta(c+kT)}{pkT}\right\}$, $q_3 = \min\left\{1, \frac{\delta}{p} + \frac{c}{c+kT}\right\}$ are the extinction probabilities for an infection started by a single infected cell, productively infected cell or a single virion, respectively [21, 22, 59, 86].

3.4.1 The T_1^*-T_2^*-V model is consistent with data and literature

The T_1^*-T_2^*-V model draws a more accurate picture of HIV infection than the simple birth-death model. However, we need to estimate more parameters. As mentioned above, estimates of the viral clearance rate c [91] have found $c = 23\,\mathrm{day}^{-1}$ in chronic infection, and a study in macaques [119] suggests that this is also a good estimate for early infection. Similarly, the death rate δ of productively infected cells in chronically infected patients is estimated to be around $1\,\mathrm{day}^{-1}$ [74], and exploring a subset of the seroconversion panels that we analyse in this paper, Ribeiro et al. [93] estimated $\delta = 0.6\,\mathrm{day}^{-1}$ using the post-peak viral load decay rate. The average rate s at which infected cells become productively infected has been estimated as $1\,\mathrm{day}^{-1}$ [25]. We further assume a virus-only inoculum and hence set $T_1^*(0) = T_2^*(0) = 0$. This leaves the three parameters kT (infection rate), p (production rate) and v_0 (virus inoculum). We use the following information to constrain these: (i) as above, we assume the virus inoculum ranges between $10 \leq v_0 \leq 15{,}000$; (ii) we calculate the infection rate kT and the production rate p such that $0.95 \leq \mathrm{P}_{\mathrm{ext}} \leq 0.995$ and that simultaneously the growth rate r of the model matches the maximal growth rate obtained from the seroconversion panels. These constraints on kT and p are described via the eigenvalues of system (3.2)

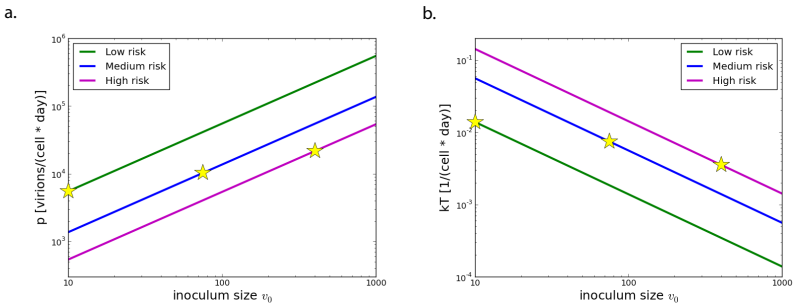


Figure 3.6: The interdependence of the parameters of the T_1^*-T_2^*-V model. From left to right, the stars indicate particular choices of parameters in three different risk scenarios: low risk and small inoculum size, medium risk and medium inoculum size, and high risk and large inoculum size, see Section 3.5. The resulting fitted values for p and kT are within the estimates from the literature. **a.** The production rate p as it depends on v_0 and risk of infection. **b.** The infection rate kT as it depends on v_0 and risk of infection.

and equation (3.4).

For the T_1^*-T_2^*-V model to be compatible with our general knowledge of early HIV infection we expect the fitted infection rate kT and production rate p to be comparable with estimates from the literature. However, since these parameters are difficult to measure experimentally, especially in early infection, there exist only very rough estimates: reports of the infection rate kT and virion production rate p both span several order of magnitude: $kT \simeq 10^{-3} - 10^{1}$ cell^{-1} day^{-1} [88], and $p \simeq$ 200-20,000 virions cell^{-1} day^{-1}[17].

Figure 3.6 shows the resulting fits for kT and p as v_0 and $\mathrm{P_{ext}}$ vary. In particular, for a large range of v_0 and $\mathrm{P_{ext}}$ the fitted parameters from the T_1^*-T_2^*-V model agree with the wide parameter ranges from the literature. Hence, unlike the simple birth-death model, the T_1^*-T_2^*-V model can be parameterized consistently with the limited data available for early HIV infection. This is the model we will apply from now on.

3.5 Risk scenarios

We choose three scenarios within the acceptable area in the parameter space (Figure 3.6) to calculate the length of the eclipse period, that is, given an infection, the length of time from exposure to the reliable detection of the HIV infection by an RNA test. Our choice of per-exposure risk is on the higher end of the large spectrum found in the literatures, because many study designs, in particular those of steady partnerships, may bias towards a low risk of transmission (see [108] for more details), and because a higher per-exposure risk of infection leads to a longer eclipse period in our model. This way the upper bounds that we will report can be thought of as a worst-case. We present results for each risk scenario, using the average of the observed maximal viral growth rates from the data:

1. *High risk large inoculum size* with per-exposure risk of infection of 5% and inoculum size of 400 virions (then: $kT = 3.6 \times 10^{-3}$ cell^{-1} day^{-1}, $p =$ 21,970 virions/day),

2. *Medium risk medium inoculum size* with per-exposure risk of infection of 2% and inoculum size of 75 virions (then: $kT = 7.5 \times 10^{-3}$ cell^{-1} day^{-1}, $p = 10,460$ virions/day),

3. *Low risk small inoculum size* with per-exposure risk of infection of 0.5% and inoculum size of 10 virions (then: $kT = 1.4 \times 10^{-2}$ cell^{-1} day^{-1}, $p = 5,621$ virions/day).

3.6 The infection bias and its consequences

An inclusion criterion for our data is that the plasma donors ultimately became infected. HIV exposures that did not lead to HIV infection are not recorded. To reflect this bias in our model, we have to restrict the stochastic process to exclude extinction. We will denote the conditioned process, which will be used to fit the data, with a tilde.

3.6.1 Mathematical derivation

For a more detailed discussion of the mathematical background and derivation of the conditioned model formulation for branching processes like the T_1^*-T_2^*-V mod-

els can be found in Chapter 2. We use Bayes' rule and the Markov property to rewrite the probability

$$\begin{aligned}\tilde{P}_{m,n,v;m_0,n_0,v_0}(t) = P((T_1^*(t),T_2^*(t),V(t)) = (m,n,v) \,|\, \\ (T_1^*(0),T_2^*(0),V(0)) = (m_0,n_0,v_0), \\ (T_1^*(\infty),T_2^*(\infty),V(\infty)) \neq (0,0,0))\end{aligned}$$

of the process conditioned on not going extinct as

$$\tilde{P}_{m,n,v;m_0,n_0,v_0}(t) = P_{m,n,v;m_0,n_0,v_0}(t)\frac{1-q_1^m q_2^n q_3^v}{1-q_1^{m_0} q_2^{n_0} q_3^{v_0}}.$$

The conditioned process has the generating function

$$\begin{aligned}\tilde{G}_{m_0,n_0,v_0}(t,z_1,z_2,z_3) &= \sum_{m,n,v=0}^{\infty} \tilde{P}_{m,n,v;m_0,n_0,v_0}(t) z_1^m z_2^n z_3^v \\ &= \frac{G_{m_0,n_0,v_0}(t,z_1,z_2,z_3) - G_{m_0,n_0,v_0}(t,q_1 z_1, q_2 z_2, q_3 z_3)}{1-q_1^{m_0} q_2^{n_0} q_3^{v_0}},\end{aligned}$$

where $G_{m_0,n_0,v_0}(t,z_1,z_2,z_3) = \sum_{m,n,v=0}^{\infty} P_{m,n,v;m_0,n_0,v_0}(t) z_1^m z_2^n z_3^v$ is the generating function of the unconditioned process which satisfies the branching property

$$G_{m_0,n_0,v_0}(t,z_1,z_2,z_3) = G_{1,0,0}(t,z_1,z_2,z_3)^{m_0} G_{0,1,0}(t,z_1,z_2,z_3)^{n_0} G_{0,1,0}(t,z_1,z_2,z_3)^{v_0}.$$

Here

$$\begin{aligned}\frac{d}{dt}G_{1,0,0} &= s(G_{0,1,0} - G_{1,0,0}), \\ \frac{d}{dt}G_{0,1,0} &= \delta(1-G_{0,1,0}) + pG_{0,1,0}(G_{0,0,1}-1), \\ \frac{d}{dt}G_{0,0,1} &= kT(G_{1,0,0} - G_{0,0,1}) + c(1-G_{0,0,1}),\end{aligned} \tag{3.5}$$

with $G_{1,0,0}(0) = z_1$, $G_{0,1,0}(0) = z_2$, $G_{0,1,0}(0) = z_3$. From the generating function we obtain the average number of virions $E[\tilde{V}(t)]$ of the conditioned process by

calculating

$$
\begin{aligned}
E[\tilde{V}(t)] &= \sum_{m,n,v} v\tilde{P}_{m,n,v;m_0,n_0,v_0}(t) = \left.\frac{\partial \tilde{G}_{m_0,n_0,v_0}(t,z_1,z_2,z_3)}{\partial z_3}\right|_{(z_1,z_2,z_3)=(1,1,1)} \\
&= \frac{1}{1-q_1^{m_0}q_2^{n_0}q_3^{v_0}}\left[\left.\frac{\partial G_{m_0,n_0,v_0}(t,z_1,z_2,z_3)}{\partial z_3}\right|_{(z_1,z_2,z_3)=(1,1,1)}\right. \\
&\qquad\qquad \left. - \left.\frac{\partial G_{m_0,n_0,v_0}(t,q_1z_1,q_2z_2,q_3z_3)}{\partial z_3}\right|_{(z_1,z_2,z_3)=(1,1,1)}\right] \\
&= \frac{f_3(t)-g_3(t)}{1-q_1^{m_0}q_2^{n_0}q_3^{v_0}}. \qquad (3.6)
\end{aligned}
$$

Here f_3 is the solution corresponding to the V variable in (3.2) above. The relevant system of equations for g_3 can be derived by direct partial differentiation of G_{m_0,n_0,v_0}, which leads to the system

$$
\begin{aligned}
\dot{g}_1 &= \frac{q_1}{q_3}kTg_3 - sg_1, & g_1(0) &= m_0q_1^{m_0}q_2^{n_0}q_3^{v_0}, \\
\dot{g}_2 &= sg_1 - \frac{\delta}{q_2}g_2, & g_2(0) &= n_0q_1^{m_0}q_2^{n_0}q_3^{v_0}, \\
\dot{g}_3 &= pq_3g_2 - (c+kT)g_3, & g_3(0) &= v_0q_1^{m_0}q_2^{n_0}q_3^{v_0}.
\end{aligned}
$$

We solve the two linear systems of differential equations numerically and then use (3.6) to calculate and eventually fit the average number of virions of the T_1^*-T_2^*-V model conditioned on non-extinction.

3.6.2 The effect of the infection bias on the estimated length of the eclipse phase

To illustrate and quantify the different risk scenarios and the effect of the infection bias we plot the fit of the mean of the T_1^*-T_2^*-V model to the data of four sample plasma donors *1018*, *PRB966*, *9018* and *PRB946*, for each risk scenario with and without conditioning (Figure 3.7) on infection. Initially the viral load drops below the inoculum size, because the inoculum contains no target cells, and hence inoculum-virions need to first infect new target cells before new virions can be produced. However, the average drop of viral load and initial rate of viral increase

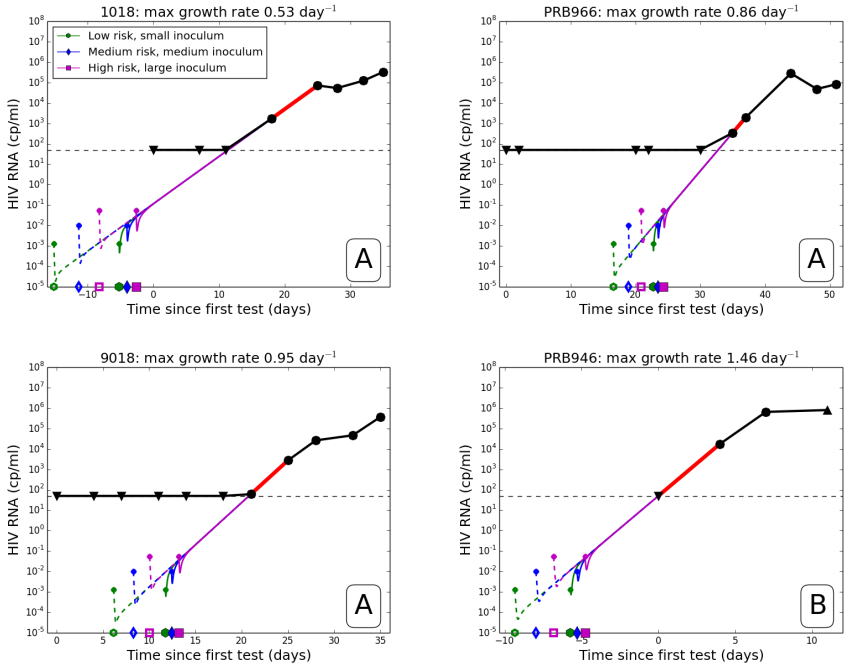


Figure 3.7: The effect of the infection bias on the estimated exposure date, as obtained by fitting the T_1^*-T_2^*-V model, unconditioned (dashed lines, open markers) and conditioned (solid lines, filled markers), for each exposure category: low risk (pentagon), medium risk (diamond) and high risk (square).

differs among the risk categories, and also depends on whether the infection bias is taken into account. The fitted eventual growth rate, however, is reached very quickly by all models, so that the estimated viral load agrees for all models well before the detection limit is reached.

The big difference between the model estimates of the conditioned and unconditioned model occurs because the risk of extinction (which we condition against) declines with population size: initially, when the number of infected target cells and viral particles is still small, there is a high chance that these populations go extinct. Once the populations are large enough, the risk of extinction becomes negligible. Therefore, the effect of conditioning on non-extinction is more pro-

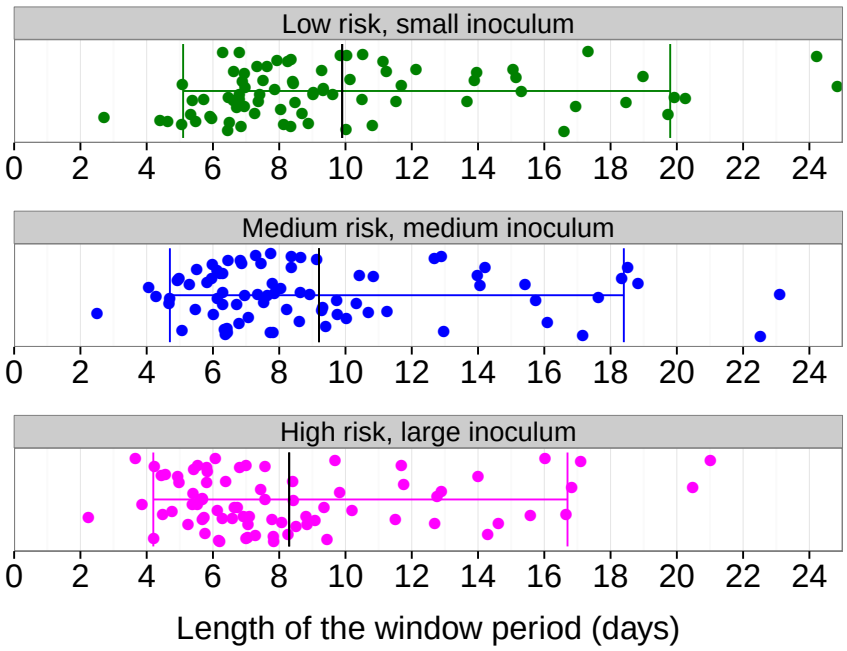


Figure 3.8: The estimated length of the eclipse period for all 78 plasma donors according to the three different exposure-risk categories, using the conditioned T_1^*-T_2^*-V model. Each dot represents the estimate for one donor. The mean and 95% confidence intervals are indicated by the vertical bars.

nounced around the time of exposure and for higher risk scenarios. Ignoring this effect and naively fitting the slope and the inoculum size neglects the bias inherent in the data and hence predicts an earlier time of exposure, which leads to an overestimate of the length of the eclipse period (see Figure 3.5)

3.7 Quantifying the length of the eclipse period on a population level

We used the conditioned three-state model to calculate the estimated lengths of the eclipse period for each plasma donor in our data set (Figure 3.8), for each risk category described above. We observe that the variability between the risk categories is much smaller than the variability within the risk categories. This

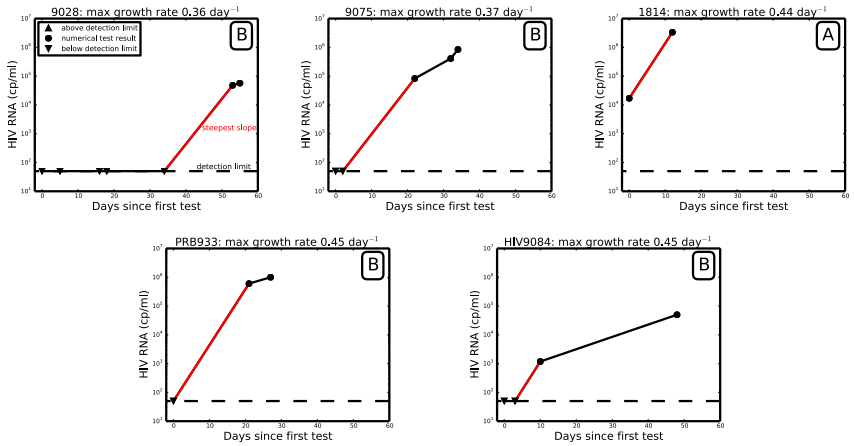


Figure 3.9: The five plasma donors with the smallest maximal growth rates are at or above the 95th percentile of the length of the eclipse period. Due to the relatively long time-interval between the RNA measurements where the maximal growth rate occurs, and the relatively high RNA loads, it is quite likely that the real maximal growth rate may have been missed for these donors.

shows that the uncertainty in the choice of risk category has only minor effects on our main results. Indeed, the average length of the eclipse period is similar between the risk groups: *high-risk* 8.3 days (95% CI: 4.2-16.7 days), *medium-risk* 9.2 days (4.7-18.4 days), *low-risk* 9.9 days (5.1-19.8 days). These confidence intervals are quite large, from about half the average value to twice the average value.

The skew to longer eclipse periods is driven by a small number of donors for whom the estimated eclipse period is very long — up to four weeks in the highest risk category. Generally for these five donors the estimated length of the eclipse period is at or above the 95% confidence interval. These are the donors with the smallest maximal viral growth rate, namely 0.36/day (*9028*), 0.37/day (*9075*), 0.44/day (*1814*), 0.45/day (*PRB933*) and 0.45/day (*HIV9084*). Recall that a small measured maximal viral growth rate leads via fitting to slow population growth and hence a longer eclipse period. Looking at the seroconversion panels of these donors (Figure 3.9) it is possible to argue that the observed maximal growth rate is an underestimation of the actual maximal growth rate: The time interval between the two

test results that account for the observed growth rate is 19 days (9028), 20 days (9075) and 21 days (PRB933), and all viral loads used to determine the maximal growth rates are greater than 4.8×10^4 RNA cp/ml. For HIV9084 the time interval between the tests was as long as 38 days, while the first viral load measurement for donor 1814 was already very high at 1.7×10^4 RNA cp/ml. Long time intervals between two tests make it more likely that the actual maximal growth rate has been missed, and so do high RNA measurements since the maximal viral growth rate is typically found between two measurements with low RNA loads. We conclude that if these outliers are actually not properly represented by their seroconversion time course data, an upper bound of 17 days for the full eclipse period is a very reasonable estimate. Eliminating these five donors from the plot in Figure 3.8 gives the following results for the length of the eclipse period: *high-risk* 7.6 days (95% CI: 4.2-14.1 days), *medium-risk* 8.4 days (4.7-15.6 days), *low-risk* 9.1 days (5.1-16.7 days).

As a final point, we also computed the distributions of the length of the eclipse using the unconditioned T_1^*-T_2^*-V model and found that the unconditioned model substantially overestimates the duration of the eclipse period. The difference in the mean over all donors was found to be 3.4 days for the *high-risk* scenario, 4.5 days for *medium-risk*, and 6.0 days for the *low-risk* scenario on average. This shows the great importance of correctly taking the infection bias in the data into account by using the conditioned model.

3.8 Probability of a false negative HIV RNA test

3.8.1 Mathematical derivation

Given a lower detection limit of dL total virions of the HIV RNA test, we calculate the probability that a plasma donor will be infected but have an undetectable viral load at time t after exposure as

$$P(\tilde{V}(t) < \mathrm{dL}) = \sum_{v=0}^{\mathrm{dL}-1} P(\tilde{V}(t) = v)$$

$$
\begin{aligned}
&= \sum_{v=0}^{\mathrm{dL}-1} \sum_{m,n=0}^{\infty} \tilde{P}_{m,n,v;m_0,n_0,v_0}(t) \\
&= \sum_{v=0}^{\mathrm{dL}-1} \frac{1}{v!} \frac{\partial^v \tilde{G}_{m_0,n_0,v_0}(t,1,1,z_3)}{\partial z_3^v}\bigg|_{z_3=0} \\
&= \sum_{v=0}^{\mathrm{dL}-1} \frac{1}{\pi} \mathrm{Re}\left(\int_0^{\pi} \tilde{G}_{m_0,n_0,v_0}(t,1,1,e^{i\varphi}) e^{-iv\varphi}\, d\varphi \right) \\
&= \frac{1}{\pi} \mathrm{Re}\left(\int_0^{\pi} \tilde{G}_{m_0,n_0,v_0}(t,1,1,e^{i\varphi}) \frac{1-e^{-i\mathrm{dL}\varphi}}{1-e^{-i\varphi}} d\varphi \right).
\end{aligned}
$$

Note that this formula only applies to plasma donors who become infected following exposure. The technical details of this approach are found in Chapter 2 and our previous work [21, 22].

From the expression above we infer the probability that a person is HIV infected despite having a negative RNA test at time t after exposure. Using Bayes' rule

$$
P(A\,|\,B) = \frac{P(A)P(B\,|\,A)}{P(A)P(B\,|\,A) + P(\overline{A})P(B\,|\,\overline{A})},
$$

we choose A as "HIV exposure leads to infection" and B as "negative HIV RNA test at time t". Using a false positive rate of α (specificity of $1-\alpha$) we obtain $P(B\,|\,\overline{A}) = 1-\alpha$. For a false negative rate of β (sensitivity of $1-\beta$) we obtain that

$$
\begin{aligned}
P(B\,|\,A) &= P(\tilde{V}(t) < \mathrm{dL}) + \beta P(\tilde{V}(t) \geq \mathrm{dL}) \\
&= 1 - (1-\beta) P(\tilde{V}(t) \geq \mathrm{dL}).
\end{aligned}
\tag{3.7}
$$

Denoting the probability that HIV exposure does not lead to infection by $P(\overline{A}) = \mathrm{P_{ext}}$, we obtain the explicit formula

$$
\begin{aligned}
&P(\text{infected despite negative RNA test at } t) \\
&\qquad = \frac{(1-\mathrm{P_{ext}})(1-(1-\beta)P(\tilde{V}(t) \geq \mathrm{dL}))}{(1-\mathrm{P_{ext}})(1-(1-\beta)P(\tilde{V}(t) \geq \mathrm{dL})) + \mathrm{P_{ext}}(1-\alpha)}.
\end{aligned}
\tag{3.8}
$$

In particular, we see that the likelihood of being infected despite a negative RNA test is bigger for tests that are not perfectly specific and not perfectly sensitive.

3.8.2 Results and implications for the interpretation of HIV test results and the timing of follow-up tests

We calculate, for each plasma donor, the probability that the viral load has increased above the detection threshold for an HIV RNA test at time t after exposure. This likelihood increases steadily from 0 (at the time of exposure) to its asymptote at 1 (since we condition our analysis successful infection), see Figure 3.10. We average these likelihood curves to obtain, under the assumption that our data gives a fair representation of the general at-risk population and given that infection will occur, an estimate for the probability of a positive HIV RNA test t days after exposure, for an unknown individual who might present at an HIV clinic. These curves are very similar for each risk category, so in Figure 3.10 we only show the likelihood curves for the low risk category, which estimates the longest eclipse periods.

Remaining risk		Day (since exposure) of negative test. Values for (α, β).			
relative	absolute	$(0,0)$	$(2\%,0)$	$(0,2\%)$	$(2\%,2\%)$
95%	0.475%	5.2	5.5	5.2	5.5
90%	0.450%	5.9	6.1	5.9	6.1
75%	0.375%	7.2	7.3	7.2	7.3
50%	0.250%	9.2	9.3	9.3	9.4
25%	0.125%	12.8	12.9	13.2	13.4
10%	0.050%	18.8	18.9	19.8	19.9
5%	0.025%	22.2	22.3	24.6	24.7

Table 3.1: Assuming a per-exposure risk of 0.5% this table shows the required waiting time until a negative HIV RNA test reduces the risk of infection to the desired level. The waiting time is longer for tests with non-zero false positive rate α and non-zero false negative rate β.

However, in the clinical scenario we would not know if the patient is actually infected or not. Hence the more applicable metric in this case is how much the confidence in non-infection increases by a negative result from an RNA test. Intuitively, after a very early negative test one can hardly assume a lower relative risk of infection, whereas a negative test several days or weeks after the exposure dramat-

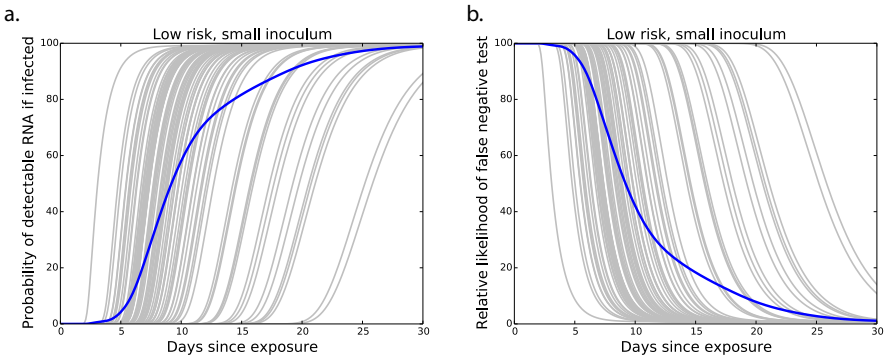


Figure 3.10: Individual and aggregate likelihood of detectable RNA as a function of days since exposure, and relative probability of a false negative test. Each grey line represents the estimate for one plasma donor. The blue curve is the average of all donors in the dataset and hence provides our best estimate for a random patient from the general population. The largest information gains are between day 5 and day 10. Plotted for *low risk, small inoculum* to provide an upper bound on the length of the eclipse period. **a.** Probability of detectable RNA t days after exposure, if infection occurs eventually. **b.** Probability of a false negative test, assuming that the test is perfectly sensitive and specific, relative to the baseline probability of infection. The results for non-perfect tests are similar and summarized in Table 3.1.

ically decreases the relative risk of infection. Our estimates of relative confidence are displayed in Figure 3.10, and also summarized in Table 3.1.

For both plots and the table in this subsection we conservatively choose the parameters for the *low risk small inoculum size* exposure scenario. This scenario presents the longest estimated eclipse periods in our model. The results for the other exposure categories are similar (not shown). We conclude that very little information would be gained by a repeated test before day five, and that the additional information gained in tests more than three weeks after exposure is very little. Therefore, given a negative HIV RNA test after a recent potential exposure, our results suggest that a test three weeks after the potential HIV exposure should

catch 95% of real HIV infections. Removing the five possible outlier cases (as discussed in the previous section) increases this estimate to 100%.

3.9 Discussion

Uncertainty about the length of the window period for HIV infection remains a source of anxiety for clinicians and patients alike. Reliable estimates are difficult to obtain, because, by definition, the virus is not detectable during that time. Efforts to close this knowledge gap have revealed important insights into the relative time period between the eclipse phases of different HIV tests [32], which has recently been refined by ourselves for newer antibody tests [103]. Our goal here was to calculate the absolute length of the eclipse phase for HIV tests that measure the viral RNA instead of antigens or antibodies and therefore promise the shortest eclipse phase. Knowing this absolute time period is critical because the relative estimates for the eclipse phases of other tests are based on it. Our results suggest that, in more than 95% of the plasma donors, this period lasts between four and 20 days, with an average of eight to ten days, depending on the assumed risk category.

To obtain these results we extracted the maximal initial viral growth rate from HIV seroconversion panels and fitted simple stochastic models, using parameter estimates from the literature. From a large choice of available mathematical models of within-host dynamics of HIV infection [87] we chose a model that is as simple as possible, but not any simpler. We chose to include the delay between infection and viral production of a target cells, compared to the slightly simpler T^*-V model, because this delay slightly affects our estimates of the length of the eclipse period (by about half a day on average, not shown), and the single additional parameter s is well-estimated in the literature.

A major limitation of the T_1^*-T_2^*-V model is that it assumes that infection happens in a well-mixed environment. However, depending on the type of exposure the inoculum has to breach various mucosal layers and spatial effects may become important. We neglected the delay before the inoculum breaches the mucosa, but since this delay is assumed to be on the order of a few hours [43, 44], we do not expect that any our results would change considerably. Similarly, the mucosal layers greatly constrain the number of donor virions that eventually reach infectible

tissues, which we accounted for by choosing a relatively small effective inoculum size in all of our risk categories. Following our principle of keeping the model simple we also chose to not explicitly include other potentially important effects, like the formation of local foci of infection and rapid viral replication in lymph nodes (see [42, 44]). If any of these effects significantly modulates the time-course of early infection away from an essentially exponential process, then our eclipse period estimates would need to be adjusted. However, in the absence of compelling evidence we believe that our simplified models are the most parsimonious choice.

Interestingly, our analysis shows that a simplistic exponential-growth model can not be used to describe early HIV infection. Further, we determined that it is important to account for the infection bias in seroconversion data: viral load growth rates are only available for plasma donors that were eventually infected, yet most exposures do not lead to infection. Conditioning our model to account for this fact, we showed that the length of the window period would be considerably overestimated if this effect was not taken into account.

Finally, we use our model to calculate how quickly the likelihood of a false negative test decreases over time since exposure. This result can help patients and clinicians alike to understand the informative value of a negative HIV RNA test. In our model the likelihood of a false negative test decreases most rapidly between 5 and 10 days after exposure. HIV RNA tests typically have a very high sensitivity, which is often reported as perfect, and also a very high, but not perfect, specificity, see [13]. Either error increases the waiting time, so we considered a relatively poor sensitivity and specificity of 98%. We further assumed here that infection could be detected at the exact moment that the viral load reaches the detection limit, whereas in reality the chance of detection increases more steadily with viral load, due to the uncertainty in the measurement and fluctuations in the viral load of the patient's blood stream. Unfortunately, we are not aware of any existing work that can be used to quantify this uncertainty. In future work, as this data becomes available, it would therefore be interesting to include a more sophisticated viral-load dependence in the likelihood of detecting the infection.

Chapter 4

Modelling the population-level impact of increasing HIV testing

4.1 Motivation and background

With an estimated 71,300 (range: 58,600 to 84,000) people living with HIV in Canada in 2011, the overall population prevalence is relatively low at 0.2% [83]. However, when stratifying by exposure category it becomes clear that the epidemic is concentrated in the men who have sex with men (MSM) community (46.7% of all HIV infections) and among injecting drug users (IDUS) (16.6%). Indeed, in British Columbia 59% of all newly diagnosed infections are identified as belonging to the MSM exposure category, and the HIV prevalence in men who have sex with men (MSM) in Vancouver is estimated to be approximately 15% [34, 77]. Many studies have been administered by the British Columbia Centre for Disease Control (BCCDC) and others to better understand the risk behaviour and disease dynamics of the MSM population, so that HIV prevention programs can better target this key risk group. A major source of such information is the largest online survey of gay and bisexual men in Canada is *Sex Now*, which has been conducted since 2002 and annually collects data from thousands of MSM [15].

The importance of the internet as a way to facilitate disease transmission was first highlighted by the 2004 Sex Now survey, which found that a majority of MSM meet sexual partners online. At the same time, this group was also found to use

the web to look for health information, showing that an online presence has the potential of unprecedented reach to a population at risk of infection. This led to BCCDC's first online HIV intervention, the *Cyber Outreach Program* in 2004, which offered sexually transmitted infection (STI)/HIV education and live chat services [33, 84]. This was also around the time when the idea of internet-based testing was conceived, which eventually led to the initiation of the Get Checked Online (GCO) program by the BCCDC.

The goal of the study in this chapter is to quantitatively predict the population impact of increasing HIV testing, with initiatives such as GCO, on the prevalence and incidence of HIV among Vancouver MSM in the next five years. This estimate will serve as a key input to future cost-effectiveness studies and, as such, will help guide the expansion of online testing in British Columbia (BC), other canadian cities and beyond. First, we by explain how the GCO program works in practice. Next, we describe how the data for our study was collected. Then we present a basic model of susceptible and HIV infected individuals that accounts for behaviour change depending on the awareness of infection. The basic model not only serves as a point of reference, but we also use it to explain the modeling techniques and key parameters that we need to identify from the data. We then show a technique for calculating the sensitivity of the model outcome as it depends on the parameter values, before we discuss the results and insights we gain from the basic model. In the following sections we introduce the full model, which incorporates different testing and risk behaviour, and describe how we obtain the additionally required parameters from the data. Finally, we summarize our results and predictions, and indicate how this work could be extended.

4.2 How Get Checked Online works

GCO is a program to improve sexual health by increasing the frequency of STI/HIV testing, using clinical resources more efficiently, and reaching greater at-risk populations, such as the youth or people from rural areas that may otherwise face barriers to accessing testing. By reducing wait times and increasing the timeliness of diagnosis GCO may increase the rate of treatment, induce risk-behaviour change and ultimately reduce onward STI/HIV transmission. On the GCO website users

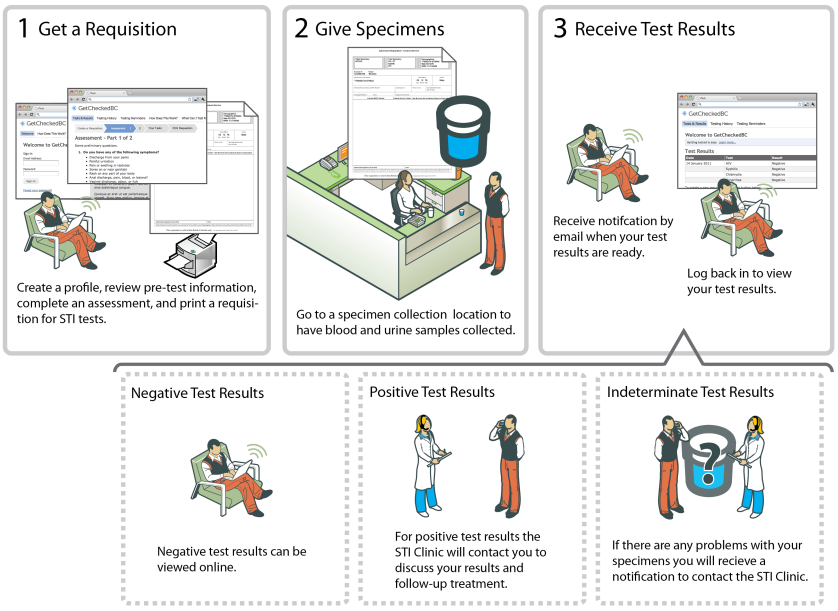


Figure 4.1: Sketch of how the GCO program works, first published in [53]. After answering a questionnaire that assesses their risk and recommends which STIS to test for, users print their requisition from home. Specimens can be given anonymously at participating lab locations, and test results can be viewed online if all tests are negative. Otherwise the STI clinic contact the user to discuss erroneous specimens as well as indeterminate or positive results.

with an account answer a questionnaire similar to what they would be asked in a STI/HIV testing clinic. Based on the answers to these questions the system compiles a Portable Document Format (PDF) document of recommended tests that the person can print and bring to a partnering laboratory where they can give the required specimens for the corresponding STI test. Once all test results are available the person gets notified to login and view their results. The results are shown only if all tests are negative. If any of the tests is positive, indeterminate, equivocal or if there is any problem with any of the specimens, no result is displayed. Instead, a message is shown to contact a clinic for follow-up and nurses will make every effort to contact the person.

It is important to note that the entire GCO process can be done anonymously.

Each login can be linked to a first and last name, but these fields are not checked and pseudonyms are explicitly allowed. Either way, the name does not appear on the printed lab requisition, instead a unique ID is used to link the test results to the corresponding person. For further privacy protection no test result is stored longer than 30 days online - results only remain available in the clinical information system which holds the complete medical record. Another benefit of GCO is a built-in email-testing-reminder that accounts for the risk category as identified in the questionnaire.

GCO launched on September 9th 2014 by email invitation only, and was fully launched in January 2015 when hand-out access codes were distributed to drop-ins and turn-aways at testing clinics, and social media marketing campaigns were run. Currently GCO is available for chlamydia, gonorrhea, syphilis, HIV as well as hepatitis C virus testing.

4.3 Survey data collection

The author was not involved in the collection of the raw data.

The study population was recruited from June 18, 2011 to March 2, 2012 (10 months) at health Initiative for Men (HIM) Sexual Health Centers in Vancouver. Eligibility conditions were that the participant was at least 19 years of age, self-identified as gay, bisexual or MSM, intended to reside in the Greater Vancouver area for the full year of follow-up, and received a negative HIV test at the recruiting HIM sexual health clinic. Of the 1141 eligible men 194 consented and 166 (14.5%) completed the baseline survey.

In total there were four *surveys* and *network grid interviews*, respectively, on day 7, day 30, day 180 and day 360 after recruitment. The surveys were online and self-administered with questions on demographics, HIV testing patterns, personal sexual history and substance use, disclosure of HIV status and risk of HIV, HIV knowledge and attitudes, mental health as well as knowledge of other campaigns targeted at MSM. The network grids were phone interviews with detailed questions about the five most recent sex partners in the past six months, such as the length of the relationship, frequency of sex, knowledge of mutual HIV status, and the types of sexual acts they engage in.

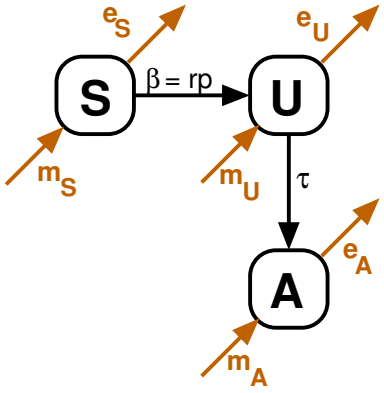


Figure 4.2: Schematics of the basic model for GCO. Individuals can be susceptible, infected and unaware, or infected and aware of their infection. The mass-action infection rate $\beta = rp$ is a product of the exposure-rate r and the per-exposure probability of HIV transmission p, and is reduced for individuals who are aware of their infection. GCO aims to increase the testing rate τ with which infected individuals become aware of their infection. The model also accounts for turnover via immigration m and emigration e in all three compartments.

The survey and grid on day 30 were mainly designed as follow-ups on the recent test at the recruiting HIM clinic and did not feature the full set of questions that are relevant in our study. We therefore choose to focus on the data from the three surveys and three network grids administered on day 7, day 180 and day 360.

Parameter values that cannot be estimated from this data are taken from the literature.

4.4 The basic model with homogeneous testing patterns and risk behavior

In the basic model we account for susceptible (denoted by S) and infected individuals, where infected individual are either unaware (U) or aware (A) of their infections, see Figure 4.2. All three groups are subject to immigration and emigration at rates m and e, respectively. We denote r the rate of "risky events", that is, events where HIV may be transmitted and we assume that r is independent of the knowledge of a person's HIV status. We further assume that a person who knows

that they are infected has a reduced risk of transmission, e.g., by using condoms, serosorting (choosing sexual partners with the same serostatus) or being on treatment. The per-encounter risk of disease transmission from unaware and aware individuals is p_U and p_A, respectively, with $p_U > p_A$. Awareness of infection is obtained at the testing rate τ.

The mathematical equations for the basic model are given by

$$\begin{aligned}
\frac{dS}{dt} &= m_S - Sr\frac{p_U U + p_A A}{S+U+A} - e_S S, & S(0) &= S_0,\\
\frac{dU}{dt} &= m_U + Sr\frac{p_U U + p_A A}{S+U+A} - (e_U+\tau)U, & U(0) &= U_0,\\
\frac{dA}{dt} &= m_A + \tau U - e_A A, & A(0) &= A_0.
\end{aligned} \tag{4.1}$$

We simplify the above by assuming a constant population size $N = S+U+A$. This is a very natural assumption considering the time-span of only a few years that we are interested in when evaluating the initial impact of GCO. Hence, we set $e = e_U = e_A = e_S$, $m_U = m_A = 0$ and $m_S = e(S+U+A)$. This way e acts as a renewal or turnover rate at that accounts for sexually active individuals. Note that this assumption slightly reduces the effectiveness of intervention due to wash-out of infected individuals. This leads to the following system of ODES,

$$\begin{aligned}
\frac{dS}{dt} &= e(U+A) - Sr\frac{p_U U + p_A A}{N}, & S(0) &= S_0,\\
\frac{dU}{dt} &= Sr\frac{p_U U + p_A A}{N} - (e+\tau)U, & U(0) &= U_0,\\
\frac{dA}{dt} &= \tau U - eA, & A(0) &= A_0,\\
\frac{d\text{NEWINFECTED}}{dt} &= Sr\frac{p_U U + p_A A}{N}, & \text{NEWINFECTED}(0) &= 0,\\
\frac{d\text{NEWDETECTED}}{dt} &= \tau U, & \text{NEWDETECTED}(0) &= 0.
\end{aligned} \tag{4.2}$$

We added the last two equations to account for the number of new infections and newly detected infections.

4.5 Parameterizing the basic model

4.5.1 The rate of risky events

We define the rate of risky events r as the number of sex acts where HIV may be transmitted per unit time. We obtain this parameter value from the grid data, where participants were asked detailed questions about their five most recent relationships in the past six months. For each relationship the participants provided the date of first and last sex, whether the relationship was ongoing, and the average frequency of sex in the past year. From this information we estimate the total number of sex acts, as the product of the number of days of the relationship in the last year and the frequency of sex. Since the provided responses to either of these quantities are often vague or approximations we estimate a generous lower and upper bound for each relationship.

Some participants reported an unreasonably high number of sex acts, even for long-term relationships, so we set the maximum number of sex acts to once per day. Similarly, some participants did not report any sex act with their sex partners, so we set the minimum number of sex acts to once per relationship, see Figure 4.3. For each participant and each survey time point we then sum the total number of all risky events over all their relationships in the past year, and average those numbers to obtain the final estimate $r = 64.5\,\text{year}^{-1}$. When a range on the length of the relationship and the number of risky events was provided we estimate a generous lower bound of $r = 38.1\,\text{year}^{-1}$ and upper bound of $r = 104.7\,\text{year}^{-1}$.

4.5.2 The per-encounter transmission probability

The likelihood of transmitting the infection depends on the type of sex acts that the partners engage in [85, 89, 107]. From the network grid we know which sexual acts the participants engaged in the most recent time they had sex with each partner, and whether a condom was used. We assume that the most recent sex is representative of the typical sex act for this relationship.

In 79% of the cases more than one type of sex act was performed at the most recent sexual encounter. Further, in more than 97% of the cases the partners engaged in at least one of the following risky acts with a substantial risk of HIV

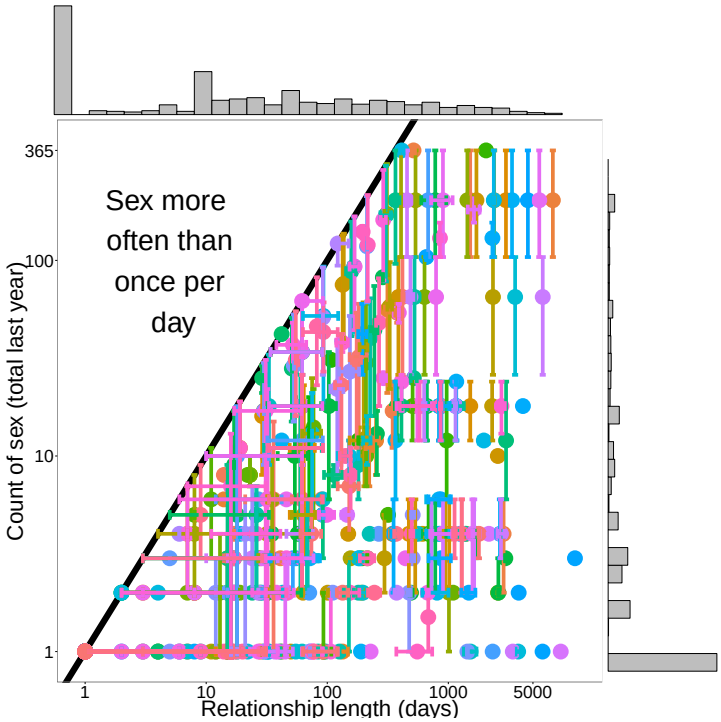


Figure 4.3: Relationship length and total number of risky events as obtained by the grid data. The horizontal and vertical error bars indicate the uncertainty in the length of the relationship and number of risky events, respectively. We chose to bound the number of risky events by once per day. The marginal histograms show that a large fraction of risk events stem from one-time sexual encounters.

transmission: receptive anal sex (with or without condom), insertive anal sex (with or without condom), receptive oral sex, insertive oral sex. A small number of participants reported receptive or insertive anal sex with and without condom at the same encounter, or that the condom came off at some point during the act. In this case we only consider the more risky version without a condom. However, participants may engage in both receptive and insertive anal intercourse, with or without receptive or insertive oral intercourse at the same encounter, see Figure 4.4.

Among the remaining 3% of cases the reported sex acts were rimming, fisting or masturbation, which are not considered to be independent risk factors, and have a negligibly small risk of HIV transmission. We discarded those events in our risk

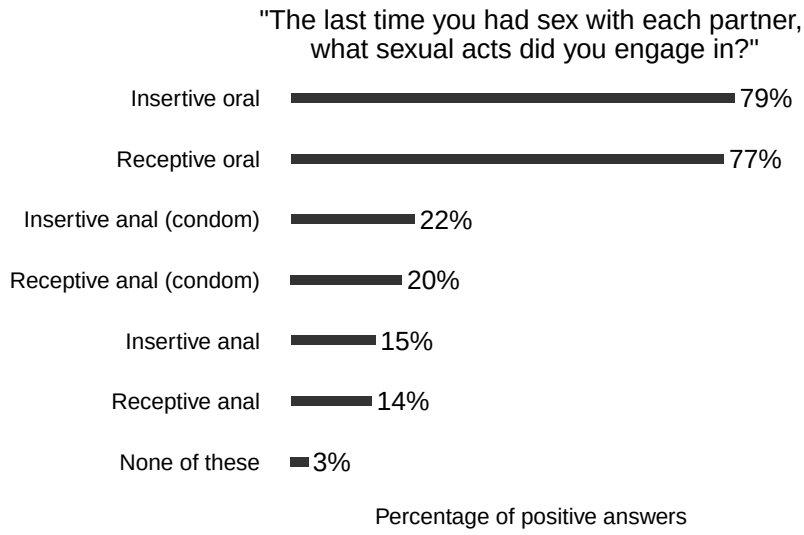


Figure 4.4: Responses to the network grid question *"The last time you had sex with each partner, what sexual acts did you engage in?"*. We use these frequencies together with the biological per-exposure probabilities from Table 4.1 to calculate the per-encounter risk of HIV transmission.

analysis.

To calculate the per-encounter risk of infection in the presence of more than one sex act per encounter, we assume that the likelihood of infection from all risk factors are independent, to find

$$\begin{aligned}
\mathbb{P}(\text{infected}) &= 1 - \mathbb{P}(\text{not infected}) \\
&= 1 - \prod_{x \in \text{sex act}} \mathbb{P}(\text{not infected by } x) \\
&= 1 - \prod_{x \in \text{sex act}} (1 - \mathbb{P}(\text{infected by } x)) \\
&= 1 - \prod_{x \in \text{sex act}} (1 - \mathbb{P}(\text{infected} \,|\, \text{engaged in } x) \times \mathbb{P}(\text{engaged in } x)).
\end{aligned}$$

The likelihood of engaging in the sex act is obtained from the grid data as described above, while the probability of infection for a given sex act is a biological parameter and can be obtained from the literature, see Table 4.1. From this data we then calculate, for each participant and study date, the per-encounter risk of HIV infec-

tion as the average of the probability of infection for a typical sexual encounter for each partner, weighted by the total number of sexual encounters with this partner. Finally, averaging over all participants and study dates we obtain the final average per-encounter risk of HIV transmission as $p_U = 0.397\%$. If we use the lower and upper bounds of the 95% confidence intervals from Table 4.1 we obtain a lower estimate of $p_U = 0.283\%$ and an upper estimate $p_U = 0.567\%$, respectively.

All the survey responses from the grid data are from HIV-negative participants who were recently tested, indicating uncertainty about their status. It is therefore reasonable to assume that they behave similarly to HIV-positive individuals that are not aware of their infection. HIV-positive individuals that are aware of their infection may reduce their rate of sexual encounters, or adjust their risk behaviour to reduce the risk of onwards transmission. For simplicity we assume that the rate of risky events r is unchanged by the knowledge of their HIV status, but that the per-encounter risk of HIV transmission from an individuals that is aware of their status p_A is reduced, e.g., by consistently using condoms. Under this assumption, p_A is calculated similarly as p_U with the only difference that for all sex acts we assume the minimal associated risk that is obtained by condom, that is, each unprotected receptive or insertive anal intercourse is replaced by receptive or insertive anal intercourse with condom, respectively. This results in a per-encounter probability of HIV infection of $p_A = 0.122\%$. Again using the 95% confidence intervals from Table 4.1 we obtain a lower estimate of $p_A = 0.087\%$ and an upper estimate $p_A = 0.175\%$, respectively.

Type of sex	**Per-exposure risk** (95% **CI)**	**Reference**
Receptive anal	1.38% (1.02%-1.86%)	[85]
Receptive anal (condom)	0.28% (0.20%-0.37%)	[89]
Insertive anal	0.11% (0.04%-0.28%)	[85]
Insertive anal (condom)	0.022% (0.008%-0.056%)	[89]
Receptive oral	0.028% (0.020%-0.037%)	[107]
Insertive oral	0.014% (0.010%-0.019%)	[107]

Table 4.1: Per-exposure probability of HIV infection for different risk behaviours. The use of condoms reduces the per-encounter risk of infection by 80% [89], while infection via receptive and insertive oral sex is estimated to be a factor of 50 and 100 times less likely than unprotected anal sex, respectively [107].

4.5.3 The testing rate

The testing rate τ, before the effect of GCO, is obtained from the survey data, more precisely, from the survey question

> "What best describes your HIV testing pattern?"

The aggregated answers from all participants on all study dates are shown in Figure 4.5. We identify two groups of testers, those with a regular testing interval and those who test irregularly and infrequently. We resist the temptation to put concrete numbers on the vague answers "After risky sex", "At new relationship" and "Other" (see Figure 4.5) as it is not clear if this group tests more or less frequent than those with a regular testing pattern. Taking the weighted average of the testing rates for the regular testers (every 3 months, every 6 months or every 12 months) we obtain a testing rate of 2.61 year^{-1}.

Since the data from our study was taken from individuals who just had an HIV test, we do not sample from the group that has never tested to date. In order to estimate the size and testing rate of this group we use data from the 2011 *Sex Now* survey [15], which reports that 18% of MSM never had an HIV test (31% of MSM under 30, 14% of MSM over 30). The fraction of MSM under 30 in the *Sex Now* study is 22% and the average age is 43 years, so if we assume that the average age of those under 30 is 25 we obtain an average age of 48 for those over 30. Further, assuming a minimum age of 15 years and that the random variable of the fraction of the population that had an HIV test by age $t+15$ years is exponentially distributed with parameter λ, we obtain the two estimates

$$\begin{aligned} \mathbb{P}(X \leq 10) &= 1 - e^{-10\lambda} = 0.69, \qquad \lambda = 0.117, \\ \mathbb{P}(X \leq 33) &= 1 - e^{-33\lambda} = 0.83, \qquad \lambda = 0.053. \end{aligned}$$

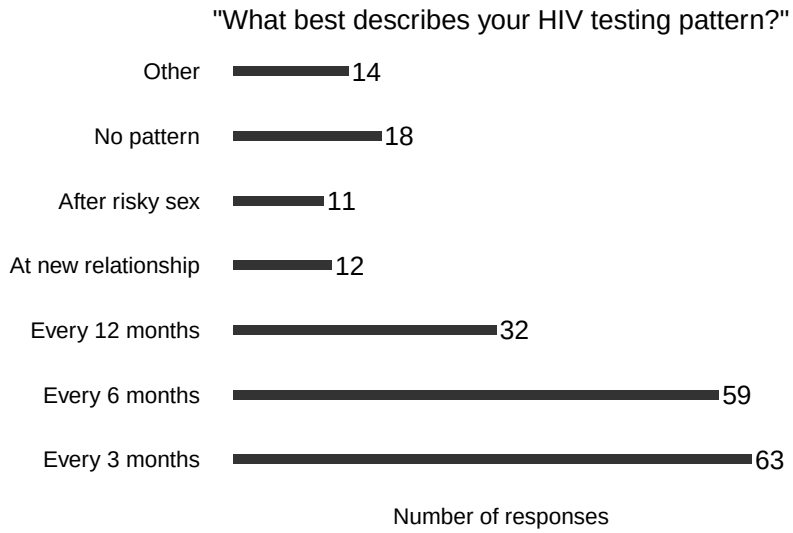


Figure 4.5: Responses to the online survey question *"What best describes your HIV testing pattern?"*. A large fraction of the population gets tested regularly, albeit at different rates, while a smaller fraction of the population has no regular testing pattern at all.

These estimates only differ by a factor of 2, so we take their average $\lambda = 0.085$ year^{-1} as the testing rate of the 18% of the population that is not represented by the participants of our survey.

Taking the weighted average of the two obtained testing rates for the regular testers and irregular testers we find a testing rate of $\tau = 2.16$ year^{-1}. We will use this as the baseline testing rate before the effect of GCO.

4.5.4 Other parameters from the literature

It is assumed that the total population size of the active Vancouver MSM community is about $20,000$, of which 18.1% are HIV infected and of those infected, 13.8% are unaware of their HIV infection (2.5% of the total population) [106]. Thus we choose the initial condition $U(0) = 20000 \times 0.025 = 500$, $A(0) = 20000 \times 0.181 - U(0) = 3120$, $S(0) = 20000 - U(0) - A(0) = 16380$. The turnover rate e is more difficult to estimate. We assume that the average time a MSM is sexually active with risk of HIV transmission is 20 years. An individual may leave the at-risk study

population by death, emigration from Vancouver, settling in a monogamous sero-concordant relationship, or by stopping to be sexually active altogether, among other things. This rough estimate yields $e = 1/20 = 0.05\,\text{year}^{-1}$.

Parameter	Interpretation	Value	Source
N	Total size of sexually active MSM population	20,000	[106]
$S(0)$	Initial number of uninfected MSM	16,380	[106]
$U(0)$	Initial number of HIV positive MSM unaware of their infection	500	[106]
$A(0)$	Initial number of HIV positive MSM aware of their infection	3,120	[106]
$1/e$	Average time of being sexual active and at-risk	20 years	text
τ	Average rate of HIV tests	$2.16\,\text{year}^{-1}$	[15], text
p_U	Average per-encounter risk of infection at encounter with unaware individual	0.397% 0.283-0.567	[85, 89, 107], text
p_A	Average per-encounter risk of infection at encounter with aware individual	0.122% 0.087-0.175	[85, 89, 107], text
r	Average number of risky encounters	$64.5\,\text{year}^{-1}$ 38.1-104.7	text

Table 4.2: Model parameters for the basic model, with 95% confidence interval or lower and upper bound, if applicable.

4.6 Sensitivity analysis - the Direct Differential method

For all our models we want to calculate how much the output variables change when the parameter values are slightly modified. Sensitivity analysis helps us to

translate uncertainties in parameter values to uncertainties in model predictions, and mainly serves two purposes: Firstly, it allows us to estimate how the effects of GCO impact the outcomes of interest, e.g., the total number of new infections. Secondly, sensitivity analysis allows us to identify those parameters that our model outcomes are most sensitive to, which, in turn, can guide the data acquisition and analysis: More care and focus should be put into getting good estimates for the parameters that impact our model predictions the most.

The straightforward way to calculate sensitivity is to repeatedly solve the system of differential equations for several parameter values and compare the results. However, when parameter changes are regarded in isolation, that is, only one parameter is varied at a time, an approximation called the *Direct Differential method* allows us to avoid repeated calculations. Instead of solving the same system many times, a slightly larger system is solved once. Let

$$\frac{d\vec{y}}{dt} = f(\vec{y},\vec{p},t), \qquad \vec{y}(0) = \vec{y}_0,$$

be an ordinary differential equation that depends on the parameters $\vec{p}$. We want to calculate how the model outcome $\vec{y}$ changes with each parameter p_j, i.e., we want to find $\vec{s}_j = \frac{\partial \vec{y}}{\partial p_j}$. To that end we set up a differential equation by taking the time-derivative:

$$\begin{aligned}
\frac{d}{dt}\vec{s}_j &= \frac{d}{dt}\frac{\partial \vec{y}}{\partial p_j} \\
&= \frac{\partial}{\partial p_j}\frac{d\vec{y}}{dt} \\
&= \frac{\partial}{\partial p_j} f(\vec{y},\vec{p},t) \\
&= \frac{\partial f}{\partial \vec{y}} \cdot \frac{\partial \vec{y}}{\partial p_j} + \frac{\partial f}{\partial p_j} \\
&= J\vec{s}_j + F_j,
\end{aligned}$$

where $J = \left(\frac{\partial f_i}{\partial \vec{y}_j}\right)_{ij}$ is the Jacobian matrix and the expression for $F_j = \begin{bmatrix} \frac{\partial f_1}{\partial p_j} \\ \vdots \\ \frac{\partial f_n}{\partial p_j} \end{bmatrix}$ can be calculated analytically. The initial condition is given by $\vec{s}_j(0) = \frac{\partial \vec{y}_j(0)}{\partial p_j}$, which is only non-zero when the initial value of the model depends on the parameters. Often the relative change $\vec{s}_j \times \frac{\vec{p}}{y_i}$is more informative than the absolute change, especially when parameters of different orders of magnitude are compared.

4.7 Results and predictions from the basic model

Here we present the results of and discuss the predictions from the basic model. We are interested in the immediate impact of GCO and hence run the model for five years only. Practical insights we could gain from studying the long-term behaviour of the model would be very limited, because the parameter values and model structure are only valid for the current state of the epidemic, and would need to be adjusted if the epidemic would grow or decrease considerably in the years to come. For this reason we also refrain from an extensive stability analysis. The model predictions, using the parameters from Section 4.5, are shown in Figure 4.6. The predicted total number of new infections and newly detected infections in the next five years is 1326 (499-3967) (265 (100-793) averaged annual) and 1666 (935-3971) (333 (197-794) averaged annual), respectively. These predictions are about twice as high as the 806 new infections found by the BCCDC from 2008-2012 [34], which we discuss below.

4.7.1 Calculating R_0

To validate our parameter values, we are interested in the basic reproduction number R_0 for the basic model. Using the next-generation method [23, 24, 49] we obtain the matrix of new infections F (not to be confused with F_j in the sensitivity analysis) and transfers of infections V, evaluated at the disease-free equilibrium $(S,U,A) = (N,0,0)$ as

$$F = \begin{bmatrix} rp_U & rp_A \\ 0 & 0 \end{bmatrix}, \qquad V = \begin{bmatrix} e+\tau & 0 \\ -\tau & e \end{bmatrix}.$$

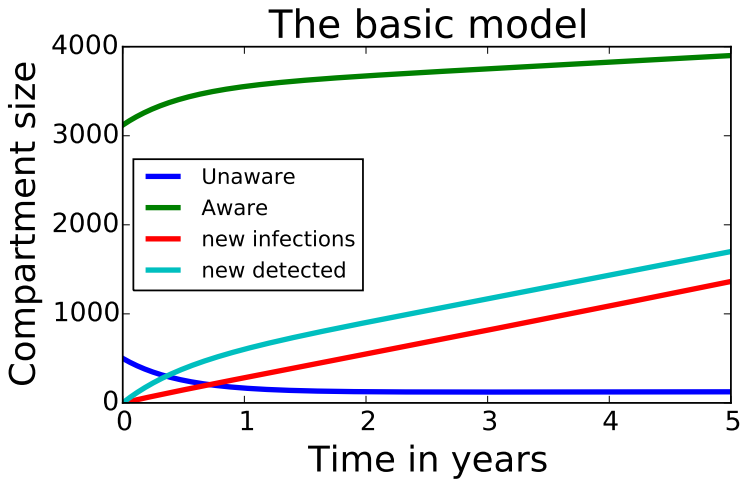


Figure 4.6: Evolution of all compartments for the next five years, using parameter values from Section 4.5. The total number of individuals is constant, hence the (roughly constant) number of susceptibles is not shown for simplicity. The basic model estimates a total of 1326 new infections and 1666 newly detected infections over five years.

Therefore, R_0 is the dominant eigenvalue of $G = FV^{-1}$, which is

$$R_0 = \frac{r}{e+\tau}\left(p_U + \frac{p_A \tau}{e}\right).$$

For the parameters obtained in Section 4.5 we find $R_0 = 1.65$ ($R_0 = 0.70$ - 3.87 for the lower and upper estimate of the contact rate r and per-encounter infection probabilities p_U and p_A, respectively). For a slow epidemic like HIV a value of R_0 close to unity means that the epidemic is relatively stable, which agrees with the observations that the number of new diagnosis of MSM in Vancouver did not change much over the last decade [34].

4.7.2 Sensitivity analysis

Applying the techniques of Section 4.6 to the basic model (4.2), we obtain that the model vector $\vec{y}$ and the parameter vector $\vec{p}$ are both 5-dimensional.

$$\vec{y} = \begin{bmatrix} S \\ U \\ A \\ \text{NEWINFECTED} \\ \text{NEWDETECTED} \end{bmatrix}, \; J = \begin{bmatrix} -r\frac{p_U U + p_A A}{N} & e - rp_U\frac{S}{N} & e - rp_A\frac{S}{N} & 0 & 0 \\ r\frac{p_U U + p_A A}{N} & rp_U\frac{S}{N} - (e+\tau) & rp_A\frac{S}{N} & 0 & 0 \\ 0 & \tau & -e & 0 & 0 \\ r\frac{p_U U + p_A A}{N} & rp_U\frac{S}{N} & rp_A\frac{S}{N} & 0 & 0 \\ 0 & \tau & 0 & 0 & 0 \end{bmatrix},$$

$$F_e = \begin{bmatrix} U+A \\ -U \\ -A \\ 0 \\ 0 \end{bmatrix}, \quad F_r = \begin{bmatrix} -S\frac{p_U U + p_A A}{N} \\ S\frac{p_U U + p_A A}{N} \\ 0 \\ S\frac{p_U U + p_A A}{N} \\ 0 \end{bmatrix},$$

$$F_{p_U} = \begin{bmatrix} -Sr\frac{U}{N} \\ Sr\frac{U}{N} \\ 0 \\ Sr\frac{U}{N} \\ 0 \end{bmatrix}, \quad F_{p_A} = \begin{bmatrix} -Sr\frac{A}{N} \\ Sr\frac{A}{N} \\ 0 \\ Sr\frac{A}{N} \\ 0 \end{bmatrix}, \quad F_\tau = \begin{bmatrix} 0 \\ -U \\ U \\ 0 \\ U \end{bmatrix}.$$

Here we used the subscript corresponding to the parameter value for simplicity of notation. The initial conditions do not depend on the parameter values and hence $\vec{s}_X(0) = 0$. This adds the following differential equations to (4.2):

$$\begin{aligned} \frac{d\vec{s}_e}{dt} &= Js_e + F_e, & s_e(0) &= 0, \\ \frac{d\vec{s}_r}{dt} &= Js_r + F_r, & s_r(0) &= 0, \\ \frac{d\vec{s}_{p_U}}{dt} &= Js_{p_U} + F_{p_U}, & s_{p_U}(0) &= 0, \\ \frac{d\vec{s}_{p_A}}{dt} &= Js_{p_A} + F_{p_A}, & s_{p_A}(0) &= 0, \\ \frac{d\vec{s}_\tau}{dt} &= Js_\tau + F_\tau, & s_\tau(0) &= 0. \end{aligned}$$

	e	r	p_U	p_A	τ
S	0.06	-0.09	-0.01	-0.08	0.01
U	-0.18	1.30	0.13	1.18	-1.06
A	-0.24	0.32	0.04	0.28	0.01
NEWINFECTED	-0.09	1.19	0.15	1.05	-0.09
NEWDETECTED	-0.08	0.84	0.11	0.73	0.03

Table 4.3: Sensitivity of the relative change in the outcome predictions of the basic model (4.2) with respect to its parameters. Positive numbers indicate that increasing the parameter leads to an increase in the final size of the corresponding compartment.

Using the parameters from Section 4.5 we numerically approximate the solution of (4.2) and calculate the sensitivity of the number of new infections to all parameter values, see Table 4.3. For example, the relative change in the predicted number of newly infected individuals with respect to the testing rate τ is -0.09. This means that increasing the testing rate by 25% leads to a relative change in NEWINFECTED by $-0.09 \times 0.25 = -0.0225$. With an estimated 265 (100-793) new infections annually, this corresponds to $0.0225 \times 265 = 6.0$ (2.3-17.8) averted HIV infections annually, or 29.8 (11.2-89.3) averted infections in five years.

4.7.3 Fitting to the epidemiological data

Up to this point we fitted each parameter individually to the available survey and grid data and computed the model predictions. This approach has the advantage that it accurately reflects the collected data as well as the consequences of the model assumptions. Although we obtain model predictions that are reasonably close to the epidemiological data, a gap remains, see Figure 4.6. This implies that some assumptions or effects have been missed or need to be reevaluated. We therefore revisit each parameter and compute the value that forces the model predictions to match the external epidemiological data. More precisely, we individually fit each parameter such that the predicted number of newly detected infections over five years matches the 806 newly detected infections by the BCCDC. We fit each parameter individually to gain an understanding of their individual impact, the reality is likely a combinations of these pure effects.

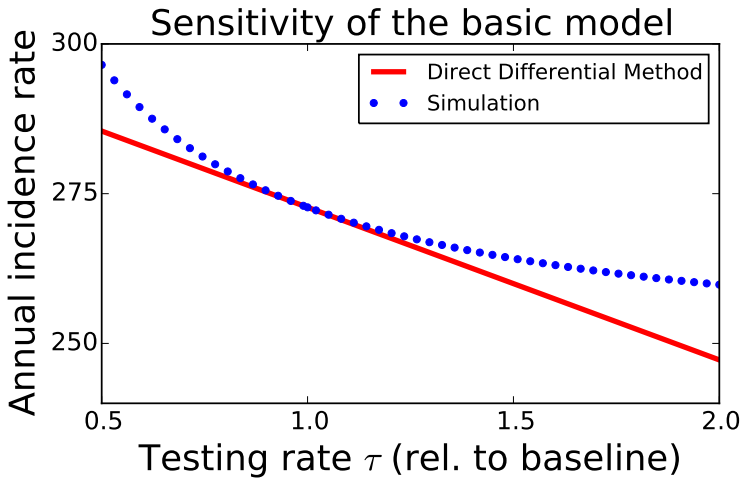


Figure 4.7: Sensitivity of the annual incidence rate as it depends on the testing rate τ, using the parameter values from Section 4.5. The linear approximation provided by the *Direct Differentiation method* is accurate for parameter values close to the initial value.

The emigration rate e would need to be increased by a factor of 11.5, from $0.05\,\text{year}^{-1}$ to $0.58\,\text{year}^{-1}$, which would mean that the average length of time of sexual activity is only $1/0.58 = 1.7$ years and implies for the basic reproduction number $R_0 = 0.20$. Instead, the contact rate r could be reduced by a factor of 3.3 from 64.5 to 19.7 risky annual sex events, which implies $R_0 = 0.51$. When varying the per-encounter risk of unaware individuals, p_U, the best fit is obtained for minimal values of p_U. At the smallest reasonable value, $p_U = p_A$, the model still predicts almost twice as many newly detected infections as desired (1552 instead of 806, $R_0 = 1.57$). On the other hand, a perfect fit can be found when reducing the per-encounter risk of aware individuals, p_A, by a factor of 4.0 from 0.122% to 0.031% ($R_0 = 0.50$). Next, a 12.9 fold reduction of the testing rate τ from $2.16\,\text{year}^{-1}$ to $0.17\,\text{year}^{-1}$, implying $R_0 = 2.39$ would also make the model predictions match BCCDC's number. Finally, decreasing the total size of the at-risk Vancouver MSM population by a factor of 2.07, which does not change $R_0 = 1.65$, also results in a perfect fit. Naturally, combinations of the adjustments above could also be considered.

4.8 Discussion of the basic model

The basic model predicts that the Vancouver MSM epidemic is going to be relatively stable for the next five years. At the current testing rate the number of individuals who are unaware of their infection is going to slowly decrease. This slowly reduces the annual incidence rate, and hence eventually leads to a slow decrease in the total number of HIV positive individuals.

From the sensitivity analysis we estimate the total number of new infections that are averted if GCO is successful at increasing the testing rate. The estimated 30 (11-89) averted infections in five years should be considered when analyzing the cost-effectiveness of GCO or similar programs. In fact, the sensitivity analysis provided here also allows a quick way to estimate the impact of other interventions that modify any of the model parameters, such as a program to reduce the per-encounter risk of infection, or the frequency of risky events.

The predicted number of newly detected infections in the next five years is about twice as high as the corresponding records in the previous five years, which implies that the basic model may not account for important effects, or that some of the collected behavioural data may be inaccurate. Hence we individually adjusted each model parameter to fit the observed size of the epidemic. For some parameters this results in values that are unlikely, but other findings suggest very reasonable adjustments: The reported rate of sexual encounters r may well be an overestimate. Further, our assumption that the main risk-reduction strategy of individuals who are aware of their infection is using condoms may be overly pessimistic. Indeed, adjusting risk behaviour, serosorting and treatment may all contribute to a larger reduction in the per-encounter risk of transmission p_A. In particular, a combination of the two effects above, reducing the number of sex partners and per-encounter risk, is very plausible and would be sufficient to explain the gap between the model predictions and epidemiological observations. Finally, the at-risk MSM population in Vancouver may be smaller than previously estimated, and reducing N by half would also be sufficient to close the gap.

The basic model assumes a very simplified view of the MSM epidemic in Vancouver. When exploring the survey and grid data it becomes apparent that risk and testing behaviour varies widely in the at-risk population. In the remainder of this

chapter we discuss models that stratify the population according to their risk and testing behaviour, resulting in more realistic models and (hopefully) more accurate predictions. Nevertheless, it is valuable to study the methodology and predictions from the basic model, as they will serve as a foundation for more sophisticated models.

4.9 The full model with heterogeneous testing patterns and risk behavior

The basic model (4.2) assumes that the Vancouver MSM population is homogeneous in their risk behaviour and testing pattern. However, when looking at the data in more detail it becomes apparent that reality is more complicated. Indeed, studying the observed testing patterns reveals that testers can be classified as frequent testers, regular testers, and non-testers. Equally importantly, the risk of infection can broadly be categorized as high and low risk. The existence of a core group or super-spreaders has been noticed and studied for HIV and other diseases in textbooks and the literature [7, 12, 24, 45, 90, 102, 118]. To reflect these observations we extend the basic model by compartmentalizing the total population into six groups, according to their risk behaviour and testing pattern, see Figure 4.8.

This stratification also implies that the effect of GCO can be modelled more realistically: By advertising and facilitating regular STI testing, GCO aims at encouraging at-risk individuals to adopt a regular testing pattern and to test frequently. Instead of modifying the individual testing rates of each compartment, we model the effect of GCO as increasing the likelihood that an individual who never tests is more likely to become a regular tester, and that regular testers are more likely to become frequent testers. As before, GCO does not affect the risk behavior.

There are two natural assumptions about how to model the contact mixing between the testing and risk groups in the full model. We assume that a person's choice of sexual partner is not directly influenced by their partner's testing pattern, but may be influenced by the partner's risk behaviour. Individuals in the same risk group may look for sex partners in similar venues or websites, and may have a similar preference for sexual acts, which may favour mixing within the same risk group. How much mixing between risk groups occurs is difficult to measure di-

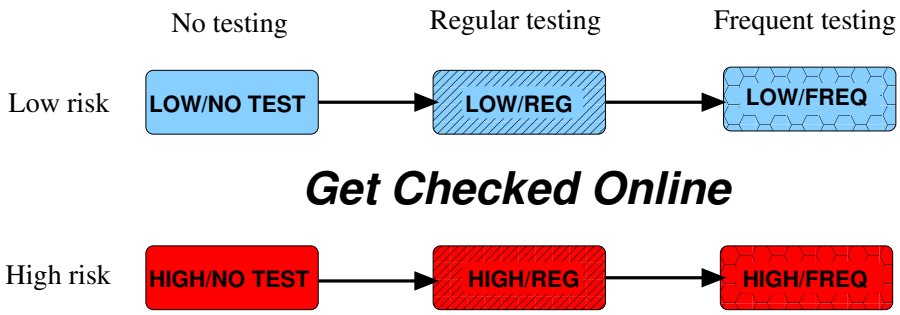


Figure 4.8: The full model accounts for two risk and three testing categories, and each resulting compartment is modeled using the basic model. By facilitating testing GCO encourages individuals to increase their testing frequency. We assume that GCO does not affect risk behavior.

rectly and can't be estimated from our data, so we assume both extreme cases to obtain a lower and upper estimate for the effect of contact mixing.

4.9.1 No contact mixing between risk groups

Here we assume that risky events only occur between individuals of the same risk category. After all, MSM in the same risk category may look for partners at similar venues or using similar other channels, and may share sexual preferences. This assumption leads to the following system of equations, for $i = 1,2$ and $j = 1,2,3$,

$$\begin{aligned}
\frac{dS_{ij}}{dt} &= e(U_{ij}+A_{ij}) - S_{ij}r_i\frac{p_U^{(i)}\sum_{k=1}^{3}U_{ik} + p_A^{(i)}\sum_{k=1}^{3}A_{ik}}{N_i} + \text{transfer}[S_{ij}],\\
\frac{dU_{ij}}{dt} &= S_{ij}r_i\frac{p_U^{(i)}\sum_{k=1}^{3}U_{ik} + p_A^{(i)}\sum_{k=1}^{3}A_{ik}}{N_i} - (e+\tau_j)U_{ij} + \text{transfer}[U_{ij}],\\
\frac{dA_{ij}}{dt} &= \tau_j U_{ij} - eA_{ij} + \text{transfer}[A_{ij}],\\
\frac{d\text{NEWINFECTED}}{dt} &= \sum_{i=1}^{2}\sum_{j=1}^{3} S_{ij}r_i\frac{p_U^{(i)}\sum_{k=1}^{3}U_{ik} + p_A^{(i)}\sum_{k=1}^{3}A_{ik}}{N_i},\\
\frac{d\text{NEWDETECTED}}{dt} &= \sum_{i=1}^{2}\sum_{j=1}^{3}\tau_j U_{ij}.
\end{aligned} \tag{4.3}$$

Here transfer$[X_{ij}]$ denotes the transfer from other subgroups, i.e.,

$$\begin{aligned}\text{transfer}[X_{ij}] = {} & h_{j-1,j}X_{i,j-1} + h_{j+1,j}X_{i,j+1} + v_{i-1,i}X_{i-1,j} + v_{i+1,i}X_{i+1,j} \\ & - (h_{j,j-1} + h_{j,j+1} + v_{i,i-1} + v_{i,i+1})X_{i,j},\end{aligned}$$

where $h_{k,\ell}$ is the rate of horizontal transition from the k-th to the ℓ-th column, and $v_{k,\ell}$ is the rate of vertical transition from the k-th to the ℓ-th row.

4.9.2 Full mixing between risk groups

Here we assume that individuals choose their sex partner independently of their or their partner's risk group. If two individuals from different risk groups mix it is reasonable to assume that the per-encounter risk of disease transmission $p_X^{(1,2)}$ is between the individual per-encounter risks, $p_X^{(1)} \le p_X^{(1,2)} \le p_X^{(2)}$. For simplicity we choose the arithmetic mean and hence define

$$p_U^{(i_1,i_2)} = \frac{p_U^{(i_1)} + p_U^{(i_2)}}{2}, \qquad p_A^{(i_1,i_2)} = \frac{p_A^{(i_1)} + p_A^{(i_2)}}{2}, \tag{4.4}$$

for $i_1, i_2 = 1,2$. With this notation we obtain the system of differential equations for the full model with between-risk-group mixing by replacing the mixing term in (4.3) with

$$Sr_i \frac{r_1\left(p_U^{(1,i)}\sum_{k=1}^3 U_{1k} + p_A^{(1,i)}\sum_{k=1}^3 A_{1k}\right) + r_2\left(p_U^{(2,i)}\sum_{k=1}^3 U_{2k} + p_A^{(2,i)}\sum_{k=1}^3 A_{2k}\right)}{r_1N_1 + r_2N_2}. \tag{4.5}$$

4.10 Parameterizing the full models from the survey and grid data

The full model requires additional parameters to those for the basic model. Instead of one average testing rate for the entire population we now consider three testing groups, each with their own testing rate. Similarly, the total average contact rate and per-encounter risk is now divided into a high risk and low risk group. We further must specify the transition rates between the various compartments, and extract suitable initial conditions. All other parameters can be carried over from

the basic model: The total population size, the total fraction of infected individuals that are aware or unaware of their infection, the population-turnover rate e and the biological per-exposure risk of HIV transmission.

4.10.1 Obtaining the testing rate τ and the horizontal transition rates

We divide the entire population into three categories according to their testing behaviour. The first group is those who never test, hence we have $\tau_1 = 0$. From Figure 4.5 we split the remaining population into regular testers, who test every 6-12 months (58.3%), and frequent testers who test every 3 months (41.7%). Taking the weighted average of the rates of those who test every 6-12 months we obtain $\tau_2 = 1.66\,\text{year}^{-1}$, while the rate for frequent testers is $4\,\text{year}^{-1}$.

To obtain the transition rate $h_{2,3}$ from the regular testers to the frequent testers we calculate the fraction of survey participants who change their testing pattern between the three different time points of the survey. Assuming a constant probability per unit time of switching from regular testing to frequent testing, leading to an exponential distribution waiting time for a switch, the fraction f of regular testers who were not frequent testers at time t satisfies

$$f = e^{-h_{2,3}t}, \qquad h_{2,3} = -\frac{\ln(f)}{t}. \tag{4.6}$$

Comparing between the three time points of the survey this gives us three estimates for $h_{2,3}$, $0.41\,\text{year}^{-1}$ (1st and 2nd time point), $0.26\,\text{year}^{-1}$ (2nd and 3rd time point), $0.20\,\text{year}^{-1}$ (1st and 3rd time point), which we average to obtain $h_{2,3} = 0.29\,\text{year}^{-1}$.

Similarly, for the rate $h_{3,2}$ at which frequent tester become regular testers, we calculate the three estimates $0.28\,\text{year}^{-1}$ (1st and 2nd time point), $0.27\,\text{year}^{-1}$ (2nd and 3rd time point) and $0.31\,\text{year}^{-1}$ (1st and 3rd time point), which we average to find $h_{3,2} = 0.28\,\text{year}^{-1}$.

Our model does not explicitly account for a direct transition between non-testers and frequent testers, we assume that individuals become regular testers for some amount of time before they transition between the extremes. To calculate the transition rates between the non-testing and regular testing group we set up a simple linear system of differential equations for the testing part of the full model.

Initially every individual starts in the never tested group N. Following the horizontal transition rules individuals in the regular testing group R are tested at rate τ_2 and individuals in the frequent testing group F are tested at rate τ_3. This leads to the following linear ODE:

$$\begin{aligned}
\frac{dN}{dt} &= h_{2,1}R - h_{1,2}N, & N(0) &= 1, \\
\frac{dR}{dt} &= h_{1,2}N + h_{3,2}F - (h_{2,1} + h_{2,3} + \tau_2)R, & R(0) &= 0, \\
\frac{dF}{dt} &= h_{2,3}R - (h_{32} + \tau_3)F, & F(0) &= 0, \\
\frac{dT}{dt} &= \tau_2 R + \tau_3 F, & T(0) &= 0.
\end{aligned} \tag{4.7}$$

From the *Sex Now* survey [15] we estimate, as in Section 4.5.3, that the average fraction of those who ever tested by age 25 and 48 is 69% and 83%, respectively. Again assuming a minimum age of 15 years this yields $T(10) = 0.69$ and $T(33) = 0.83$. We use the unknown transition parameters $h_{1,2}$ and $h_{2,1}$ to fit the solution of the ODE above with these constrains.

The fit is sensitive to the forward transition rate $h_{1,2}$, but insensitive to the backwards transition rate $h_{2,1}$. This allows us to estimate $h_{1,2} = 0.08\,\text{year}^{-1}$ with confidence, but $h_{2,1}$ is more difficult to estimate. A natural simplification would be to set $h_{2,1} = 0$ by arguing that the number of individuals who were regular testers and then stopped testing altogether is relatively small. Another natural assumption is to set $h_{2,1} = h_{1,2}$ and assume that individuals are just as likely to become regular testers as they are to loose the regular testing pattern. The squared difference between the model prediction and the real testing rates is almost identical for either choice of $h_{2,1}$, and we choose $h_{2,1} = h_{1,2} = 0.08\,\text{year}^{-1}$.

4.10.2 Obtaining the rate of risky events and the vertical transition rates

In order to split the population into a low-risk and high-risk group we first calculate, as in Section 4.5, the per-encounter risk as reported in the grid, p_U, for each participant and partner. A participant's total force of infection is then calculated as $\beta = \sum_{\text{partners}} r^{\text{partner}} p_U^{\text{partner}}$. Using a two-means clustering algorithm [48]

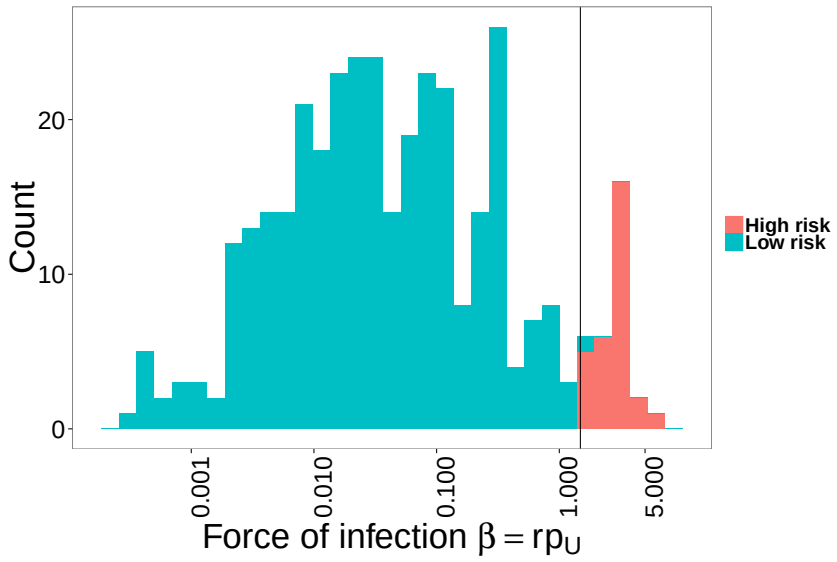


Figure 4.9: Per-participant force of infection $\beta = \sum_{\text{partners}} r^{\text{partner}} p_U^{\text{partner}}$. We use a two-means clustering algorithm to assign each individual into a low-risk or high-risk group according to their force of infection from this bimodal distribution. Notice the different orders of magnitude as highlighted by the logarithmic scale on the x axis.

we define the cut-off between the low and high risk group as the decision boundary $\beta = 1.485\,\text{year}^{-1}$, see Figure 4.9. This cutoff implies that 8.3% of the study population is classified as high risk. Averaging the corresponding risk parameters for each group yields for the low-risk group $p_U^{(1)} = 0.306\%$ (0.219%-0.445%), $p_A^{(1)} = 0.108\%$ (0.077%-0.155%), $r_1 = 51.7\,\text{year}^{-1}$ ($30.5\,\text{year}^{-1}$-$83.5\,\text{year}^{-1}$) and for the high-risk group $p_U^{(2)} = 1.391\%$ (1.013%-1.914%), $p_A^{(2)} = 0.286\%$ (0.209%-0.394%), $r_2 = 205.5\,\text{year}^{-1}$ ($122.7\,\text{year}^{-1}$-$337.7\,\text{year}^{-1}$). The ranges are obtained by using the 95% confidence intervals from Table 4.1.

With this we calculate the vertical transition rates between the risk groups in the same way as in Section 4.10.1 above, which yields $\nu_{1,2} = 0.137\,\text{year}^{-1}$ and $\nu_{2,1} = 2.373\,\text{year}^{-1}$.

Parameter	Interpretation	Value	Source
N	Total size of sexually	20,000	[106]

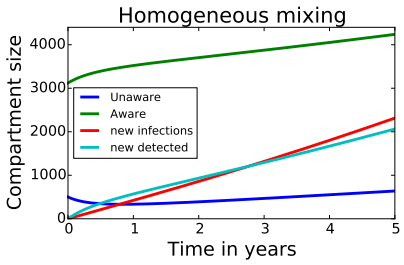
Parameter	Interpretation	Value	Source
	active MSM population		
$S(0)$	Initial number of uninfected MSM	16,380	[106]
$U(0)$	Initial number of HIV positive MSM unaware of their infection	500	[106]
$A(0)$	Initial number of HIV positive MSM aware of their infection	3,120	[106]
$f_{1,1}$	Initial fraction of MSM who are low risk and do not test	16.5%	[106], text
$f_{1,2}$	Initial fraction of MSM who are low risk and test regularly	43.8%	[106], text
$f_{1,1}$	Initial fraction of MSM who are low risk and test frequently	31.3%	[106], text
$f_{2,1}$	Initial fraction of MSM who are high risk and do not test	1.5%	[106], text
$f_{2,2}$	Initial fraction of MSM who are high risk and test regularly	4.0%	[106], text
$f_{2,3}$	Initial fraction of MSM who are high risk and test frequently	2.8%	[106], text
$1/e$	Average time of being sexual active and at-risk	20 years	text
τ_1	Average rate of HIV tests (non-testers)	$0\,\text{year}^{-1}$	text
τ_2	Average rate of HIV tests (regular testers)	$1.66\,\text{year}^{-1}$	text
τ_3	Average rate of HIV tests (frequent testers)	$4\,\text{year}^{-1}$	text
$h_{1,2}$	Average transition rate from non-tester to regular tester	$0.08\,\text{year}^{-1}$	[15], text
$h_{2,1}$	Average transition rate from regular tester to non-tester	$0.08\,\text{year}^{-1}$	[15], text

Parameter	Interpretation	Value	Source
$h_{2,3}$	Average transition rate from regular to frequent tester	$0.29\,\text{year}^{-1}$	text
$h_{3,2}$	Average transition rate from frequent to regular tester	$0.28\,\text{year}^{-1}$	text
r_1	Average number of risky encounters (low-risk group)	$51.7\,\text{year}^{-1}$ 30.5-83.5	[85, 89, 107], text
r_2	Average number of risky encounters (high-risk group)	$205.5\,\text{year}^{-1}$ 122.7-337.7	[85, 89, 107], text
$p_U^{(1)}$	Average per-encounter risk among unaware individual (low-risk group)	0.306% 0.219-0.445	[85, 89, 107], text
$p_U^{(2)}$	Average per-encounter risk among unaware individual (high-risk group)	1.391% 1.013-1.914	[85, 89, 107] text
$p_A^{(1)}$	Average per-encounter risk among aware individual (low-risk group)	0.108% 0.077-0.155	[85, 89, 107], text
$p_A^{(2)}$	Average per-encounter risk among aware individual (high-risk group)	0.286% 0.209-0.394	[85, 89, 107], text
$v_{1,2}$	Average transition rate from low to high risk	$0.137\,\text{year}^{-1}$	text
$v_{2,1}$	Average transition rate from high to low risk	$2.373\,\text{year}^{-1}$	text

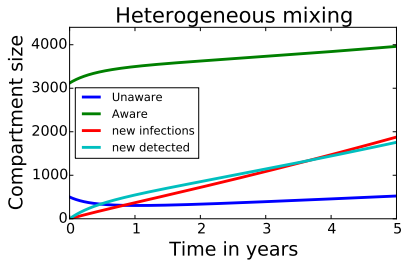
Table 4.4: Model parameters for the full model, with 95% confidence interval or lower and upper bound, if applicable.

4.11 Results

As discussed for the basic model, the focus of our modelling work is to estimate the immediate impact of GCO, so we again focus on the first five years of the model



(a) Estimates from the full model with homogeneous mixing, using the parameters from Section 4.10. The model predicts a total of 2314 new infections in the next five years.



(b) Estimates from the full model with homogeneous mixing, using the parameters from Section 4.10. The model predicts a total of 1878 new infections in the next five years.

Figure 4.10: Model estimates for the spread of HIV prevalence in the next five years, without GCO, for the full models with homogeneous and heterogeneous mixing.

predictions. Both mixing assumptions result in reasonable projections for the total number of new infections in the next five years, 2314 (384-7559) (463 (77-1512) average annual) for the model with homogeneous mixing and 1878 (333-6612) (376 (67-1322) average annual) for the model with heterogeneous mixing. The model predictions for the number of newly detected infections in the five year are 2063 (696-5844) (413 (139-1169) average annual) for the model with homogeneous mixing and 1757 (660-5081) (351 (132-1016) average annual) for the model with heterogeneous mixing. As with the basic model this is about twice as large as the 806 newly detected infections that the BCCDC reported between 2008-20012 [34].

4.11.1 The basic reproduction number R_0

Calculating the basic reproduction number for the full model is more complicated than for the basic model, because we obtain larger and more complicated matrices. We are able to derive the matrices of new infections F and transfers of infections V, but the analytical expression of the next generation matrix G and R_0 become too complicated to be useful. By exploiting the block structure of the matrices F and V we are able to simplify the required computations for R_0, which we ultimately

solve numerically.

Let $\mathbf{X} = (U_{11}, U_{12}, U_{13}, U_{21}, U_{22}, U_{23}, A_{11}, A_{12}, A_{13}, A_{21}, A_{22}, A_{23})$ be the vector of all infectious compartments. The matrix of transfers of infections is the same for both mixing assumptions. It can be written as a block matrix

$$V = \begin{bmatrix} -\Sigma - \Gamma + \Delta & \mathbf{0}_{6\times 6} \\ \Gamma & -\Sigma + \Delta \end{bmatrix}.$$

Here $\Sigma = eI_{6\times 6}$, $\Gamma = \mathrm{diag}(\tau_1, \tau_2, \tau_3, \tau_1, \tau_2, \tau_3)$ and Δ is the matrix that stems from the horizontal and vertical transfer terms,

$$\Delta = \begin{bmatrix} \Delta_{1,1} & \Delta_{1,2} \\ \Delta_{2,1} & \Delta_{2,2} \end{bmatrix},$$

where $\Delta_{1,2} = v_{21} I_{3\times 3}$, $\Delta_{2,1} = v_{12} I_{3\times 3}$,

$$\Delta_{1,1} = \begin{bmatrix} -h_{12} - v_{12} & h_{21} & 0 \\ h_{12} & -h_{21} - h_{23} - v_{12} & h_{32} \\ 0 & h_{23} & -h_{32} - v_{12} \end{bmatrix},$$

$$\Delta_{2,2} = \begin{bmatrix} -h_{12} - v_{21} & h_{21} & 0 \\ h_{12} & -h_{21} - h_{23} - v_{21} & h_{32} \\ 0 & h_{23} & -h_{32} - v_{21} \end{bmatrix}.$$

Similar to the calculations for the basic model no new infectious arrive in the aware compartments, so that the bottom half of F is all zeros. Hence the matrix of new infections has the form

$$F = \begin{bmatrix} F_{11} & F_{12} \\ \mathbf{0}_{6\times 6} & \mathbf{0}_{6\times 6} \end{bmatrix}.$$

For the homogeneous model we have

$$F_{11} = \begin{bmatrix} \frac{S_{11}r_1}{N_1}p_U^{(1)} & \frac{S_{11}r_1}{N_1}p_U^{(1)} & \frac{S_{11}r_1}{N_1}p_U^{(1)} & 0 & 0 & 0 \\ \frac{S_{12}r_1}{N_1}p_U^{(1)} & \frac{S_{12}r_1}{N_1}p_U^{(1)} & \frac{S_{12}r_1}{N_1}p_U^{(1)} & 0 & 0 & 0 \\ \frac{S_{13}r_1}{N_1}p_U^{(1)} & \frac{S_{13}r_1}{N_1}p_U^{(1)} & \frac{S_{13}r_1}{N_1}p_U^{(1)} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{S_{21}r_2}{N_2}p_U^{(2)} & \frac{S_{21}r_2}{N_2}p_U^{(2)} & \frac{S_{21}r_2}{N_2}p_U^{(2)} \\ 0 & 0 & 0 & \frac{S_{22}r_2}{N_2}p_U^{(2)} & \frac{S_{22}r_2}{N_2}p_U^{(2)} & \frac{S_{22}r_2}{N_2}p_U^{(2)} \\ 0 & 0 & 0 & \frac{S_{23}r_2}{N_2}p_U^{(2)} & \frac{S_{23}r_2}{N_2}p_U^{(2)} & \frac{S_{23}r_2}{N_2}p_U^{(2)} \end{bmatrix},$$

$$F_{12} = \begin{bmatrix} \frac{S_{11}r_1}{N_1}p_A^{(1)} & \frac{S_{11}r_1}{N_1}p_A^{(1)} & \frac{S_{11}r_1}{N_1}p_A^{(1)} & 0 & 0 & 0 \\ \frac{S_{12}r_1}{N_1}p_A^{(1)} & \frac{S_{12}r_1}{N_1}p_A^{(1)} & \frac{S_{12}r_1}{N_1}p_A^{(1)} & 0 & 0 & 0 \\ \frac{S_{13}r_1}{N_1}p_A^{(1)} & \frac{S_{13}r_1}{N_1}p_A^{(1)} & \frac{S_{13}r_1}{N_1}p_A^{(1)} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{S_{21}r_2}{N_2}p_A^{(2)} & \frac{S_{21}r_2}{N_2}p_A^{(2)} & \frac{S_{21}r_2}{N_2}p_A^{(2)} \\ 0 & 0 & 0 & \frac{S_{22}r_2}{N_2}p_A^{(2)} & \frac{S_{22}r_2}{N_2}p_A^{(2)} & \frac{S_{22}r_2}{N_2}p_A^{(2)} \\ 0 & 0 & 0 & \frac{S_{23}r_2}{N_2}p_A^{(2)} & \frac{S_{23}r_2}{N_2}p_A^{(2)} & \frac{S_{23}r_2}{N_2}p_A^{(2)} \end{bmatrix},$$

and for the model with heterogeneous mixing we have

$$F_{11} = \frac{1}{r_1N_1 + r_2N_2} \begin{bmatrix} S_{11}r_1^2p_U^{(1)} & S_{11}r_1^2p_U^{(1)} & S_{11}r_1^2p_U^{(1)} \\ S_{12}r_1^2p_U^{(1)} & S_{12}r_1^2p_U^{(1)} & S_{12}r_1^2p_U^{(1)} \\ S_{13}r_1^2p_U^{(1)} & S_{13}r_1^2p_U^{(1)} & S_{13}r_1^2p_U^{(1)} \\ S_{21}r_2r_1p_U^{(12)} & S_{21}r_2r_1p_U^{(12)} & S_{21}r_2r_1p_U^{(12)} \\ S_{22}r_2r_1p_U^{(12)} & S_{22}r_2r_1p_U^{(12)} & S_{22}r_2r_1p_U^{(12)} \\ S_{23}r_2r_1p_U^{(12)} & S_{23}r_2r_1p_U^{(12)} & S_{23}r_2r_1p_U^{(12)} \end{bmatrix}$$

$$\begin{bmatrix} S_{11}r_1r_2p_U^{(12)} & S_{11}r_1r_2p_U^{(12)} & S_{11}r_1r_2p_U^{(12)} \\ S_{12}r_1r_2p_U^{(12)} & S_{12}r_1r_2p_U^{(12)} & S_{12}r_1r_2p_U^{(12)} \\ S_{13}r_1r_2p_U^{(12)} & S_{13}r_1r_2p_U^{(12)} & S_{13}r_1r_2p_U^{(12)} \\ S_{21}{r_2}^2p_U^{(2)} & S_{21}{r_2}^2p_U^{(2)} & S_{21}{r_2}^2p_U^{(2)} \\ S_{22}{r_2}^2p_U^{(2)} & S_{22}{r_2}^2p_U^{(2)} & S_{22}{r_2}^2p_U^{(2)} \\ S_{23}{r_2}^2p_U^{(2)} & S_{23}{r_2}^2p_U^{(2)} & S_{23}{r_2}^2p_U^{(2)} \end{bmatrix},$$

$$F_{12} = \frac{1}{r_1 N_1 + r_2 N_2} \begin{bmatrix} S_{11} r_1^2 p_A^{(1)} & S_{11} r_1^2 p_A^{(1)} & S_{11} r_1^2 p_A^{(1)} \\ S_{12} r_1^2 p_A^{(1)} & S_{12} r_1^2 p_A^{(1)} & S_{12} r_1^2 p_A^{(1)} \\ S_{13} r_1^2 p_A^{(1)} & S_{13} r_1^2 p_A^{(1)} & S_{13} r_1^2 p_A^{(1)} \\ S_{21} r_2 r_1 p_A^{(12)} & S_{21} r_2 r_1 p_A^{(12)} & S_{21} r_2 r_1 p_A^{(12)} \\ S_{22} r_2 r_1 p_A^{(12)} & S_{22} r_2 r_1 p_A^{(12)} & S_{22} r_2 r_1 p_A^{(12)} \\ S_{23} r_2 r_1 p_A^{(12)} & S_{23} r_2 r_1 p_A^{(12)} & S_{23} r_2 r_1 p_A^{(12)} \end{bmatrix}$$
$$\begin{bmatrix} S_{11} r_1 r_2 p_A^{(12)} & S_{11} r_1 r_2 p_A^{(12)} & S_{11} r_1 r_2 p_A^{(12)} \\ S_{12} r_1 r_2 p_A^{(12)} & S_{12} r_1 r_2 p_A^{(12)} & S_{12} r_1 r_2 p_A^{(12)} \\ S_{13} r_1 r_2 p_A^{(12)} & S_{13} r_1 r_2 p_A^{(12)} & S_{13} r_1 r_2 p_A^{(12)} \\ S_{21} {r_2}^2 p_A^{(2)} & S_{21} {r_2}^2 p_A^{(2)} & S_{21} {r_2}^2 p_A^{(2)} \\ S_{22} {r_2}^2 p_A^{(2)} & S_{22} {r_2}^2 p_A^{(2)} & S_{22} {r_2}^2 p_A^{(2)} \\ S_{23} {r_2}^2 p_A^{(2)} & S_{23} {r_2}^2 p_A^{(2)} & S_{23} {r_2}^2 p_A^{(2)} \end{bmatrix}.$$

With this notation $\dot{\mathbf{X}} = (F+V)\mathbf{X}$ for either model and the next generation matrix $G = -FV^{-1}$ is quickly calculated to be

$$G = -\begin{bmatrix} (F_{11} - F_{12}(-\Sigma+\Delta)^{-1}\Gamma)(-\Sigma-\Gamma+\Delta)^{-1} & F_{12}(-\Sigma+\Delta)^{-1} \\ \mathbf{0}_{6\times 6} & \mathbf{0}_{6\times 6} \end{bmatrix}.$$

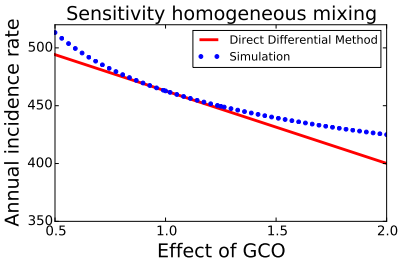
Note that all inverse matrices exist as the inverse of strictly diagonally dominant matrices. The set of eigenvalues of an upper triangular block matrix is identical to the set of eigenvalues of the diagonal blocks, hence the basic reproductive number R_0 is the dominant eigenvalue of the upper-left matrix,

$$R_0 = \max\left\{\sigma\left(-(F_{11} - F_{12}(-\Sigma+\Delta)^{-1}\Gamma)(-\Sigma-\Gamma+\Delta)^{-1}\right)\right\}. \tag{4.8}$$

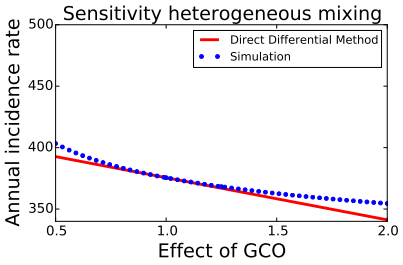
We calculate this expression numerically, which yields for the parameter values obtain in Section 4.10 a reproductive number of $R_0^{\text{hom}} = 2.73$ for the model with homogeneous mixing, and $R_0^{\text{het}} = 2.50$ for the model with heterogeneous mixing.

4.11.2 Sensitivity

In the basic model we model the effect of GCO as increasing the average population testing rate. In the full model we can be more specific: By facilitating testing



(a) Sensitivity of the annual incidence rate on the effect of GCO for the full model with homogeneous mixing.



(b) Sensitivity of the annual incidence rate on the effect of GCO for the full model heterogeneous mixing.

Figure 4.11: The effect of GCO is that it increases the horizontal forward transition rates and decreases the horizontal backward transition rates of the full model. This increasing the rate at which individuals become regular and frequent testers, and decreases the rate at which people switch to a less frequent testing behaviour.

and by setting automatic reminders GCO achieves its increase in the testing rate by encouraging individuals who are currently not testing regularly to pick up a regular testing pattern, and by encouraging individuals with a regular testing pattern become frequent testers. In our model this effect is reflected by the horizontal transition rates between the three testing compartments. We therefore model the effect of GCO as increasing the horizontal forward transition rates, and decreasing the horizontal backward transition rates by a certain factor.

It is possible to calculate the sensitivity by using the Direct Differential method, but this approach does not scale well with the number of equations and variables. Instead we calculate the sensitivity by running the same model repeatedly, varying one parameter value at a time, and comparing the outcome. We are interested in the total number of averted infections in the first five years, so we compare the model predictions of the baseline parameters with those from the model where the horizontal transition rates have been modified, see Figure 4.11.

If GCO manages to increase the horizontal forward transition rates by 25% and decrease the horizontal backward transition rate to 80%, the model with homogeneous mixing predicts that 13.3 (1.6-34.3) new HIV infections could be averted annually (66.4 (8.0-171.5) averted infections in five years), while the model with

heterogeneous mixing predicts a decrease by 7.3 (1.1-42.0) annual infections (36.4 (5.4-210.1) averted infections in five years).

4.12 Discussion and future work

In this chapter we use mathematical models to predict the impact of Get Checked Online (GCO), a novel program by the British Columbia Centre for Disease Control (BCCDC) to facilitate testing for sexually transmitted infections (STIS). We first use a basic model where we average all risk and testing behaviour parameters, which we then extend into a full model with various risk and testing groups, using two natural assumptions for the mixing between the risk groups. Our models are based on parameters that we extract from BCCDC data collected from HIV negative MSM who tested at a health Initiative for Men (HIM) sexual health clinic in Vancouver between 2011 and 2012. We give a detailed derivation of how the model parameters are extracted from the raw data set. The infection rate β is split into a rate of risky encounters and a per-encounter probability of infection, both of which are obtained independently from the data. Using a simple clustering algorithm we identify and parameterize a low risk and high risk group in the MSM population. The resulting average risk parameters differ by more than one order of magnitude, which has important consequences on the model predictions. For the testing rate we adjust for the fact that all study participants were recruited at a testing clinic.

We then calculate the basic reproductive number of the Vancouver HIV epidemic in the MSM community, which predicts that the epidemic is still growing, albeit rather slowly. Using sensitivity analysis we estimate the impact of GCO on the number of new infections in the first five years of the program. All models predict that even a modest increase in the testing rates would result in a valuable reduction of the number of new HIV infections.

We validate our models by comparing the predicted number of newly detected infection with those reported by the BCCDC in recent years. Our models consistently overestimate the reported numbers by a factor of 2, which indicates that either BCCDC's numbers underestimate the size of the epidemic, or that our model parameters do not accurately represent the entire active MSM community. We therefore fit the parameter values that we extracted from our data such that the model

predictions match the epidemiological observations. Our findings are that the difference in risk-behaviour between individuals who are unaware and aware of their infection be larger than we anticipated, that is, aware individuals apply effective risk-reduction strategies. An alternative explanation is that the size of the at-risk population may be smaller than previously estimated. Of course, in reality there may be a combination of these effects.

It is curious that the full model, which distinguished between two risk groups and three testing groups, predicts a larger number of infections and a larger basic reproductive number. This effect is highlighted when the contact mixing is limited to partners from the same risk groups. This suggests that the epidemic is driven by a small but very active high-risk group. Such a finding may have tremendous impact on how to fight the disease spread, in particular if this core group is difficult to identify and reach with interventions. This effect is well known in epidemiology and has been described in detail in many articles and standard textbooks, e.g., [7, 12, 45, 90, 102, 118]. For the Vancouver epidemic this result underscores the importance of identifying and reaching out to those at the highest risk of infection, which, if achieved, can have a great positive effect on the health of the entire population.

Our models are based on many assumptions and simplifications which offer room for future work. To begin with, a key model parameter, the number of sexually active MSM in Vancouver, is difficult to estimate and not very well known in the literature. More important than the total potential population size is the number of individuals who are sexually active and at risk, which is even less well defined.

Further, our ODE models do not account for the formation of partnerships. This key aspect of human relationships, and hence of the HIV transmission network, is extremely challenging to get meaningful data on. ODE models with partnership formation are not uncommon in the modelling of epidemics, see e.g., [24], but would quickly make the resulting system of differential equations too complicated to handle conveniently, and would most likely violate the underlying assumption of all ODE models that each subgroup should be "large".

A natural extension of the work we presented here is an agent-based model, where each agent represents a MSM that draws its parameters from those extracted from our dataset or others. This could naturally allow for partner formation, and

could easily account for a dependence of the rate of risky activities and the HIV testing rate. Such a model could also incorporate the effect of GCO in ways that are not possible with ODE models, such as the spread via word-of-mouth or by reaching out to a user's sexual partners.

Future work should also seek to incorporate data from HIV positive individuals, in particular how their risk behaviour changed when they learned about their infection. Further, the population could be split into more distinct groups, such as recently infected individuals with an increased risk of transmission, and individuals on treatment which a reduced risk of transmission.

Chapter 5

Assessing the optimal virulence of malaria-targeting mosquito pathogens: a mathematical study of engineered Metarhizium anisopliae

The content of this chapter and the corresponding Appendix C stem from a course project in *MATH 561: Mathematics of Infectious Diseases and Immunology* taught in 2011 at UBC, the results of which were published in *Malaria Journal* [66]. I am the lead author of the publication, which is joint authored with Michael Lindstrom, Anja Gumpinger, Jielin Zhu and Daniel Coombs. The journal article is distributed under the terms of the Creative Commons Attribution License (CC-BY) and has been modified slightly to fit the style and context of this thesis.

Abstract:

Background: *Metarhizium anisopliae* is a naturally occurring fungal pathogen of mosquitoes. Recently, *Metarhizium* has been engineered to act against malaria by directly killing the disease agent within mosquito vectors and also effectively

blocking onward transmission. It has been proposed that efforts should be made to minimize the virulence of the fungal pathogen, in order to slow the development of resistant mosquitoes following an actual deployment.

Results: Two mathematical models were developed and analysed to examine the efficacy of the fungal pathogen. It was found that, in many plausible scenarios, the best effects are achieved with a reduced or minimal pathogen virulence, even if the likelihood of resistance to the fungus is negligible. The results for both models depend on the interplay between two main effects: the ability of the fungus to reduce the mosquito population, and the ability of fungus-infected mosquitoes to compete for resources with non-fungus-infected mosquitoes.

Conclusions: The results indicate that there is no obvious choice of virulence for engineered *Metarhizium* or similar pathogens, and that all available information regarding the population ecology of the combined mosquito-fungus system should be carefully considered. The models provide a basic framework for examination of anti-malarial mosquito pathogens that should be extended and improved as new laboratory and field data become available.

5.1 Background

The major route of malaria transmission to humans is via blood feeding of female *Anopheles* mosquitoes (principally *Anopheles gambiae* and *Anopheles funestus*). Therefore, major efforts have been made to control mosquito populations in areas where malaria is prevalent. When first introduced, chemical insecticides were very efficient at reducing malaria prevalence in humans, although not without environmental damage. However, resistance has been observed to develop rapidly to broadly-used insecticides and there is a lack of new chemical agents [30, 105]. For this reason, fungal entomopathogens have been under extensive investigation as alternatives for mosquito control [11, 98]. Diverse fungal pathogens of mosquitoes exist in nature and additionally can be genetically modified to generate desirable properties.

The focus of this work is the application of engineered fungal pathogens of

mosquitoes that can neutralize or kill malarial sporozoites in the mosquito vector itself, preventing onward transmission to humans. This is motivated by the recent development of an engineered (*Metarhizium anisopliae*) fungus strain [31]. *Metarhizium* is a natural parasite of mosquitoes that infects through direct contact with the insect cuticle, and therefore is appropriate for control strategies based on local spraying indoors or baited traps. Recombinant strains of *Metarhizium* have been designed to (i) block attachment of malarial sporozoites to salivary glands of the mosquito; and (ii) neutralize or kill *Plasmodium falciparum* directly within the mosquito hemolymph. Under laboratory conditions, these engineered pathogens were found to substantially reduce sporozoite counts in the salivary glands of mosquitoes, compared to both fungus-uninfected mosquitoes and wild-type *Metarhizium* [31]. This raises the possibility of producing a biological agent which targets the malaria parasite within the mosquito, and is thus able to disrupt the transmission cycle and reduce the prevalence of malaria in humans.

Interestingly, because fungal pathogens such as *Metarhizium* do not kill infected mosquitoes until their later life-stages, after the majority of mosquito reproduction has occurred, it is believed that the selection pressure for mosquito resistance is quite low and therefore resistance should develop slowly even under widespread deployment [104]. Mosquito resistance to fungal biopesticides of this type has not been reported to date. Nonetheless, the concern of emerging resistance leads Fang *et al.* [31] to argue against engineering *Metarhizium* to kill mosquitoes faster. This argument stands in contrast to a strategy where fungal species are applied in conjunction with chemical pesticides to reduce the overall numbers of mosquitoes (discussed in [11]).

However, even in the absence of developing resistance, an interesting question remains about how to optimize *Metarhizium* or similar agents in terms of virulence against mosquitoes: should one expect a high virulence agent to outperform an alternative low-virulence strain? Indeed, if a high-virulence mosquito pathogen strain has worse performance than a low-virulence strain even in the absence of resistance, the threat of resistance additionally counts against it. (In this context, performance is measured in terms of reducing human malaria prevalence.) This chapter presents simple mathematical models to investigate this question. It will be shown that the optimal choice of virulence level is not obvious, and depends

quite sensitively on the details of the complex ecological system. In certain circumstances, it is expected to be preferable to apply a low-virulence agent which penetrates the mosquito population, and is thus able to reduce the prevalence of malaria parasites in mosquitoes and humans, even though the overall population level of mosquitoes is not strongly impacted. Under other circumstances, a high-virulence fungus that more effectively reduces the total mosquito population is expected to be preferable.

Mathematical modelling has been used in the study of malaria and as a tool for evaluating possible control strategies for over a century. The first mathematical models for malaria transmission were pioneered by Ross in 1911 [94] and developed further by Macdonald [73] and Anderson and May [7]. More recent mathematical models distinguish between exposed and infectious humans and mosquitoes reflect the fact that humans can become (temporarily) immune, treated or vaccinated, allow spatial heterogeneity, or include time-dependent parameters to account for environmental factors such as rainfall and humidity (see for example [3, 19, 64, 99]). The current work presents results from simple models for this new agent and potential improvements for future work are described.

5.2 Methods

Two prototype mathematical models are proposed that take into account the most fundamental properties of the malaria parasite, transmission between human host and mosquito vector, and the fungal mosquito pathogen. To derive the models in a simple form, it is necessary to make a number of simplifying assumptions. These assumptions could be relaxed in future versions of the model, at the expense of analytical and intuitive understanding of the model results. Some possibilities for this work are described in the conclusions.

5.2.1 Simplified model set-up and assumptions

1. The total human population, H, is taken to be constant. This is roughly equivalent to supposing that the population dynamics of mosquitoes, malaria parasites and fungal pathogens equilibrate rapidly compared to the human demographic timescale. The (time-dependent) fraction of humans that are

able to infect mosquitoes with the malaria parasite is denoted by $h(t)$ and these patients leave the infected group at rate ρ (for instance due to treatment, recovery or death). The human population is taken to be homogeneous in all other ways and the incubation period in the human host is assumed to be negligible compared to the duration of infectiousness.

2. The mosquito population is represented in three parts: uninfected (susceptible), infected with the malaria parasite, and fungus-infected. These three populations are denoted by $S(t)$, $I(t)$, and $F(t)$, respectively. It is necessary to make some simplifying assumptions concerning the population dynamics of mosquitoes, with and without the malaria parasite or the fungal pathogen. A commonly simplified model for biological populations is that of logistic growth [28]. This model has been applied in several previous works on malaria [18, 19, 35, 80]. Here, it is assumed that, in the absence of the malaria parasite and the fungus, the mosquito population experiences logistic growth with innate growth rate $\tilde{\kappa}$ and carrying capacity $\tilde{P}$. All mosquitoes also suffer a natural background death with rate denoted by μ, which is assumed to be independent of infection with the malaria parasite [97]. An alternative model of mosquito population dynamics, that includes larval and adult stages, is described below.

3. It is assumed that every mosquito has a human biting rate of β, irrespective of its infection status, and that if the human host is infectious, the bite always transmits the malaria parasite to the mosquito. On the other hand if an infected mosquito bites an uninfected human, the human will develop transmissible malaria with probability γ.

4. The fungal pathogen is constantly applied to the environment, and causes mosquitoes to transition from the S and I classes to the F class continuously, at constant rate α.

5. The fungal virulence (additional death of mosquitos due to fungal infection) is denoted by σ. To explicitly compare the natural mosquito death rate μ with the fungal virulence σ, one may rewrite the classical logistic growth

equation $dS/dt = \tilde{\kappa}S(1-S/\tilde{P})$ in the form presented below, where $\kappa = \tilde{\kappa}+\mu$ and $P = \tilde{P}\kappa/(\kappa-\mu)$.

6. The fungus is assumed to be perfectly, permanently and immediately effective at blocking malaria transmission to and from fungus-infected mosquitoes.

Under these assumptions, the following system of ordinary differential equations can be put forward (t denotes time):

$$\begin{aligned}
\frac{dh}{dt} &= \beta\gamma\frac{I}{H}(1-h) - \rho h, \\
\frac{dS}{dt} &= \kappa(S+I+F)\left(1-\frac{S+I+F}{P}\right) - \beta Sh - (\mu+\alpha)S, \\
\frac{dI}{dt} &= \beta Sh - (\mu+\alpha)I, \\
\frac{dF}{dt} &= \alpha(S+I) - (\mu+\sigma)F.
\end{aligned} \tag{5.1}$$

A model schematic is given in Figure 5.1.

5.2.2 Life-stage-structured mosquito model and assumptions

Recent field studies strongly indicate that mosquito populations are controlled via density-dependent regulation at the larval stage of development [100, 112]. The simplified model above was derived under the assumption that these effects can be captured via a logistic model for the adult mosquito population, since the larval population is not modelled explicitly. This simplification allows for fairly clean analytical results, but the biological system may have been oversimplified. Additionally, vertical transmission is a feature of certain fungal symbionts of insects [36]. Therefore, a possible improvement of the engineered fungus would be a vertically transmissible variant, that could be passed from mosquitoes to their offspring. To accommodate the additional realism of larval competition for resources, and to consider vertical transmission, an alternative model is now put forward. This model is built under the following set of additional assumptions:

1. Adult mosquitoes produce new larvae at a constant rate κ_c, irrespective of infections with fungus or malaria parasites. The offspring of fungal-infected

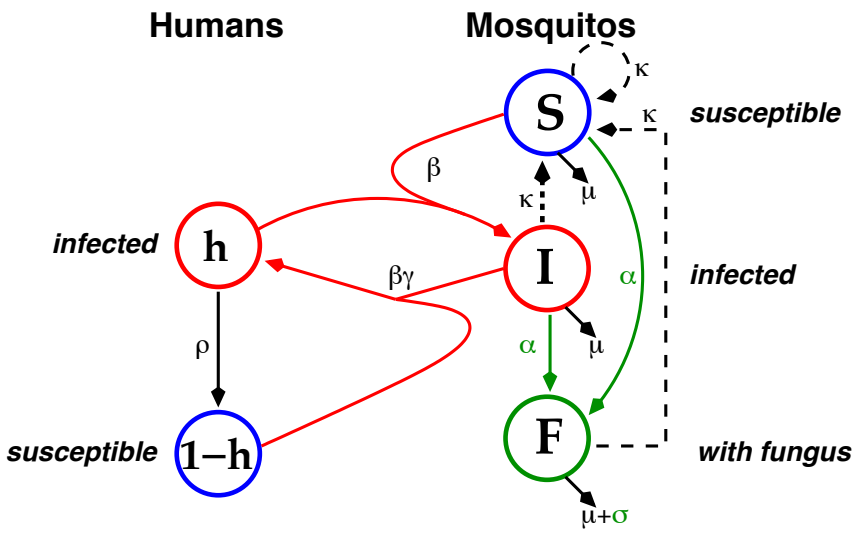


Figure 5.1: Simplified model. The fraction of infected humans in the population is given by h. Infection of susceptible mosquitoes S occurs at mass-action rate β while removal/recovery occurs at a rate ρ. Infected mosquitoes I infect susceptible humans with probability γ, given an encounter. Malaria infection does not significantly alter the mosquito death rate μ, but fungus infection, occurring with rate α reflecting the intensity of fungal application, increases the death rate by σ. Since neither malaria nor the fungus is transmitted vertically in the simplified model all new-born mosquitoes (indicated by κ) are susceptible.

mosquitoes are themselves infected with probability ξ, representing the probability of vertical transmission of the fungal pathogen. The populations of fungus-uninfected and fungus-infected larvae are denoted by $L_S(t)$ and $L_F(t)$ respectively. In the absence of information about larval infectibility, it is assumed that larvae are not directly infected by the fungus (although they may be infected via vertical transmission), and that the fungus has no detrimental effects in the larval stage.

2. Larvae compete for resources in a density-dependent manner, with the intensity of competition determined by the parameter κ_L.

3. Larvae mature into adult mosquitoes at rate m. In contrast to the simplified

model, adult mosquitoes do not directly compete for resources.

Under these additional assumptions, a new life-stage-structured model can be put forward:

$$\begin{aligned}
\frac{dh}{dt} &= \beta\gamma I\frac{(1-h)}{H} - \rho h,\\
\frac{dL_S}{dt} &= \kappa_c(S+I+(1-\xi)F) - \kappa_L(L_S+L_F)L_S - mL_S,\\
\frac{dL_F}{dt} &= \kappa_c\xi F - \kappa_L(L_S+L_F)L_F - mL_F,\\
\frac{dS}{dt} &= mL_S - \beta Sh - (\mu+\alpha)S,\\
\frac{dI}{dt} &= \beta Sh - (\mu+\alpha)I,\\
\frac{dF}{dt} &= mL_F + \alpha(S+I) - (\mu+\sigma)F.
\end{aligned} \tag{5.2}$$

A model schematic is given in Figure 5.2.

5.3 Results and Discussion

The goal is now to study the level of malaria in humans (h) as a function of the fungal parameters: the spraying rate α, fungal virulence σ and fungal vertical transmissibility ξ. For the simplified model, analytical results will be presented and applied to find the optimal values of α and σ in terms of reducing malaria in humans. For the life-stage-structured model, the situation is more complex and results will be shown only for particular numerical values of the parameters. In both models, it is found that application of a highly virulent fungus as biopesticide may not be the best strategy.

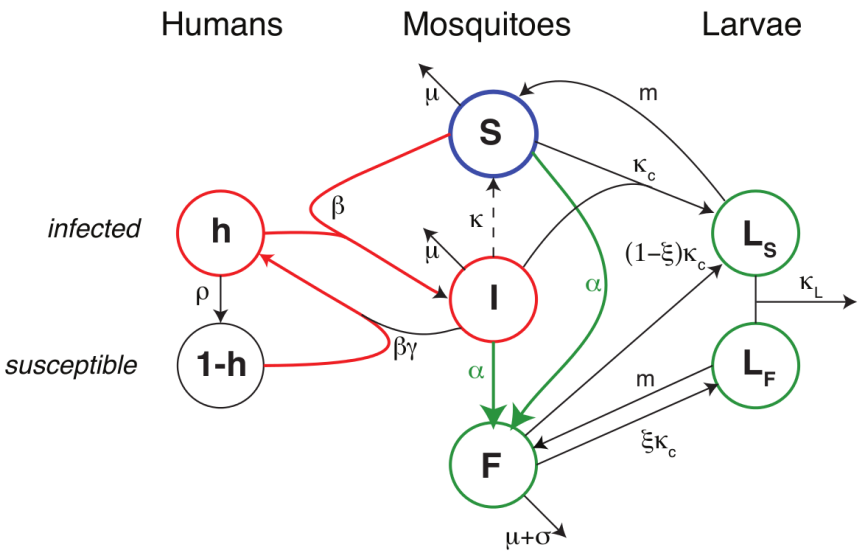


Figure 5.2: Life-stage structured model. In this model mosquitoes produce larvae at rate κ_c. Larvae may be fungal carriers (L_F) or fungus-free (L_S) and undergo density-dependent competition (see text) with intensity parameter κ_L. Larvae mature to produce adult mosquitoes at rate m. The parameter ξ determines the degree of vertical transmissibility of the fungal pathogen.

5.3.1 The simplified model with no fungus

If there is no fungus present in the system, the simplified model simplifies to the following:

$$\begin{aligned}
\frac{dh}{dt} &= \beta\gamma\frac{I}{H}(1-h)-\rho h,\\
\frac{dS}{dt} &= \kappa(S+I)\left(1-\frac{S+I}{P}\right)-\beta Sh-\mu S,\\
\frac{dI}{dt} &= \beta Sh-\mu I.
\end{aligned} \tag{5.3}$$

This system has three relevant equilibria: (1) the trivial equilibrium $(h,S,I) = (0,0,0)$ (no mosquitoes, no malaria); (2) the malaria-free equilibrium $(h,S,I) =$

$(0,(1-\frac{\mu}{\kappa})P,0)$ (mosquitoes but no malaria) and (3) the endemic equilibrium

$$\begin{aligned} h &= \frac{\beta^2\gamma P(\kappa-\mu)-H\kappa\mu\rho}{\beta(H\kappa\rho+\beta\gamma P(\kappa-\mu))}, \\ S &= \frac{\mu(H\kappa\rho+\beta\gamma P(\kappa-\mu))}{\beta\gamma\kappa(\beta+\mu)}, \\ I &= \frac{\beta^2\gamma P(\kappa-\mu)-H\kappa\mu\rho}{\beta\gamma\kappa(\beta+\mu)}. \end{aligned} \tag{5.4}$$

From the malaria-free equilibrium it is observed that $\kappa > \mu$ is the basic requirement for there to be mosquitoes in the system. This corresponds to the basic mosquito growth rate κ being large enough to outweigh mosquito death (and with no mosquitoes there can be no malaria).

The key quantity determining whether malaria will exist in the system is the basic reproduction number for the no-fungus system, $R_0^{(\mathrm{NF})} = \beta\sqrt{\gamma S_m/(\rho\mu H)}$, where $S_m = P(1-\frac{\mu}{\kappa})$ is the mosquito population at the malaria-free steady state. This quantity can be calculated using the next-generation method [24]. The square root occurs in this formula because there are two steps per rounds of transmission, mosquito to human to mosquito, and the next-generation R_0 is, by definition, a per-step quantity. The parameter β is not included in the square root because it determines the transmissibility human-vector and vector-human. It can also be shown that if $R_0^{(\mathrm{NF})} < 1$ then the malaria-free equilibrium is locally stable and the endemic equilibrium does not exist, while if $R_0^{(\mathrm{NF})} > 1$ then the endemic equilibrium exists and is locally stable.

5.3.2 Equilibrium analysis of the simplified model

The simplified model including the fungus (5.1) supports three distinctive equilibria, which can be defined as the *trivial*, *malaria-free* and *endemic* equilibria. The following results on the local stability of these equilibria can then be proven (see Appendix C for the proofs, which are lengthy but straightforward):

1. The trivial equilibrium $(h,S,I,F) = (0,0,0,0)$ is locally asymptotically stable if and only if $\kappa < \frac{(\mu+\alpha)(\mu+\sigma)}{\mu+\alpha+\sigma}$. This condition is analogous to the condition $\kappa > \mu$ for the model with no fungus, described above.

2. The malaria-free equilibrium

$$
\begin{aligned}
h_m &= 0, \\
S_m &= \frac{P(\mu+\sigma)}{\mu+\alpha+\sigma}\left(1-\frac{(\mu+\alpha)(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)}\right), \\
I_m &= 0, \\
F_m &= \frac{P\alpha}{\mu+\alpha+\sigma}\left(1-\frac{(\mu+\alpha)(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)}\right)
\end{aligned}
$$

exists in the positive plane and is locally asymptotically stable if and only if both of the following conditions are fulfilled:

$$
\kappa > \frac{(\mu+\alpha)(\mu+\sigma)}{\mu+\alpha+\sigma} \quad \text{and} \quad R_0 = \beta\sqrt{\frac{\gamma S_m}{\rho(\mu+\alpha)H}} < 1.
$$

The first condition corresponds to the survival of the mosquito population (exactly as for the trivial equilibrium). The second condition states that the basic reproductive number for malaria, R_0, must be subcritical.

3. The endemic equilibrium

$$
\begin{aligned}
h_e &= \frac{R_0^2-1}{R_0^2+\beta/(\mu+\alpha)}, \\
S_e &= \left(1-\frac{\beta}{\mu+\alpha+\beta}\frac{R_0^2-1}{R_0^2}\right)S_m, \\
I_e &= \frac{\beta}{\mu+\alpha+\beta}\frac{R_0^2-1}{R_0^2}S_m, \\
F_e &= F_m
\end{aligned}
$$

exists in the positive plane and is locally asymptotically stable if and only if both of the following conditions are fulfilled:

$$
\kappa > \frac{(\mu+\alpha)(\mu+\sigma)}{\mu+\alpha+\sigma} \quad \text{and} \quad R_0 = \beta\sqrt{\frac{\gamma S_m}{\rho(\mu+\alpha)H}} > 1.
$$

By comparing R_0 for the simplified model to the equivalent no-fungus quantity ($R_0^{(\mathrm{NF})}$), it is possible to see the role of the spraying rate α in reducing the reproductive number of the malaria parasite. The virulence of the fungus (parameter σ) does not affect R_0, because it is assumed that mosquitoes that are infected with the fungus are not able to transmit the malaria parasite.

5.3.3 Optimizing the virulence and application of the fungal pathogen in the simplified model

The two parameters of the model that are in principle under control in a real application are α (the application/spraying rate) and σ (the parameter controlling fungal virulence). Our goal is to choose α and σ such that the endemic malaria prevalence in humans

$$h_e = h_e(\alpha,\sigma) = \frac{R_0^2(\alpha,\sigma)-1}{R_0^2(\alpha,\sigma)+\beta/(\mu+\alpha)},$$

is minimized. The first and most intuitive result is that h_e is a monotonic decreasing function of α, which means that the more the mosquitoes are exposed to the fungus, the lower the endemic malaria prevalence is in humans. Increasing the fungus application rate is therefore always beneficial. This result is numerically illustrated in Figure 5.3. It can further be shown (see Appendix C) that the total number of mosquitoes $(S+I+F)$ also decreases when α increases, unless the fungal virulence σ is zero, in which case the total number of mosquitoes is independent of α.

There are two possible effects of continual fungus application that have to be considered in answering the question of how to find the optimal fungal virulence σ:

1. *The biopesticide effect:* Increasing the fungal virulence will lead to a decrease in the total number of mosquitoes $(S+I+F)$. This can be confirmed mathematically for the presented models (see Appendix C).

2. *The competition effect:* Fungus-infected mosquitoes F do not contribute directly to the transmission of the malaria parasite to humans. However, they do compete for resources with all the non-fungus-infected mosquitoes.

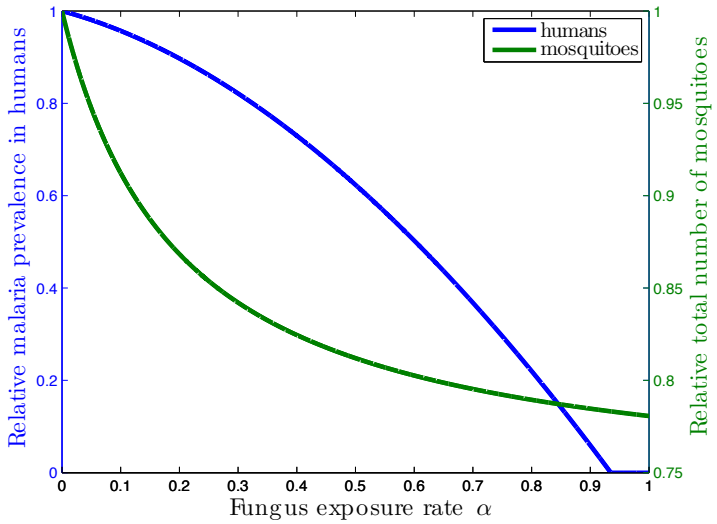


Figure 5.3: Prevalence of malaria in humans and total number of mosquitoes for varying fungus deployment rate. The steady state human malaria prevalence and the total mosquito population, both relative to baseline, are plotted against the fungus exposure intensity α, with fixed fungal pathogen virulence $\sigma = 0.1$ day^{-1}. As expected, quantities decrease when α increases. Note that, once α is large enough so that $R_0 < 1$ and malaria is eradicated, h_e remains constant at 0. Full details of all other chosen parameter values are given in Appendix C.

Therefore, if the fungal virulence is low, then the population will have a higher fraction of non-malaria-carrying mosquitoes and the force of infection of the malaria parasite on humans could plausibly be reduced.

Hence there is a tradeoff between the two mechanisms above: the total number of mosquitoes could be minimized by a high-virulence fungus (biopesticide effect), but what is critical for the prevalence in humans is the total number of mosquitoes infected with the malaria pathogen - and this might be minimized (via the competition effect) by a low-virulence fungus.

In order to determine the balance of these two effects, one must carefully analyse the mathematical model. It is then possible to establish a critical threshold for the relation between the mosquito innate growth rate κ and the natural mosquito

death rate μ that makes one or the other argument stronger. Specifically, it can be shown that the malaria prevalence in the human population $h_e(\alpha,\sigma)$ and the total number of infected mosquitoes $I_e(\alpha,\sigma)$ are both *maximized* if

$$\sigma = \sigma^* \equiv \frac{(\kappa - 2\mu)}{2 - \kappa/(\mu+\alpha)}.$$

This result indicates how to best design the fungus: To avoid the worst case $\sigma = \sigma^*$, the virulence σ must be chosen far away from σ^*. Note that while $\sigma = \sigma^*$ maximizes the malaria prevalence in humans as well as the total number of malaria-infected mosquitoes, it does not maximize or minimize the total number of mosquitoes. This result reveals the complexity of the tradeoff between the biopesticide and the competition effect, and indicates the importance of a good understanding of the mosquito population as well as the fungus-mosquito interaction, in designing the optimal pathogen for deployment.

5.3.4 Optimal virulence depends on background mosquito growth rate

Continuing the analysis of the previous section, two interesting cases can be distinguished: (i) If σ^* is negative, then σ is necessarily greater than σ^* and hence virulence should always be *maximized* to reduce malaria prevalence in humans (indicating that the biopesticide effect is stronger than the competition effect); (ii) If σ^* is positive, however, then a very large σ or a very low σ is desirable, but not an intermediate value. In the second case, this shows that the choice is either to use the fungal pathogen as a biopesticide (high virulence) or a distributed anti-*Plasmodium* agent (low virulence) in the mosquito population.

The choice depends on the relation between the mosquito growth rate κ, the mosquito death rate μ and the fungal exposure rate α. If the growth rate κ is small compared to the mosquito death rate μ (specifically, if $\kappa < 2\mu$) then $\sigma^* < 0$ for any value of α, and hence a larger fungal virulence σ is always desirable. This means that the fungus should be used as a biopesticide for slow-growing mosquito populations. On the other hand, if the mosquito population is relatively fast-growing ($\kappa > 2\mu$), then σ^* is negative for $0 < \alpha < (\kappa - 2\mu)/2$, while σ^* is positive for $\alpha > (\kappa - 2\mu)/2$. Parameter estimation (see Appendix C) yields that $\kappa \approx 5\mu$, im-

plying that in most real mosquito populations, the latter scenario is much more realistic. Hence, if the fungus can be applied at a sufficient rate, then minimizing the fungal virulence σ has the most beneficial effect of malaria prevalence in humans.

Figure 5.4 illustrates these principles for particular parameter choices. The human malaria prevalence $h_e(\alpha, \sigma)$ is plotted, relative to the baseline prevalence when no fungus is applied, and hence indicating how malaria management can be optimized. The left column shows the slow-growing result ($\kappa < 2\mu$) where the optimal strategy is to use the fungal pathogen as a biopesticide. The right column shows the more surprising result where, if the fungus deployment rate is sufficient, it is preferable to select a low-virulence fungus strain. In all cases, verifying the first result of this section, it can be seen that a higher fungus-exposure rate α is always desirable.

5.3.5 Equilibrium analysis of the life-stage-structured model

The investigation of the simplified model has shown that depending on the ecology of the mosquito population, it can be expected that there are circumstances where the optimal virulence of the fungal pathogen can be zero. However, competition between mosquitoes in the simplified model takes place in the adult stage, but recent evidence indicates that competition occurs at the larval stage. To examine this effect it is necessary to use the more complex, life-stage-structured model (5.2), which includes the larval stage. In this section the effects of vertical transmission of fungus will also be examined.

5.3.6 Life-stage-structured model without vertical transmission of fungus

By setting $\xi = 0$ in the second model, it is possible to investigate the model without vertical transmission. Similar to the simplified model, it can be shown that this model supports three easily explained equilibria, corresponding to (i) the trivial equilibrium with no mosquitoes and no malaria, (ii) a malaria-free equilibrium with mosquitoes but no malaria, and (iii) an endemic equilibrium state. Provided that the malaria-free equilibrium with mosquitoes exists, it can be shown that the

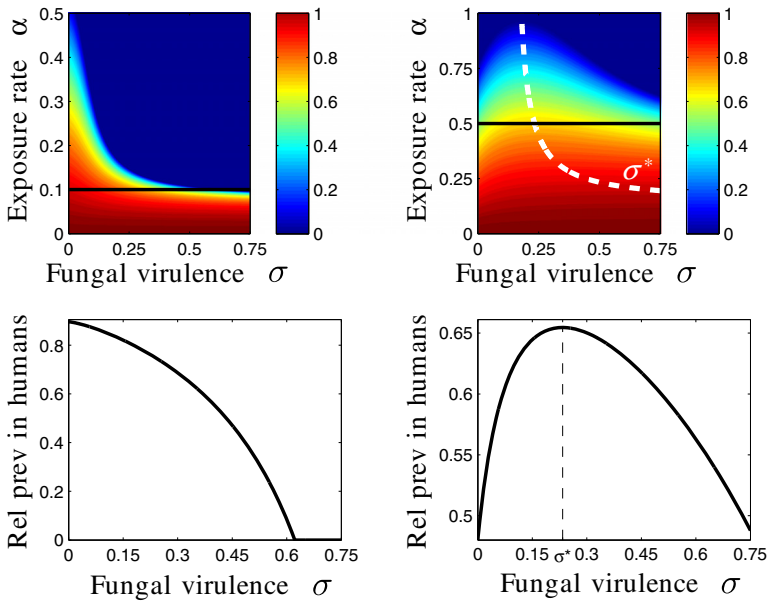


Figure 5.4: Malaria prevalence in humans varying over fungal pathogen virulence and deployment rate. The heat maps (top row) indicate the endemic malaria prevalence in humans h_e, relative to baseline where no fungus is applied, for the given parameters, the graphs (bottom row) show a one-dimensional projection of the heat maps, for a fixed fungal application (spraying) rate. Two distinct cases can be distinguished: (i) If $\kappa - 2\mu < 0$ (left column) then human malaria prevalence is a decreasing function of fungal virulence σ and deployment rate α. (ii) If $\kappa - 2\mu > 0$ (right column), the curve describing the worst-case fungal virulence σ^* is superimposed, indicating a non-monotonic relationship between human malaria prevalence and σ. Top row: The dark blue region indicates $R_0 < 1$ and hence $h_e = 0$. Bottom row: one-dimensional slice through the heat map above for a fixed value of $\alpha = 0.1$ day^{-1} (left) and $\alpha = 0.5$ day^{-1} (right), as indicated by the black line in the heat map above. Here, $\mu = 0.1$ day^{-1}, while $\kappa = 0.18$ day^{-1} (left) and $\kappa = 0.48$ day^{-1} (right). All other parameter values are as given in Appendix C.

endemic equilibrium state exists and is stable given the following reproductive number condition:

$$R_0 = \sqrt{\frac{\beta^2\gamma m^2((\kappa_c-\mu)(\mu+\alpha+\sigma)-\alpha\sigma)}{\kappa_L\rho(\mu+\alpha)^3(\mu+\sigma)H}} = \beta\sqrt{\frac{\gamma S_0}{\rho(\alpha+\mu)H}} > 1,$$

where

$$S_0 = \frac{m^2\left((\kappa_c-\mu)(\mu+\alpha+\sigma)-\alpha\sigma\right)}{\kappa_L(\mu+\alpha)^2(\mu+\sigma)}$$

is the equilibrium number of mosquitoes in the population in the absence of the malaria parasite. Under this condition, the fraction of the human population infected with malaria, at the endemic equilibrium, is given by the same function of R_0 as before (reflecting the very similar structures of the two models),

$$h(\sigma) = \frac{R_0^2-1}{R_0^2+\beta/(\mu+\alpha)}.$$

The derivative of this function with respect to σ is found to have no zeroes and is always negative for $R_0 > 1$. This indicates that the human prevalence of malaria is a strictly decreasing function of σ and therefore *the best strategy for the life-stage-structured model with no vertical transmission of fungus is to increase the fungal virulence as much as possible*. This result (illustrated in Figure 5.5, top left panel where $\xi = 0$) stands in contrast to the previous section where more nuanced conclusions were drawn, and is a consequence of the fact that the competition effect is occurring at the larval stage, while the biopesticide effect is occurring at the adult stage. Since it is adult mosquitoes, not larvae, that act as vectors for the malaria parasite, only the biopesticide effect can reduce malaria incidence in this version of the model.

5.3.7 Life-stage-structured model with vertical transmission of fungus

Vertical transmission is a feature of fungal symbionts of insects [36] and it seems natural to consider this in the context of the model. The life-stage-structured model,

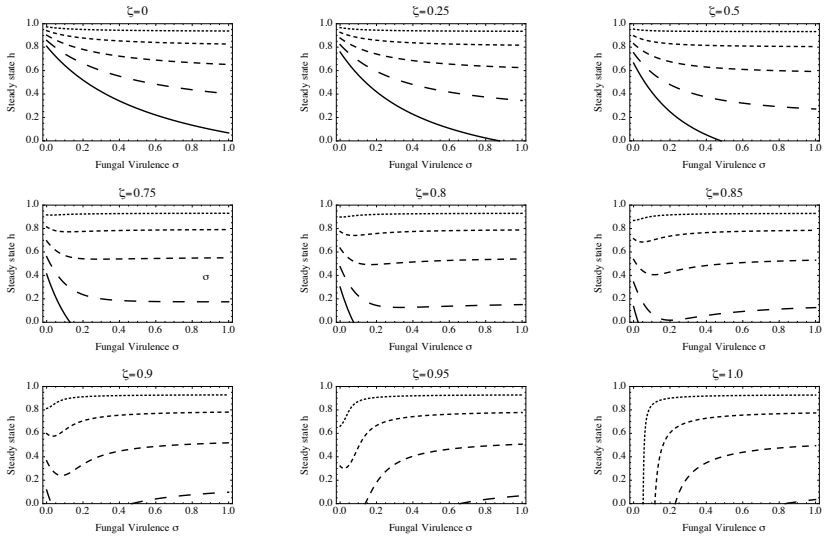


Figure 5.5: Malaria prevalence in humans varying over vertical transmission, deployment rate and fungal virulence. Each subfigure shows the endemic malaria prevalence in humans, relative to the no-fungus baseline, plotted against the fungal virulence σ, for five different levels of the fungal application (spraying) rate α. The different values of α in each subfigure are $\{0.05, 0.1, 0.15, 0.2, 0.25\}$ days^{-1}, corresponding to curves from top to bottom, and the length of the dashing increases with α. Each subfigure represents a different value of the vertical transmissibility ξ from $\xi = 0$ (no vertical transmission) to $\xi = 1$ (perfect vertical transmission). All other parameter values are as given in Appendix C.

where the parameter ξ is the vertical transmission fraction, reflects this possibility and allows the previous analysis to be repeated. In this case, the analytical expressions for the steady states of the full model can be obtained using a computer algebra system, but are too long to usefully give here. Instead we explore this scenario numerically, see Figure 5.5. The basic analytical results remain: there are three potential equilibria of the model, but only one has mosquitoes and malaria. Further, although the model is resistant to analytical exploration, it is easy to choose parameters and work numerically.

Figure 5.5 shows the equilibrium fraction of infected humans as a function of the fungal virulence σ and application rate α, across a range of possible vertical

transmission probabilities ξ. It is observed that if vertical transmission is unlikely then the benefit of the fungus is maximized by high virulence, but when vertical transmission becomes more likely, an intermediate or low virulence fungus would be favourable. In fact, in some cases a highly vertically transmissible fungus ($\xi \geq 0.85$ in Figure 5.5) is predicted to be able to eliminate malaria altogether, provided the virulence is low, but not too low. In all cases it should be noted that, as is to be expected, increasing the fungal application rate α is always beneficial.

These results indicate how the biopesticide effect of high virulence, acting at the adult stage, can be dominated by the competition effect, provided the adult and larval stages are coupled sufficiently strong via vertical transmission of the fungus. In this case, highly virulent fungus are found to be less effective because they prevent the fungus-infected mosquitoes from generating fungus-carrying larvae. This reduces competitive inhibition of fungus-uninfected larvae, thus reducing the overall effectiveness of the fungus in preventing malaria parasite infection of mosquitoes, and so ultimately increasing the prevalence of malaria in humans.

5.4 Conclusions

In this chapter, mathematical models have been employed to investigate the effectiveness of a fungal pathogen that blocks malaria transmission in mosquitoes to reduce malaria prevalence in humans. Unsurprisingly, all models indicate that malaria prevalence in humans could be reduced substantially by such a counter measure, and that the mosquito exposure rate to the fungus should be maximized to reduce malaria prevalence. However, the main interest of the results is to demonstrate that the optimal design of an agent that can simultaneously kill mosquitoes, *and malaria parasites within mosquitoes* depends quite sensitively on the details of a complex ecological system.

The first model, in which competition between mosquitoes occurred at the adult level, showed that in fast growing mosquito populations the fungal pathogen should be engineered to have low virulence. This result is independent of the possibility of mosquito resistance developing to the fungal pathogen and can be understood in very simple terms: fungus-infected mosquitoes do not directly contribute to the malaria epidemic, but competitively hamper the introduction/survival of new

susceptible mosquitoes. If mosquito resistance to fungal biopesticides arises in the field, this argument would additionally be strengthened.

In the second model, competition occurs at the larval stage of development. This model predicted that, in the absence of vertical transmission of the fungal pathogen, a highly virulent (biopesticidal) fungus would be desirable. However, the addition of reliable vertical transmission to this model significantly alters the predictions. If the fungal pathogen could be engineered in this way, then a high virulence would be highly detrimental to its efficacy as an anti-malarial strategy.

The key observation that should be drawn is that finding the optimal properties of agents that can reduce the malaria parasite incidence in mosquitoes is not an easy task, since several direct and indirect effects need to be parameterized and balanced, leading to conclusions that depend on the details of the mosquito population, fungal pathogen, and environment.

In order to obtain a straightforward analytical treatment, several simplifying assumptions were made throughout. It was assumed that the fungus could completely block onward malaria transmission. Modifying this assumption in the simplified model leads to an increased favourability of virulence, but the authors believe that this issue will best be resolved in a future model where the time dependence of the fungus-mosquito interaction is explicitly analysed. Also, by assuming a constant mass-action biting coefficient, human reactions to avoid mosquito bites (bed nets or indoor residual spraying) were neglected. This was done to study the pure effect of the fungus interaction, but is likely not a realistic representation of a real-life setting. Furthermore, the human population was modelled as a homogeneous population, neglecting for example co-infections, age structure, and previous malaria history. All of the above have an important impact on the malaria epidemic: co-infections increase the severity of each disease, children are much more vulnerable to malaria, and previous malaria infections can lead to temporal immunity. The incubation period of the malaria parasite and of the fungus in mosquitos were also neglected. This simplification particularly affects the mosquito population, where the incubation period is of about the same order as the life expectancy. Future work will include extending the current model to add more details of mosquito and malaria parasite life history in an age- and life-stage-dependent model of the mosquito.

Horizontal fungal transfer between mosquitoes was also not considered here. This effect has been observed for a number of fungal symbionts of insects and could play a role similar to vertical transmission in enhancing the effectiveness of the fungal pathogen. Horizontal transfer (possibly mediated through the environment) between mosquitoes would also be expected to increase the effectiveness of fungal spread, and reduce the necessary level of fungal application/spraying in the environment, which might otherwise have to be very intense. Further, in assuming a fairly simple logistic growth model, seasonal effects such as rainfall and humidity were neglected. These effects would lead to temporal variations in the mosquito population growth rate and carrying capacity, as well as unknown possible effects on the fungal pathogen. A final topic for possible future work is to consider spatial heterogeneities such as breeding sites and human habitat.

Despite these limitations, the models are useful in defining an argument for a minimal virulence of the antimalarial fungal pathogens. In future work, as field studies of *Metarhizium* or similar agents are completed, the parameterization of the model can be improved, and new models that allow insights into the potential of a large-scale deployment of such controls for malaria and other mosquito-borne diseases can be developed.

Chapter 6

Concluding remarks

Mathematical biology is a young and rapidly evolving field. At the same time, it is also very broad, using a wide range of mathematical techniques to answer an even wider range of theoretical and practical questions. The collaboration of all parts of the diverse community of mathematical biologists is necessary to fully unleash the enormous potential that lies in studying biological systems and phenomena with the rigor and thoroughness of the language of nature itself – mathematics.

Being a young and diverse field, however, also entails challenges. While there are many examples of fruitful collaborations between mathematical biologists and experimentalists (a few are highlighted in the introduction), the usefulness of mathematical modelling is not yet universally acknowledged. The author has always regarded preaching mathematical biology as part of being a mathematical biologist, and so my Ph.D. work and ultimately this thesis are aimed at proving the usefulness of mathematical biology by applying mathematical tools to obtain practical insights into current problems of global scale, HIV and malaria. The questions that led to the research presented in this thesis were, directly or indirectly, brought to us by collaborators in the public health sector and even mycologists, so non-mathematicians. Therefore the focus of this work lies in the modelling of the underlying biological system, as well as the interpretation and implications of the model predictions, instead of advancing the theory behind the mathematical tools that we are using.

Having said that, this thesis starts with a thorough yet application-focused introduction to continuous-time branching processes and their use in modelling within-host viral dynamics. We derive classical results from first principles, and combine techniques from various mathematical fields in new ways to express our model results in forms that can readily and efficiently be computed. In particular, with our approach the full probability density function can be computed efficiently, avoiding the necessity to approximate this distribution by sampling using the popular Gillespie algorithm. Our theoretical results extend to continuous-time multi-type branching processes, that can also be conditioned on non-extinction. This extension becomes very useful in our application to early within-host HIV infection in the first case study. Unfortunately, our technique is limited to branching processes, and is hence not applicable to a large class of non-linear processes.

In the first of three case studies we model the early dynamics of within-host HIV infection in order to estimate the length of the eclipse phase, that is, the length of time between exposure and detectability of infection. The uncertainty about this length and confidence in an early negative test is a source of anxiety for clinicians and patients alike. Therefore we collaborated with the British Columbia Centre for Disease Control (BCCDC) to develop mathematical models that can estimate the length of the eclipse phase, quantify the uncertainty of an early negative test and provide a guide for the timing of follow-up testing. Because of the stochastic nature of these clinical questions we chose a stochastic model, using the previously developed techniques for continuous-time multi-type branching processes. We highlight the importance of conditioning our models on non-extinction to match the inclusion criterion for how the data was collected.

We predict that the length of the eclipse phase may be slightly shorter than previously assumed, that the largest information gains are before day 10 after exposure, and that infection can practically be excluded if a test as short as three weeks after exposure is negative. To our knowledge these are the first rigorous results that quantify this uncertainty, and we hope that our work helps clinicians and patients to alleviate their anxieties and to help schedule follow-up tests. Reliable data on the early phase of HIV infection is extremely difficult to collect, so we chose the simplest model that fits the relatively little data that is available. As our

understanding of the early dynamics of HIV infection improves, and as more data becomes available, it would be worthwhile to add more features to our model and refine our results.

In our second case study we investigate the population-level impact of increasing the HIV testing in the Vancouver men who have sex with men (MSM) community. The hope is that by encouraging more testing more individuals become aware of their infection, and as a consequence reduce their risk-behaviour to avoid spreading the disease. It has long been observed by our collaborators at the BCCDC that MSM use the internet to find sex partners and to seek health information, so it is natural to use the internet to facilitate HIV testing and the delivery of results. This project, called Get Checked Online (GCO), has been developed for many years, and quantitative data on individual risk and testing behaviour has been collected. This data allows the development and parameterization of mathematical models, which in turn allows us to predict the potential impact of this intervention.

We start with a basic model of a homogeneous population, and show how the model parameters can be derived from the raw data. Since the goal of GCO is to increase the overall testing rate we use sensitivity analysis to estimate how much the model predictions depend on the testing rate, as well as other model parameters. We find that our model slightly overestimates the observed size of the epidemic, and suggest adjustments to the behavioural parameters to match the epidemiological observations. Looking at the data more closely reveals that the population is highly heterogeneous in their risk and testing behaviour, so we extend our basic model to reflect this observation. Our updated findings confirm a well-studied phenomenon in epidemiology that a relatively small high-risk group can make an epidemic much more severe than if this risk is distributed equally in the population. Our models are designed to lay the ground work for more sophisticated, agent-based models and to ultimately help to optimize the effect of GCO and calculate its cost-effectiveness. It has been thrilling to bring the insights from mathematical models to a long-term initiative that will improve the health of the subpopulation of Vancouver that is most burdened by the HIV epidemic.

Our third case study is motivated by a recent success in developing new anti-malarial agents. Instead of killing the mosquito vectors, this team of mycologists

successfully engineered a fungus that effectively blocks malaria transmission from infected mosquitoes to humans. Engineering such a fungus raises the question whether it would be detrimental or beneficial to also add virulence. Decreasing the total number of mosquitoes is beneficial in general, but the non-infectious mosquitoes also compete for resources with the infectious mosquitoes, so not removing them at an increased rate could have an even greater positive effect. Mathematical modelling is the ideal tool to predict the outcome of either strategy, and so we developed mathematical models to describe this non-linear system. We find that, under most conditions, the best outcomes are achieved by minimizing the virulence, even if resistance to the fungus is ignored. The next logical step in the development of the fungus is to test it outside of laboratory conditions in a field trial. In this case our models could be adjusted to the conditions of the trial, and it would be exciting to compare the model predictions with the observations in the field.

We hope that the work in this thesis is inspiring to mathematical biologists eager to apply their skills to real-world data and problems, as well as to non-mathematicians who might be skeptical about the value of the insights that mathematical models can provide. The author's work as a mathematical biologists has always been most exciting when working with messy real-world data sets and debating with non-mathematicians about the best tools and approaches to address those pressing current problems that started the collaboration in the first place. Time will tell if the models presented in this thesis will add to the long list of success stories of mathematical biology, but there is no doubt that this list is only going to increase as mathematical biology continues to grow and prove its usefulness in being a valuable tool to understand complex dynamics and worldwide diseases.

Bibliography

[1] M.-R. Abrahams, J. Anderson, E. Giorgi, C. Seoighe, K. Mlisana, L.-H. Ping, G. Athreya, F. Treurnicht, B. Keele, N. Wood, et al. Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. *J. Virol.*, 83(8):3556–3567, 2009. → pages 56, 62, 63

[2] M. Adelekea, C. Mafianaa, A. Idowua, S. Sam-Woboa, and O. Idowua. Population dynamics of indoor sampled mosquitoes and their implication in disease transmission in Abeokuta, south-western Nigeria. *J Vector Borne Dis*, 47:33–38, 2010. → pages xix, 196, 197, 198

[3] M. Al-Arydah and R. Smith? Controlling malaria with indoor residual spraying in spatially heterogeneous environments. *Math. Biosci. and Eng.*, 8(4):889–914, 2011. → pages 119

[4] M. Al-Arydah and R. Smith. Controlling malaria with indoor residual spraying in spatially heterogenous environments. *Mathematical biosciences and engineering: MBE*, 8(4):889–914, 2011. → pages 8

[5] T. M. Allen, L. Mortara, B. R. Mothé, M. Liebl, P. Jing, B. Calore, M. Piekarczyk, R. Ruddersdorf, D. H. O'Connor, X. Wang, et al. Tat-vaccinated macaques do not control simian immunodeficiency virus SIVmac239 replication. *J. Virol.*, 76(8):4108–4112, 2002. → pages 57

[6] R. Anderson, editor. *Population Dynamics of Infectious Diseases: Theory and Applications*. Chapman and Hall, London, 1982. → pages 8

[7] R. Anderson and R. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, London, 1991. → pages 100, 114, 119

[8] K. J. Bar, H. Li, A. Chamberland, C. Tremblay, J. P. Routy, T. Grayson, C. Sun, S. Wang, G. H. Learn, C. J. Morgan, et al. Wide variation in the

multiplicity of HIV-1 infection among injection drug users. *J. Virol.*, 84 (12):6241–6247, 2010. → pages 56, 62, 63

[9] B. Bayati, P. Chatelain, and P. Koumoutsakos. D-leaping: Accelerating stochastic simulation algorithms for reactions with delays. *Journal of Computational Physics*, 228(16):5908–5916, 2009. → pages 16

[10] C. Berticat, G. Boquien, M. Raymond, and C. Chevillon. Insecticide resistance genes induce a mating competition cost in Culex pipiens mosquitoes. *Genet Res*, 79(1):41–7, Feb 2002. → pages 8

[11] S. Blanford, B. H. K. Chan, N. Jenkins, D. Sim, R. J. Turner, A. F. Read, and M. B. Thomas. Fungal pathogen reduces potential for malaria transmission. *Science*, 308(5728):1638–41, Jun 2005. doi:10.1126/science.1108423. → pages 117, 118

[12] M. Boily, C. Lowndes, and M. Alary. The impact of HIV epidemic phases on the effectiveness of core group interventions: insights from mathematical models. *Sexually transmitted infections*, 78(suppl 1):i78–i90, 2002. → pages 100, 114

[13] M. Burgard, S. Blanche, C. Jasseron, P. Descamps, M.-C. Allemon, N. Ciraru-Vigneron, C. Floch, B. Heller-Roussin, E. Lachassinne, F. Mazy, et al. Performance of HIV-1 DNA or HIV-1 RNA tests for early diagnosis of perinatal HIV-1 infection during anti-retroviral prophylaxis. *J. Pediatrics*, 160(1):60–66, 2012. → pages 79

[14] A. M. Carias, S. McCoombe, M. McRaven, M. Anderson, N. Galloway, N. Vandergrift, A. J. Fought, J. Lurain, M. Duplantis, R. S. Veazey, et al. Defining the interaction of HIV-1 with the mucosal barriers of the female reproductive tract. *J. Virol.*, 87(21):11388–11400, 2013. → pages 63

[15] C. B. R. Centre. Sex Now-2010 Survey Report. 2011. → pages 80, 90, 92, 104, 106

[16] A. Chatterjee, K. Mayawala, J. S. Edwards, and D. G. Vlachos. Time accelerated Monte Carlo simulations of biological networks using the binomial τ-leap method. *Bioinformatics*, 21(9):2136–2137, 2005. → pages 16

[17] H. Y. Chen, M. Di Mascio, A. S. Perelson, D. D. Ho, and L. Zhang. Determination of virus burst size in vivo using a single-cycle SIV in rhesus macaques. *Proc. Natl. Acad. Sci USA*, 104(48):19079–19084, 2007. → pages 67

[18] N. Chitnis, J. M. Cushing, and J. M. Hyman. Bifurcation analysis of a mathematical model for malaria transmission. *SIAM J. Appl. Math.*, 67: 24–45, 2006. → pages 8, 120, 195, 196, 197

[19] N. Chitnis, J. M. Hyman, and J. M. Cushing. Determining important parameters in the spread of malaria through the sensitivity analysis of a mathematical model. *Bull Math Biol*, 70(5):1272–96, Jul 2008. doi:10.1007/s11538-008-9299-0. → pages 8, 119, 120, 195, 197

[20] M. S. Cohen, G. M. Shaw, A. J. McMichael, and B. F. Haynes. Acute HIV-1 infection. *New England Journal of Medicine*, 364(20):1943–1954, 2011. → pages xi, 6

[21] J. M. Conway and D. Coombs. A stochastic model of latently infected cell reactivation and viral blip generation in treated HIV patients. *PLoS Comp Biol*, 7(4):e1002033, 2011. → pages 12, 24, 66, 75

[22] J. M. Conway, B. P. Konrad, and D. Coombs. Stochastic analysis of pre-and postexposure prophylaxis against HIV infection. *SIAM J. Appl. Math.*, 73(2):904–928, 2013. → pages 12, 54, 66, 75

[23] O. Diekmann, J. Heesterbeek, and J. A. Metz. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of mathematical biology*, 28(4):365–382, 1990. → pages 94

[24] O. Diekmann, H. Heesterbeeck, and T. Britton. *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton University Press, 2012. → pages 94, 100, 114, 125

[25] N. M. Dixit, M. Markowitz, D. D. Ho, and A. S. Perelson. Estimates of intracellular delay and average drug efficacy from viral load data of HIV-infected individuals under antiretroviral therapy. *Antivir. Ther*, 9: 237–246, 2004. → pages 66

[26] R. M. Donovan, A. J. Sedgewick, J. R. Faeder, and D. M. Zuckerman. Efficient stochastic simulation of chemical kinetics networks using a weighted ensemble of trajectories. *The Journal of chemical physics*, 139 (11):115105, 2013. → pages 16

[27] J. L. Doob. Markoff chains–denumerable case. *Transactions of the American Mathematical Society*, pages 455–473, 1945. → pages 15

[28] L. Edelstein-Keshet. *Mathematical Models in Biology*. Society for Industrial and Applied Mathematics, 2nd edition, 2005. → pages 120

[29] P. Emau, Y. Jiang, M. B. Agy, B. Tian, G. Bekele, and C.-C. Tsai. Post-exposure prophylaxis for SIV revisited: animal model for HIV prevention. *AIDS Res Ther*, 3:29, 2006. → pages 5

[30] A. Enayati and J. Hemingway. Malaria management: past, present, and future. *Annu Rev Entomol*, 55:569–91, 2010. doi:10.1146/annurev-ento-112408-085423. → pages 117

[31] W. Fang, J. Vega-Rodríguez, A. K. Ghosh, M. Jacobs-Lorena, A. Kang, and R. J. St Leger. Development of transgenic fungi that kill human malaria parasites in mosquitoes. *Science*, 331(6020):1074–7, Feb 2011. doi:10.1126/science.1199115. → pages 8, 118

[32] E. W. Fiebig, D. J. Wright, B. D. Rawal, P. E. Garrett, R. T. Schumacher, L. Peddada, C. Heldebrant, R. Smith, A. Conrad, S. H. Kleinman, et al. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. *AIDS*, 17 (13):1871–1879, 2003. → pages xii, 56, 57, 78

[33] B. C. for Disease Control. STI/HIV Annual Report 2004. 2004. → pages 81

[34] B. C. for Disease Control. HIV in British Columbia: Annual Surveillance Report 2013. 2015. → pages 80, 94, 95, 108

[35] D. Gao and S. Ruan. A multi-patch malaria model with logistic growth populations. *SIAM J. Appl. Math.*, 72(3):819–841, 2012. → pages 120

[36] C. M. Gibson and M. S. Hunter. Extraordinarily widespread and fantastically complex: comparative biology of endosymbiotic bacterial and fungal mutualists of insects. *Ecol Lett*, 13(2):223–34, Feb 2010. doi:10.1111/j.1461-0248.2009.01416.x. → pages 121, 132

[37] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361, 1977. → pages 15

[38] D. T. Gillespie and L. R. Petzold. Improved leap-size selection for accelerated stochastic simulation. *The Journal of Chemical Physics*, 119 (16):8229–8234, 2003. → pages 16

[39] N. S. Goel and N. Richter-Dyn. *Stochastic models in biology*. Elsevier, 1974. → pages 11, 13, 23

[40] R. H. Gray, M. J. Wawer, R. Brookmeyer, N. K. Sewankambo, D. Serwadda, F. Wabwire-Mangen, T. Lutalo, X. Li, T. VanCott, and T. C. Quinn. Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda. *Lancet*, 357 (9263):1149–1153, 2001. → pages 62

[41] R. E. Haaland, P. A. Hawkins, J. Salazar-Gonzalez, A. Johnson, A. Tichacek, E. Karita, O. Manigart, J. Mulenga, B. F. Keele, G. M. Shaw, et al. Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. *PLoS Pathogens*, 5 (1):e1000274, 2009. → pages 56, 62, 63

[42] A. Haase. Population biology of HIV-1 infection: viral and CD4+ T cell demographics and dynamics in lymphatic tissues. *Ann. Rev. Immunol.*, 17 (1):625–656, 1999. → pages 79

[43] A. T. Haase. Targeting early infection to prevent HIV-1 mucosal transmission. *Nature*, 464(7286):217–223, 2010. → pages 78

[44] A. T. Haase. Early events in sexual transmission of HIV and SIV and opportunities for interventions. *Ann. Rev. Medicine*, 62:127–139, 2011. → pages 78, 79

[45] K. P. Hadeler and C. Castillo-Chávez. A core group model for disease transmission. *Mathematical biosciences*, 128(1):41–55, 1995. → pages 100, 114

[46] T. E. Harris. *The theory of branching processes*. Courier Corporation, 2002. → pages 11

[47] G. Harrison et al. Mosquitoes, malaria and man: A history of the hostilities since 1880. *Mosquitoes, malaria and man: a history of the hostilities since 1880.*, 1978. → pages 8

[48] J. A. Hartigan. *Clustering algorithms*. John Wiley & Sons, Inc., 1975. → pages 104

[49] J. Heffernan, R. Smith, and L. Wahl. Perspectives on the basic reproductive ratio. *Journal of the Royal Society Interface*, 2(4):281–293, 2005. → pages 94

[50] A. D. Herrera Reyes. Mathematical models for immunodeficiency virus : post-treatment and memory activation. Master's thesis, University of British Columbia, 2012. → pages 12

[51] A. D. Herrera Reyes and B. P. Konrad. Early HIV Infection. Final project MATH 607E: Topics in Numerical Analysis, 2010. → pages 12

[52] D. D. Ho, A. U. Neumann, A. S. Perelson, W. Chen, J. M. Leonard, M. Markowitz, et al. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature*, 373(6510):123–126, 1995. → pages 2

[53] S. T. Hottes, J. Farrell, M. Bondyra, D. Haag, J. Shoveller, and M. Gilbert. Internet-Based HIV and Sexually Transmitted Infection Testing in British Columbia, Canada: Opinions and Expectations of Prospective Clients. *J Med Internet Res*, 14(2):e41, Mar 2012. doi:10.2196/jmir.1948. URL http://www.jmir.org/2012/2/e41/. → pages xv, 82

[54] J. R. Ickovics, J. A. Druley, E. L. Grigorenko, A. C. Morrill, S. E. Beren, and J. Rodin. Long-term effects of HIV counseling and testing for women: Behavioral and psychological consequences are limited at 18 months posttest. *Health Psychology*, 17(5):395, 1998. → pages 58

[55] D. D. Jenkins and G. D. Peterson. Aess: accelerated exact stochastic simulation. *Computer Physics Communications*, 182(12):2580–2586, 2011. → pages 16

[56] R. John, T. Ephraim, and A. Andrew. Reduced susceptibility to pyrethroid insecticide treated nets by the malaria vector Anopheles gambiae s.l. in western Uganda. *Malaria Journal*, 7(1):92, 2008. ISSN 1475-2875. doi:10.1186/1475-2875-7-92. URL http://www.malariajournal.com/content/7/1/92. → pages 8

[57] H. A. Joseph, A. M. Fasula, R. L. Morgan, A. Stuckey, M. E. Alvarez, A. Margolis, D. Stratford, and S. W. Dooley Jr. "The Anticipation Alone could Kill You": Past and Potential Clients' Perspectives on HIV Testing in Non-Health Care Settings. *AIDS Education and Prevention*, 23(6): 577–594, 2011. → pages 58

[58] B. Kakkilaya. Ronald Ross. http://www.malariasite.com/ronald-ross/. [Online; accessed 22-June-2015]. → pages 3

[59] S. Karlin and H. Taylor. A first course in stochastic processes, 1975. → pages 11, 66

[60] S. Karlin and H. M. Taylor. *A first course in stochastic processes*. Academic press, 1975. → pages 11, 13

[61] B. F. Keele, E. E. Giorgi, J. F. Salazar-Gonzalez, J. M. Decker, K. T. Pham, M. G. Salazar, C. Sun, T. Grayson, S. Wang, H. Li, et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. USA*, 105(21):7552–7557, 2008. → pages 56, 62, 63

[62] B. F. Keele, H. Li, G. H. Learn, P. Hraber, E. E. Giorgi, T. Grayson, C. Sun, Y. Chen, W. W. Yeh, N. L. Letvin, et al. Low-dose rectal inoculation of rhesus macaques by SIVsmE660 or SIVmac251 recapitulates human mucosal infection by HIV-1. *J. Exp. Med.*, 206(5):1117–1134, 2009. → pages 56, 58, 62, 63

[63] S. H. Kleinman, N. Lelie, and M. P. Busch. Infectivity of human immunodeficiency virus-1, hepatitis C virus, and hepatitis B virus and risk of transmission by transfusion. *Transfusion*, 49(11):2454–2489, 2009. → pages 51, 154

[64] J. Koella and R. Antia. Epidemiological models for the spread of anti-malarial resistance. *Malaria Journal*, 2:3, 2003. → pages 8, 119, 196, 197, 198

[65] B. P. Konrad, R. J. Smith, and F. Lutscher. Gender differences in heterosexual transmission of HIV in urban and rural populations. In J. M. Tchuenche and Z. Mukandavire, editors, *Advances in disease epidemiology*. Nova Science Publishers Inc, New York, 2009. → pages 3

[66] B. P. Konrad, M. Lindstrom, A. Gumpinger, J. Zhu, and D. Coombs. Assessing the optimal virulence of malaria-targeting mosquito pathogens: a mathematical study of engineered Metarhizium anisopliae. *Malaria Journal*, 13(1):11, 2014. → pages iv, 116, 183

[67] M. Kot. *Elements of mathematical ecology*. Cambridge University Press, 2001. → pages 13, 19

[68] H. Kuwahara and I. Mura. An efficient and exact stochastic simulation method to analyze rare events in biochemical systems. *The Journal of chemical physics*, 129(16):165101, 2008. → pages 16

[69] R. Landovitz. Occupational and nonoccupational postexposure prophylaxis for HIV in 2009. *Topics in HIV medicine*, 17(3):104–108, 2008. → pages 57

[70] J. Larkin, S. Flicker, R. Koleszar-Green, S. Mintz, M. Dagnino, and C. Mitchell. HIV risk, systemic inequities, and Aboriginal youth: widening the circle for HIV prevention programming. *Can. J. Public Health*, pages 179–182, 2007. → pages 58

[71] H. Li, K. J. Bar, S. Wang, J. M. Decker, Y. Chen, C. Sun, J. F. Salazar-Gonzalez, M. G. Salazar, G. H. Learn, C. J. Morgan, et al. High multiplicity infection by HIV-1 in men who have sex with men. *PLoS Pathogens*, 6(5):e1000890, 2010. → pages 56, 62, 63

[72] H. Lu and P. Li. Stochastic projective methods for simulating stiff chemical reacting systems. *Computer Physics Communications*, 183(7):1427–1442, 2012. → pages 16

[73] G. MacDonald. *The epidemiology and control of Malaria*. Oxford University Press, London, 1957. → pages 8, 119

[74] M. Markowitz, M. Louie, A. Hurley, E. Sun, M. Di Mascio, A. S. Perelson, and D. D. Ho. A novel antiviral intervention results in more accurate assessment of human immunodeficiency virus type 1 replication dynamics and T-cell decay in vivo. *J. Virol.*, 77(8):5037–5038, 2003. → pages 66

[75] A. J. Marozsan, E. Fraundorf, A. Abraha, H. Baird, D. Moore, R. Troyer, I. Nankja, and E. J. Arts. Relationships between infectious titer, capsid protein levels, and reverse transcriptase activities of diverse human immunodeficiency virus type 1 isolates. *Journal of virology*, 78(20): 11130–11141, 2004. → pages 51, 154

[76] A. B. McDermott, J. Mitchen, S. Piaskowski, I. De Souza, L. J. Yant, J. Stephany, J. Furlott, and D. I. Watkins. Repeated low-dose mucosal simian immunodeficiency virus SIVmac239 challenge results in the same viral and immunological kinetics as high-dose challenge: a model for the evaluation of vaccine efficacy in nonhuman primates. *J. Virol.*, 78(6): 3140–3144, 2004. → pages 57

[77] C. McInnes, E. Druyts, S. Harvard, M. Gilbert, M. Tyndall, V. Lima, E. Wood, J. Montaner, and R. Hogg. HIV/AIDS in Vancouver, British Columbia: a growing epidemic. *Harm Reduction Journal*, 6(1):5, 2009. ISSN 1477-7517. doi:10.1186/1477-7517-6-5. URL http://www.harmreductionjournal.com/content/6/1/5. → pages 80

[78] J. W. Mellors, A. Munoz, J. V. Giorgi, J. B. Margolick, C. J. Tassoni, P. Gupta, L. A. Kingsley, J. A. Todd, A. J. Saah, R. Detels, et al. Plasma

viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. *Annals of internal medicine*, 126(12):946–954, 1997. → pages 5

[79] E. Mjolsness, D. Orendorff, P. Chatelain, and P. Koumoutsakos. An exact accelerated stochastic simulation algorithm. *The Journal of chemical physics*, 130(14):144110, 2009. → pages 16

[80] G. A. Ngwa and W. S. Shu. A mathematical model for endemic malaria with variable human and mosquito populations. *Math Comput Model*, 32: 747–763, 2000. → pages 120

[81] D. F. Nixon, S. M. Donahoe, W. M. Kakimoto, R. V. Samuel, K. J. Metzner, A. Gettie, T. Hanke, P. A. Marx, and R. I. Connor. Simian immunodeficiency virus-specific cytotoxic T lymphocytes and protection against challenge in rhesus macaques immunized with a live attenuated simian immunodeficiency virus vaccine. *Virology*, 266(1):203–210, 2000. → pages 57

[82] C. M. Obermeyer and M. Osborn. The utilization of testing and counseling for HIV: a review of the social and behavioral evidence. *Journal Information*, 97(10), 2007. → pages 58

[83] P. H. A. of Canada. HIV/AIDS Epi Updates: NationalHIVPrevalence and Incidence Estimates for 2011. 2014. → pages 80

[84] G. S. Ogilvie, D. L. Taylor, T. Trussler, R. Marchand, M. Gilbert, A. Moniruzzaman, and M. L. Rekart. Seeking sexual partners on the internet: a marker for risky sexual behaviour in men who have sex with men. *Canadian Journal of Public Health/Revue Canadienne de Sante'e Publique*, pages 185–188, 2008. → pages 81

[85] P. Patel, C. B. Borkowf, J. T. Brooks, A. Lasry, A. Lansky, and J. Mermin. Estimating per-act HIV transmission risk: a systematic review. *AIDS*, 28 (10):1509–1519, 2014. → pages 57, 62, 63, 86, 89, 92, 107

[86] J. E. Pearson, P. Krapivsky, and A. S. Perelson. Stochastic theory of early viral infection: continuous versus burst production of virions. *PLoS Comput Biol*, 7(2):e1001058, 2011. doi:10.1371/journal.pcbi.1001058. → pages 66

[87] A. S. Perelson and R. M. Ribeiro. Modeling the within-host dynamics of HIV infection. *BMC biology*, 11(1):96, 2013. → pages 78

[88] A. S. Perelson, D. E. Kirschner, and R. De Boer. Dynamics of HIV infection of CD4+T cells. *Math. Biosci.*, 114(1):81–125, 1993. → pages 67

[89] S. D. Pinkerton and P. R. Abramson. Effectiveness of condoms in preventing HIV transmission. *Social science & medicine*, 44(9): 1303–1312, 1997. → pages x, 86, 89, 90, 92, 107

[90] F. A. Plummer, N. Nagelkerke, S. Moses, J. Ndinya-Achola, J. Bwayo, and E. Ngugi. The importance of core groups in the epidemiology and control of hiv-1 infection. *AIDS (London, England)*, 5:S169, 1991. → pages 100, 114

[91] B. Ramratnam, S. Bonhoeffer, J. Binley, A. Hurley, L. Zhang, J. E. Mittler, M. Markowitz, J. P. Moore, A. S. Perelson, and D. D. Ho. Rapid production and clearance of HIV-1 and hepatitis C virus assessed by large volume plasma apheresis. *Lancet*, 354(9192):1782–1785, 1999. → pages 64, 66

[92] R. R. Regoes. The role of exposure history on HIV acquisition: insights from repeated low-dose challenge studies. *PLoS Comp Biol*, 8(11): e1002767, 2012. → pages 58

[93] R. M. Ribeiro, L. Qin, L. L. Chavez, D. Li, S. G. Self, and A. S. Perelson. Estimation of the initial viral growth rate and basic reproductive number during acute HIV-1 infection. *J. Virol.*, 84(12):6096–6102, 2010. → pages xii, 58, 59, 60, 61, 66

[94] R. Ross. *The prevention of malaria*. John Murray, London, 1911. → pages 7, 119

[95] R. Ross. The prevention of malaria murray. *London, UK*, 1911. → pages 3

[96] P. Rusert, M. Fischer, B. Joos, C. Leemann, H. Kuster, M. Flepp, S. Bonhoeffer, H. F. Günthard, and A. Trkola. Quantification of infectious HIV-1 plasma viral load using a boosted in vitro infection protocol. *Virology*, 326(1):113–129, 2004. → pages 51, 154

[97] I. Sangare, Y. Michalakis, B. Yameogo, R. Dabire, I. Morlais, and A. Cohuet. Studying fitness cost of Plasmodium falciparum infection in malaria vectors: validation of an appropriate negative control. *Malaria Journal*, 12(1):2, 2013. ISSN 1475-2875. doi:10.1186/1475-2875-12-2. URL http://www.malariajournal.com/content/12/1/2. → pages 120

[98] E.-J. Scholte, K. Ng'habi, J. Kihonda, W. Takken, K. Paijmans, S. Abdulla, G. F. Killeen, and B. G. J. Knols. An entomopathogenic fungus for control of adult African malaria mosquitoes. *Science*, 308:1641–1643, 2005. → pages 117

[99] D. L. Smith, F. E. McKenzie, R. W. Snow, and S. I. Hay. Revisiting the basic reproductive number for malaria and its implications for malaria control. *PLoS Biol*, 5(3):e42, Mar 2007. doi:10.1371/journal.pbio.0050042. → pages 8, 119

[100] D. L. Smith, T. A. Perkins, L. S. Tusting, T. W. Scott, and S. W. Lindsay. Mosquito Population Regulation and Larval Source Management in Heterogeneous Environments. *PLoS One*, 8(8):e71247, 2013. → pages 121

[101] R. Smith? and S. Hove-Musekwa. Determining Effective Spraying Periods to Control Malaria via Indoor Residual Spraying in Sub-Saharan Africa. *Journal of Applied Mathematics and Decision Sciences*, page 745463, 2008. doi:10.1155/2008/745463. → pages 8, 195, 196, 197, 198

[102] R. A. Stein. Super-spreaders in infectious diseases. *International Journal of Infectious Diseases*, 15(8):e510–e513, 2011. → pages 100, 114

[103] D. Taylor, M. Durigon, H. Davis, C. Archibald, B. Konrad, D. Coombs, M. Gilbert, D. Cook, M. Krajden, T. Wong, et al. Probability of a false negative HIV antibody test result during the window period: a tool for pre-and post-test counselling. *Int. J. STD & AIDS*, page 0956462414542987, 2014. → pages xii, 5, 56, 59, 60, 78

[104] M. Thomas and A. Read. Can fungal biopesticides control malaria? *Nature Microbiology Reviews*, 5:377–383, 2007. → pages 118

[105] J.-F. Trape, A. Tall, N. Diagne, O. Ndiath, A. B. Ly, J. Faye, F. Dieye-Ba, C. Roucher, C. Bouganali, A. Badiane, F. D. Sarr, C. Mazenot, A. Touré-Baldé, D. Raoult, P. Druilhe, O. Mercereau-Puijalon, C. Rogier, and C. Sokhna. Malaria morbidity and pyrethroid resistance after the introduction of insecticide-treated bednets and artemisinin-based combination therapies: a longitudinal study. *Lancet Infect Dis*, 11(12): 925–32, Dec 2011. doi:10.1016/S1473-3099(11)70194-3. → pages 117

[106] T. Trussler, P. Banks, R. Marchand, W. Robert, R. Gustafson, R. Hogg, M. Gilbert, and the ManCount Survey Team. ManCount Sizes-up the Gaps: a sexual health survey of gay men in Vancouver. 2010. → pages 91, 92, 105, 106

[107] B. Varghese, J. E. Maher, T. A. Peterman, B. M. Branson, and R. W. Steketee. Reducing the risk of sexual HIV transmission: quantifying the per-act risk for HIV on the basis of choice of partner, sex act, and condom use. *Sexually transmitted diseases*, 29(1):38–43, 2002. → pages x, 86, 89, 90, 92, 107

[108] P. L. Vernazza, J. J. Eron, S. A. Fiscus, and M. S. Cohen. Sexual transmission of HIV: infectiousness and prevention. *AIDS*, 13(2):155–166, 1999. → pages 56, 63, 68

[109] J. M. Vulule, R. F. Beach, F. K. Atieli, J. M. Roberts, D. L. Mount, and R. W. Mwangi. Reduced susceptibility of Anopheles gambiae to permethrin associated with the use of permethrin-impregnated bednets and curtains in Kenya. *Med Vet Entomol*, 8(1):71–5, Jan 1994. → pages 8

[110] M. J. Wawer, R. H. Gray, N. K. Sewankambo, D. Serwadda, X. Li, O. Laeyendecker, N. Kiwanuka, G. Kigozi, M. Kiddugavu, T. Lutalo, et al. Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. *J. Inf. Dis.*, 191(9):1403–1409, 2005. → pages 62

[111] X. Wei, S. K. Ghosh, M. E. Taylor, V. A. Johnson, E. A. Emini, P. Deutsch, J. D. Lifson, S. Bonhoeffer, M. A. Nowak, B. H. Hahn, et al. Viral dynamics in human immunodeficiency virus type 1 infection. *Nature*, 373 (6510):117–122, 1995. → pages 2

[112] M. T. White, J. T. Griffin, T. S. Churcher, N. M. Ferguson, M.-G. Basánez, and A. C. Ghani. Modelling the impact of vector control interventions on Anopheles gambiae population dynamics. *Parasites and Vectors*, 4:153, 2011. → pages 121

[113] WHO. Fact Sheet 1 HIV/AIDS: the infection. http://www.who.int/hiv/abouthiv/en/fact_sheet_hiv.htm. [Online; accessed 25-November-2014]. → pages 56

[114] WHO. World malaria report 2014, 2014. → pages 7, 8

[115] D. P. Wilson, J. Kahn, and S. M. Blower. Predicting the epidemiological impact of antiretroviral allocation strategies in kwazulu-natal: the effect of the urban–rural divide. *Proceedings of the National Academy of Sciences*, 103(38):14228–14233, 2006. → pages 3

[116] C. Worthington and T. Myers. Desired elements of HIV testing services: test recipient perspectives. *AIDS patient care and STDs*, 16(11):537–548, 2002. → pages 58

[117] C. Worthington and T. Myers. Factors underlying anxiety in HIV testing: Risk perceptions, stigma, and the patient-provider power dynamic. *Qualitative Health Research*, 13(5):636–655, 2003. → pages 58

[118] J. A. Yorke, H. W. Hethcote, and A. Nold. Dynamics and control of the transmission of gonorrhea. *Sexually transmitted diseases*, 5(2):51–56, 1978. → pages 100, 114

[119] L. Zhang, R. M. Ribeiro, J. R. Mascola, M. G. Lewis, G. Stiegler, H. Katinger, A. S. Perelson, and M. P. Davenport. Effects of antibody on viral kinetics in simian/human immunodeficiency virus infection: implications for vaccination. *J. Virol.*, 78(10):5520–5522, 2004. → pages 64, 66

[120] W. Zhou, X. Peng, Z. Yan, and Y. Wang. Accelerated stochastic simulation algorithm for coupled chemical reactions with delays. *Computational biology and chemistry*, 32(4):240–242, 2008. → pages 16

Appendix A

The T^*-V-W model with infectious and non-infectious virions

The T^*-V-W model accounts for the fact that the viral replication cycle is error-prone. Only a fraction Q of the virions that are produced by infected target cells are infectious, where Q is estimated to be on the order of $Q \approx 10^{-2} - 10^{-4}$ [63, 75, 96]. Non-infectious virions, denoted as W, are not able to infect new target cells and hence are harmless. The model can be written in ODE form as

$$\begin{aligned} \dot{T}^* &= kTV - \delta T^*, \\ \dot{V} &= pQT^* - (c + kT)V, \\ \dot{W} &= p(1-Q)T^* - cW, \end{aligned} \tag{A.1}$$

with $T^*(0) = n_0$, $V(0) = v_0$, and $W(0) = w_0$.

We investigate the probability that the population consists of n infected cells, v infectious virions and w non-infectious virions at time t given the initial population size of n_0 infected cells, v_0 infectious virions and w_0 non-infectious virions:

$$P(n,v,w,n_0,v_0,w_0,t) = \mathbb{P}((T^*,V,W)(t) = (n,v,w) \,|\, (T^*,V,W)(0) = (n_0,v_0,w_0)).$$

A.1 The master equations

For convenience of notation we omit arguments that are unchanged from n, v, w, n_0, v_0, w_0, t, that is, e.g. $P(n-1,v+1)$ is shorthand notation for $P(n-1,v+1,w,n_0,v_0,w_0,t)$. Using this notation and the same techniques as before we arrive at the forward master equation

$$\begin{aligned}\frac{dP}{dt} &= kT((v+1)P(n-1,v+1)+pQ(nP(v-1)-P)+\delta((n+1)P(n+1)-nP)\\ &\quad +c((v+1)P(v+1)-vP)+p(1-Q)(nP(w-1)-nP)\\ &\quad +c((w+1)P(w+1)-wP),\end{aligned} \tag{A.2}$$

and correspondingly the backward equation

$$\begin{aligned}\frac{dP}{dt} &= kTv_0(P(n_0+1,v_0-1)-P)+pQn_0(P(v_0+1)-P)+\delta n_0(P(n_0-1)-P)\\ &\quad +cv_0(P(v_0-1)-P)+p(1-Q)n_0(P(w_0+1)-P)+cv_0(P(w_0-1)-P)\end{aligned} \tag{A.3}$$

A.2 The probability generating function

Define the probability generating function $G:\mathbb{N}_0^3\times[0,\infty)\times\overline{\mathbb{D}}^3\to\overline{\mathbb{D}}$ via

$$(n_0,v_0,w_0,t,z_1,z_2,z_3)\mapsto\sum_{n,v,w=0}^{\infty}P(n,v,w,n_0,v_0,w_0,t)z_1^nz_2^vz_3^w.$$

Then, as before multiplying the forward master equation (A.2) by $z_1^nz_2^vz_3^w$ and summing over all n, v, w we obtain the following PDE:

$$\begin{aligned}\frac{\partial G}{\partial t} &= \Big[pQz_1(z_2-1)+p(1-Q)z_1(z_3-1)+\delta(1-z_1)\Big]\frac{\partial}{\partial z_1}G\\ &\quad +\Big[kT(z_1-z_2)+c(1-z_2)\Big]\frac{\partial}{\partial z_2}G+\Big[c(1-z_3)\Big]\frac{\partial}{\partial z_3}G\\ G_{n_0,v_0,w_0}(0,z_1,z_2,z_3) &= z_1^{n_0}z_2^{v_0}z_3^{w_0}.\end{aligned} \tag{A.4}$$

From the backward equation (A.3) we obtain the following ODE, where we

omit arguments that are unchanged from n_0, v_0, w_0, t, z_1, z_2, z_3.

$$\begin{aligned}\frac{dG}{dt} &= kTv_0(G(n_0+1,v_0-1)-G)+pQn_0(G(v_0+1)-G)\\ &\quad+\delta n_0(G(n_0-1)-G)+cv_0(G(v_0-1)-G)\\ &\quad+p(1-Q)n_0(G(w_0+1)-G)+cw_0(G(w_0-1)-G),\end{aligned}$$

which we rewrite with $G_1 = G_{1,0,0}$, $G_2 = G_{0,1,0}$, $G_3 = G_{0,0,1}$ and the branching property to

$$\begin{aligned}\frac{dG_1}{dt} &= pQG_1(G_2-1)+p(1-Q)G_1(G_3-1)+\delta(1-Q)G_1(G_3-1)+\delta(1-G_1)\\ \frac{dG_2}{dt} &= kT(G_1-G_2)+c(1-G_2)\\ \frac{dG_3}{dt} &= c(1-G_3)\end{aligned} \tag{A.5}$$

with $G_j(0) = z_j$.

A.3 Calculating total viral load

In order to calculate the total viral load $\mathbb{P}(V(t)+W(t)=\ell)$ we require a careful but straightforward calculation. For ease of notation we omit the initial conditions arguments n_0, v_0, w_0 in the probability P and the probability generating function G. Then, using the definition of the probability generating function we see that

$$\begin{aligned}&\mathbb{P}(V(t)+W(t)=\ell)\\ &\quad=\sum_{k=0}^{\ell}\mathbb{P}(V=\ell-k,W=k)\\ &\quad=\sum_{k=0}^{\ell}\sum_{n=0}^{\infty}P(n,\ell-k,k,t)\\ &\quad=\sum_{k=0}^{\ell}\sum_{n=0}^{\infty}\frac{1}{(\ell-k)!}P(n,\ell-k,k,t)1^n(\ell-k)!\\ &\quad=\sum_{k=0}^{\ell}\sum_{n,v=0}^{\infty}\frac{1}{(\ell-k)!}P(n,\ell-k,k,t)1^n v(v-1)\cdots(v-\ell-k+1)z_2^{v-\ell-k}\Big|_{z_2=0}\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=0}^{\ell} \sum_{n,v=0}^{\infty} \frac{1}{(\ell-k)!} \frac{\partial^{\ell-k}}{\partial z_2^{\ell-k}} P(n,\ell-k,k,t) 1^n z_2^v \Big|_{z_2=0} \\
&= \sum_{k=0}^{\ell} \sum_{n,v=0}^{\infty} \Bigg(\frac{1}{(\ell-k)!} \frac{\partial^{\ell-k}}{\partial z_2^{\ell-k}} \\
&\qquad \sum_{w=0}^{\infty} \frac{1}{k!} P(n,\ell-k,k,t) 1^n z_2^v w(w-1)\cdots(w-k+1) z_3^{w-k} \Big|_{(z_2,z_3)=(0,0)} \Bigg) \\
&= \sum_{k=0}^{\ell} \sum_{n,v,w=0}^{\infty} \frac{1}{(\ell-k)!} \frac{\partial^{\ell-k}}{\partial z_2^{\ell-k}} \frac{1}{k!} \frac{\partial^k}{\partial z_3^k} P(n,\ell-k,k,t) 1^n z_2^v z_3^w \Big|_{(z_2,z_3)=(0,0)} \\
&= \sum_{k=0}^{\ell} \frac{1}{(\ell-k)!} \frac{1}{k!} \frac{\partial^{\ell}}{\partial z_2^{\ell-k} z_3^k} \sum_{n,v,w=0}^{\infty} P(n,\ell-k,k,t) z_1^n z_2^v z_3^w \Big|_{(z_1,z_2,z_3)=(1,0,0)} \\
&= \sum_{k=0}^{\ell} \frac{1}{(\ell-k)!} \frac{1}{k!} \frac{\partial^{\ell} G(t,1,z_2,z_3)}{\partial z_2^{\ell-k} z_3^k} \Big|_{(z_2,z_3)=(0,0)}
\end{aligned}
$$

To calculate the latter, we need to use the Cauchy integral formula twice, i.e., use a double integral.

$$
\begin{aligned}
\mathbb{P}(V=\ell-k, W=k) &= \frac{1}{(\ell-k)!} \frac{1}{k!} \frac{\partial^{\ell} G(t,1,z_2,z_3)}{\partial z_2^{\ell-k} z_3^k} \Big|_{(z_2,z_3)=(0,0)} \\
&= \frac{1}{(\ell-k)!} \frac{\partial^{\ell-k}}{\partial z_2^{\ell-k}} g(z_2) \Big|_{z_2=0},
\end{aligned}
$$

where

$$
\begin{aligned}
g(z_2) &= \frac{1}{k!} \frac{\partial^k f_{z_2}(z_3)}{\partial z_3^k} \Big|_{z_3=0} = \frac{1}{2\pi i} \int_C \frac{f_{z_2}(w)}{w^{k+1}} dw = \frac{1}{2\pi} \int_0^{2\pi} f_{z_2}(e^{i\varphi}) e^{-ik\varphi} d\varphi \\
&= \frac{1}{2\pi} \int_0^{2\pi} G(t,1,z_2,e^{i\varphi}) e^{-ik\varphi} d\varphi,
\end{aligned}
$$

and hence

$$
\mathbb{P}(V=\ell-k, W=k) = \frac{1}{(\ell-k)!} \frac{\partial^{\ell-k} g(z_2)}{\partial z_2^{\ell-k}} \Big|_{z_2=0}
$$

$$
\begin{aligned}
&= \frac{1}{2\pi}\int_0^{2\pi} g(e^{i\psi})e^{-i(\ell-k)\psi}d\psi \\
&= \frac{1}{2\pi}\int_0^{2\pi}\frac{1}{2\pi}\int_0^{2\pi} G(t,1,e^{i\psi},e^{i\varphi})e^{-ik\varphi}d\varphi e^{-i(\ell-k)\psi}d\psi \\
&= \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi} G(t,1,e^{i\psi},e^{i\varphi})e^{-ik\varphi}e^{-i(\ell-k)\psi}d\varphi d\psi \\
&= \frac{1}{\pi^2}\mathrm{Re}\int_0^{\pi}\int_0^{\pi} G(t,1,e^{i\psi},e^{i\varphi})e^{-ik\varphi}e^{-i(\ell-k)\psi}d\varphi d\psi,
\end{aligned}
$$

since $\overline{G(t,1,z_2,z_3)} = G(t,1,\overline{z_2},\overline{z_3})$. Therefore, summing over all possible values of k we obtain

$$
\begin{aligned}
\mathbb{P}(V(t)+W(t)=\ell) &= \sum_{k=0}^{\ell} P(V=\ell-k, W=k) \\
&= \frac{1}{\pi^2}\mathrm{Re}\int_0^{\pi}\int_0^{\pi} G(t,1,e^{i\psi},e^{i\varphi})\sum_{k=0}^{\ell} e^{-ik\varphi}e^{-i(\ell-k)\psi}d\varphi d\psi \\
&= \frac{1}{\pi^2}\mathrm{Re}\int_0^{\pi}\int_0^{\pi} G(t,1,e^{i\psi},e^{i\varphi})e^{-i\ell\psi}\frac{1-e^{-i(\ell+1)(\varphi-\psi)}}{1-e^{-i(\varphi-\psi)}}d\varphi d\psi.
\end{aligned} \tag{A.6}
$$

A.4 Mean behaviour of the stochastic process

We calculate the mean number of virus particles $V+W$ at time t. By the linearity of the expected value we obtain

$$
\begin{aligned}
E[V+W] = E[V]+E[W] &= \sum_{v=0}^{\infty} v\mathbb{P}(V=v) + \sum_{w=0}^{\infty} w\mathbb{P}(W=w) \\
&= \sum_{n,w=0}^{\infty}\sum_{v=0}^{\infty} P(n,v,w,t)1^n v 1^{v-1}1^w + \sum_{n,v=0}^{\infty}\sum_{w=0}^{\infty} P(n,v,w,t)1^n 1^v w 1^{w-1} \\
&= \left[\left(\frac{\partial}{\partial z_2}+\frac{\partial}{\partial z_3}\right)G(t,z_1,z_2,z_3)\right]\Big|_{(z_1,z_2,z_3)=(1,1,1)}.
\end{aligned}
$$

We now find an ODE for $E[V+W]$. The time derivative of the partial derivatives will be calculated using the PDE formulation (A.4) for G, and swapping the order

of differentiation.

$$\begin{aligned}
\frac{\partial}{\partial t}\frac{\partial G(t,z_1,z_2,z_3)}{\partial z_2} &= \frac{\partial}{\partial z_2}\frac{\partial G(t,z_1,z_2,z_3)}{\partial t} \\
&= \frac{\partial}{\partial z_2}\left[\left[pQz_1(z_2-1)+p(1-Q)z_1(z_3-1)+\delta(1-z_1)\right]\frac{\partial G}{\partial z_1}\right. \\
&\qquad \left. +\left[kT(z_1-z_2)+c(1-z_2)\right]\frac{\partial G}{\partial z_2}+c(1-z_3)\frac{\partial G}{\partial z_3}\right] \\
&= pQz_1\frac{\partial G}{\partial z_1} \\
&\quad +\left[pQz_1(z_2-1)+p(1-Q)z_1(z_3-1)+\delta(1-z_1)\right]\frac{\partial^2 G}{\partial z_1 z_2} \\
&\quad -(c+kT)\frac{\partial G}{\partial z_2}+\left[kT(z_1-z_2)+c(1-z_2)\right]\frac{\partial^2 G}{\partial z_2^2} \\
&\quad +c(1-z_3)\frac{\partial^2 G}{\partial z_2 z_3}.
\end{aligned}$$

Evaluated at $(z_1,z_2,z_3)=(1,1,1)$ we obtain

$$\begin{aligned}
\left.\frac{\partial}{\partial t}\frac{\partial G(t,z_1,z_2,z_3)}{\partial z_2}\right|_{(z_1,z_2,z_3)=(1,1,1)} &= pQ\left.\frac{\partial G(t,z_1,z_2,z_3)}{\partial z_1}\right|_{(z_1,z_2,z_3)=(1,1,1)} \\
&\quad -(c+kT)\left.\frac{\partial G(t,z_1,z_2,z_3)}{\partial z_2}\right|_{(z_1,z_2,z_3)=(1,1,1)}
\end{aligned}$$

and similarly

$$\begin{aligned}
\left.\frac{\partial}{\partial t}\frac{\partial G(t,z_1,z_2,z_3)}{\partial z_1}\right|_{(z_1,z_2,z_3)=(1,1,1)} &= -\delta\left.\frac{\partial G(t,z_1,z_2,z_3)}{\partial z_1}\right|_{(z_1,z_2,z_3)=(1,1,1)} \\
&\quad +kT\left.\frac{\partial G(t,z_1,z_2,z_3)}{\partial z_2}\right|_{(z_1,z_2,z_3)=(1,1,1)} \\
\left.\frac{\partial}{\partial t}\frac{\partial G(t,z_1,z_2,z_3)}{\partial z_3}\right|_{(z_1,z_2,z_3)=(1,1,1)} &= p(1-Q)\left.\frac{\partial G(t,z_1,z_2,z_3)}{\partial z_1}\right|_{(z_1,z_2,z_3)=(1,1,1)} \\
&\quad -c\left.\frac{\partial G(t,z_1,z_2,z_3)}{\partial z_3}\right|_{(z_1,z_2,z_3)=(1,1,1)}
\end{aligned}$$

Hence, the time-dependent functions $f_1(t) = \frac{\partial G(t,z_1,z_2,z_3)}{\partial z_1}\Big|_{(z_1,z_2,z_3)=(1,1,1)}$, $f_2(t) = \frac{\partial G(t,z_1,z_2,z_3)}{\partial z_2}\Big|_{(z_1,z_2,z_3)=(1,1,1)}$ and $f_3(t) = \frac{\partial G(t,z_1,z_2,z_3)}{\partial z_3}\Big|_{(z_1,z_2,z_3)=(1,1,1)}$ satisfy the original linear ODE (A.1):

$$\begin{aligned}
\dot{f}_1 &= -\delta f_1 + kT f_2, \\
\dot{f}_2 &= pQf_1 - (c+kT)f_2, \\
\dot{f}_3 &= p(1-Q)f_1 - cf_3,
\end{aligned}$$

with $f_1(0) = n_0$, $f_2(0) = v_0$, $f_3(0) = w_0$. From the solution to this linear ODE we obtain

$$E[V(t) + W(t)] = f_2(t) + f_3(t).$$

A.5 Calculating cumulative probabilities

Building on earlier results we use (A.6) to quickly calculate the probability that the total viral load is below the detection limit dL:

$$\begin{aligned}
&P(V+W < \mathrm{dL}) \\
&\quad = \sum_{\ell=0}^{\mathrm{dL}-1} P(V+W=\ell) \\
&\quad = \sum_{\ell=0}^{\mathrm{dL}-1} \frac{1}{\pi^2}\,\mathrm{Re}\int_0^\pi \int_0^\pi G(1,e^{i\psi},e^{i\varphi})e^{-i\ell\psi}\frac{1-e^{-i(\ell+1)(\varphi-\psi)}}{1-e^{-i(\varphi-\psi)}}d\varphi d\psi \\
&\quad = \frac{1}{\pi^2}\,\mathrm{Re}\int_0^\pi \int_0^\pi \frac{G(1,e^{i\psi},e^{i\varphi})}{1-e^{-i(\varphi-\psi)}}\sum_{\ell=0}^{\mathrm{dL}-1} e^{-i\ell\psi}(1-e^{-i(\ell+1)(\varphi-\psi)})d\varphi d\psi \\
&\quad = \frac{1}{\pi^2}\,\mathrm{Re}\int_0^\pi \int_0^\pi \frac{G(1,e^{i\psi},e^{i\varphi})}{1-e^{-i(\varphi-\psi)}}\left(\frac{1-e^{i\mathrm{dL}\psi}}{1-e^{-i\psi}} - e^{-i(\varphi-\psi)}\frac{1-e^{-i\mathrm{dL}\varphi}}{1-e^{-i\varphi}}\right)d\varphi d\psi.
\end{aligned}$$

A.6 The extinction probability, basic reproduction number R_0 and initial growth rate

Since non-infectious virus particles W can not infect new target cells, extinction is guaranteed for each lineage of W. On the other hand, the T^*-V cycle is equivalent to the T^*-V model in Section 2.11, so we expect the same extinction probabilities for a T^* and V lineage, respectively. To calculate the extinction probabilities q_1, q_2 and q_3 we consider the steady states of the ODE (A.5) to find

$$q_1 = q_2 = q_3 = 1, \quad \text{or} \quad q_1 = \frac{\delta}{pQ}\frac{c+kT}{kT}, \quad q_2 = \frac{\delta}{pQ} + \frac{c}{c+kT}, \quad q_3 = 1.$$

The quantities q_1, q_2 and q_3 are the probabilities that an infection that started with a single T^*, V or W goes extinct, respectively.

Similarly to the T^*-V model the basic reproduction number R_0 for this model is the expected number of new infected cells that are descendants of a single infected cell. Since infectious virions infect target cells at rate kT and have an average life-span on $1/(c+kT)$, while target cells produce infectious virions at rate pQ over their average life span $1/\delta$ we obtain

$$R_0 = \frac{pQkT}{\delta(c+kT)}.$$

Next, we get the innate growth rate r as the largest real part of the eigenvalues of the matrix given by

$$\begin{bmatrix} \dot{T}^* \\ \dot{V} \\ \dot{W} \end{bmatrix} = \begin{bmatrix} -\delta & kT & 0 \\ pQ & -c-kT & 0 \\ (1-Q)p & 0 & -c \end{bmatrix} \begin{bmatrix} T^* \\ V \\ W \end{bmatrix}$$

with characteristic polynomial

$$(-\lambda - c)(\lambda^2 + (\delta + c + kT)\lambda + \delta(c+kT) - pQkT).$$

For $R_0 > 1$ this polynomial has two negative and one positive root, so that in this

case the innate growth rate is given by

$$r = \frac{1}{2}\left[-\delta - c - kT + \sqrt{(\delta - c - kT)^2 + 4pQkT}\right]. \tag{A.7}$$

A.7 Conditioning on non-extinction

Following the same steps as for the previous models we obtain the probability generating function $\tilde{G}$ of the process conditioned on non-extinction as

$$\tilde{G}(n_0, v_0, w_0, t, z_1, z_2 z_3) = \frac{G(n_0, v_0, w_0, t, z_1, z_2, z_3) - G(n_0, v_0, w_0, t, q_1 z_1, q_2 z_2, q_3 z_3)}{1 - q_1^{n_0} q_2^{v_0} q_3^{w_0}}$$

A.8 Mean behaviour and cumulative probabilities for the conditioned T^*-V-W model

Let $\tilde{V}$ and $\tilde{W}$ be the random variable of the number of infectious and non-infectious virus particles in the conditioned process, respectively.

$$\begin{aligned} E[\tilde{V} + \tilde{W}] &= \left[\left(\frac{\partial}{\partial z_2} + \frac{\partial}{\partial z_3}\right)\tilde{G}(t, z_1, z_2, z_3)\right]\Big|_{\mathbf{z}=\mathbf{1}} \\ &= \frac{1}{1 - q_1^{n_0} q_2^{v_0} q_3^{w_0}}\left[\left(\frac{\partial}{\partial z_2} + \frac{\partial}{\partial z_3}\right) G(z_1, z_2, z_3)\right]\Big|_{\mathbf{z}=\mathbf{1}} \\ &\quad - \frac{1}{1 - q_1^{n_0} q_2^{v_0} q_3^{w_0}}\left[\left(\frac{\partial}{\partial z_2} + \frac{\partial}{\partial z_3}\right) G(t, q_1 z_1, q_2 z_2, q_3 z_3)\right]\Big|_{\mathbf{z}=\mathbf{1}} \end{aligned}$$

We know the former summand as the mean of the unconditioned process. To calculate the latter define

$$g_j(t) = \left.\frac{\partial G(n_0, v_0, w_0, t, q_1 z_1, q_2 z_2, q_3 z_3)}{\partial z_j}\right|_{\mathbf{z}=\mathbf{1}}$$

for $j = 1, 2, 3$ and note that

$$g_j(t) = \left.\frac{\partial G(n_0, v_0, w_0, t, q_1 z_1, q_2 z_2, q_3 z_3)}{\partial z_j}\right|_{\mathbf{z}=\mathbf{1}}$$

$$= q_j \left. \frac{\partial G(n_0, v_0, w_0, t, x_1, x_2, x_3)}{\partial x_j} \right|_{(x_1,x_2,x_3)=(q_1,q_2,q_3)}.$$

Taking the time-derivative and using the PDE formulation (A.4) for G, we obtain after swapping the partial derivatives

$$\begin{aligned}
\frac{\partial}{\partial t} g_1(t) &= q_1 \frac{\partial}{\partial t} \left. \frac{\partial G(n_0, v_0, w_0, t, x_1, x_2, x_3)}{\partial x_1} \right|_{(x_1,x_2,x_3)=(q_1,q_2,q_3)} \\
&= q_1 \frac{\partial}{\partial x_1} \Big\{ \left[pQx_1(x_2-1) + p(1-Q)x_1(x_3-1) + \delta(1-x_1) \right] \frac{\partial G}{\partial x_1} \\
&\quad + \left[kT(x_1-x_2) + c(1-x_2) \right] \frac{\partial G}{\partial x_2} + c(1-x_3) \frac{\partial G}{\partial x_3} \Big\} \Big|_{(x_1,x_2,x_3)=(q_1,q_2,q_3)} \\
&= q_1 \Big\{ \left(pQ(x_2-1) + p(1-Q)(x_3-1) - \delta \right) \frac{\partial G}{\partial x_1} + (\dots) \frac{\partial^2 G}{\partial x_1^2} \\
&\quad + kT \frac{\partial G}{\partial x_2} + (\dots) \frac{\partial^2 G}{\partial x_1 \partial x_2} + 0 \frac{\partial G}{\partial x_3} + (\dots) \frac{\partial^2 G}{\partial x_1 \partial x_3} \Big\} \Big|_{(x_1,x_2,x_3)=(q_1,q_2,q_3)} \\
&= q_1 \left(pQ \left(\frac{\delta}{pQ} + \frac{c}{c+kT} - 1 \right) - \delta \right) \frac{\partial G}{\partial x_1} + q_1 kT \left. \frac{\partial G}{\partial x_2} \right|_{(x_1,x_2,x_3)=(q_1,q_2,q_3)} \\
&= -\frac{\delta}{q_1} g_1 + \frac{q_1}{q_2} kT g_2 \\
\frac{\partial}{\partial t} g_2(t) &= q_2 \frac{\partial}{\partial t} \left. \frac{\partial G(n_0, v_0, w_0, t, x_1, x_2, x_3)}{\partial x_2} \right|_{(x_1,x_2,x_3)=(q_1,q_2,q_3)} \\
&= q_2 \frac{\partial}{\partial x_2} \Big\{ \left[pQx_1(x_2-1) + p(1-Q)x_1(x_3-1) + \delta(1-x_1) \right] \frac{\partial G}{\partial x_1} \\
&\quad + \left[kT(x_1-x_2) + c(1-x_2) \right] \frac{\partial G}{\partial x_2} + c(1-x_3) \frac{\partial G}{\partial x_3} \Big\} \Big|_{(x_1,x_2,x_3)=(q_1,q_2,q_3)} \\
&= q_2 \Big\{ pQx_1 \frac{\partial G}{\partial x_1} + (\dots) \frac{\partial^2 G}{\partial x_1 \partial x_2} + (-kT-c) \frac{\partial G}{\partial x_2} \\
&\quad + (\dots) \frac{\partial^2 G}{\partial x_2^2} + 0 \frac{\partial G}{\partial x_3} + (\dots) \frac{\partial^2 G}{\partial x_3 \partial x_2} \Big\} \Big|_{(x_1,x_2,x_3)=(q_1,q_2,q_3)} \\
&= pQq_2 g_1 - (c+kT) g_2 \\
\frac{\partial}{\partial t} g_3(t) &= q_3 \frac{\partial}{\partial t} \left. \frac{\partial G(n_0, v_0, w_0, t, x_1, x_2, x_3)}{\partial x_3} \right|_{(x_1,x_2,x_3)=(q_1,q_2,q_3)} \\
&= q_3 \frac{\partial}{\partial x_3} \Big\{ \left[pQx_1(x_2-1) + p(1-Q)x_1(x_3-1) + \delta(1-x_1) \right] \frac{\partial G}{\partial x_1}
\end{aligned}$$

$$
\begin{aligned}
&\quad + [kT(x_1 - x_2) + c(1 - x_2)] \frac{\partial G}{\partial x_2} + c(1 - x_3) \frac{\partial G}{\partial x_3} \Big\} \Big|_{(x_1,x_2,x_3)=(q_1,q_2,q_3)} \\
&= q_3 \Big\{ p(1-Q)x_1 \frac{\partial G}{\partial x_1} + (\dots) \frac{\partial^2 G}{\partial x_1 \partial x_2} + 0 \frac{\partial G}{\partial x_2} \\
&\qquad + (\dots) \frac{\partial^2 G}{\partial x_2^2} - c \frac{\partial G}{\partial x_3} + (\dots) \frac{\partial^2 G}{\partial x_3 \partial x_2} \Big\} \Big|_{(x_1,x_2,x_3)=(q_1,q_2,q_3)} \\
&= p(1-Q) q_3 g_1 - c g_3.
\end{aligned}
$$

The second-order partial derivatives vanish because the factor multiplying them is zero by the property of q_1, q_2 and q_3. For the initial conditions, we find

$$
\begin{aligned}
g_1(0) &= n_0 q_1^{n_0} q_2^{v_0} q_3^{w_0}, \\
g_2(0) &= v_0 q_1^{n_0} q_2^{v_0} q_3^{w_0}, \\
g_3(0) &= w_0 q_1^{n_0} q_2^{v_0} q_3^{w_0},
\end{aligned}
$$

so that we obtain $g_2(t)$ and $g_3(t)$ by solving the ODE

$$
\begin{aligned}
\dot{g}_1(t) &= \frac{q_1}{q_2} kT g_2 - \frac{\delta}{q_1} g_1, & g_1(0) &= n_0 q_1^{n_0} q_2^{v_0} q_3^{w_0} \\
\dot{g}_2(t) &= pQq_2 g_1 - (c + kT) g_2, & g_2(0) &= v_0 q_1^{n_0} q_2^{v_0} q_3^{w_0} \\
\dot{g}_3(t) &= p(1-Q) q_3 g_1 - c g_3, & g_3(0) &= w_0 q_1^{n_0} q_2^{v_0} q_3^{w_0}
\end{aligned}
$$

Putting these results together we find the expected value of the total viral load for the T^*-V-W model, conditioned on non-exinction:

$$
E[\tilde{V}(t) + \tilde{W}(t)] = \frac{f_2(t) + f_3(t) - g_2(t) - g_3(t)}{1 - q_1^{n_0} q_2^{v_0} q_3^{w_0}}. \tag{A.8}
$$

Appendix B

Additional material for early HIV infection

B.1 Preparing the dataset of patient timecourses

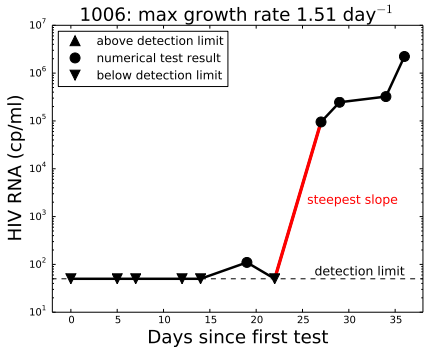
In the pre-processing stage we merge all datasets in such a way that there remains only one time-series for each patient.

We merged all datasets first, for a total of 427 time series (with duplicates). We are interested in measuring the maximal viral growth rate during the ramp-up phase immediately after infection. Hence, from the 427 time series we first removed 60 time series that were unsuitable because they only contained a single data point or recorded a high viral load at the first (not necessarily positive) test (more than 25,000 HIV RNA copies per milliliter), indicating that the ramp-up phase has been missed. Next, we removed obvious duplicate entries, and combined complementary time series where disjoint subsets of the plasma samples were tested with different HIV tests. Many HIV RNA tests have upper limits of detection, which can differ for different tests. Unfortunately, measurements above the detection limit are only sometimes labeled as such, other times the upper limit of detection is reported like an accurate measurement. In cases where two or more HIV RNA tests were taken on the same sample and had identical numerical values and at least one of the results was labeled as above the detection limit of the test, we interpreted this as a result above the limit of detection of both tests. This was the case in 21 samples.

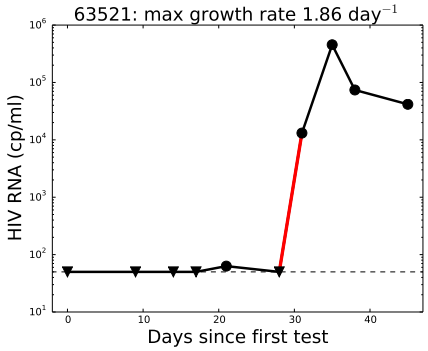
When several HIV RNA tests of the same sample were recorded as below the detection limit but did not match, as is the case for 36 time series, we took the geometric mean to preserve the distribution of measurement error, which is proportional to the logarithm of the value.

The remaining 129 time series still had some overlap where several tests were taken for the same patient at the same times, but the measured values were not identical or compatible in the sense above. In particular, some tests were not able to detect HIV RNA as early as others, or some tests only provided an upper bound where other tests measured a definite and different numerical value. We ignore tests that detected HIV RNA later, or that only provide an upper bound when a different test on the same plasma samples found a definite numerical value because we believe that these are less accurate tests. As a final tie-breaker we chose the time series of the test that was used on more plasma samples. Finally, we disregard time series where fewer plasma samples were tested. After this cleaning we obtain 94 unique time series of 94 patients with measure or averaged HIV RNA measurement at several time points. In 7 of these no meaningful viral growth rate could be calculated, either because HIV RNA was undetectable in all recorded tests, or because the viral load decreased initially, indicating that the first test was taken after the peak viral load. We further removed 5 time series where only two positive test results were available and they were more than 25 days apart, because it is unlikely that both tests were taken during the ramp-up stage. This leaves 82 patients, of which 4 more were excluded by inspection, as follows:

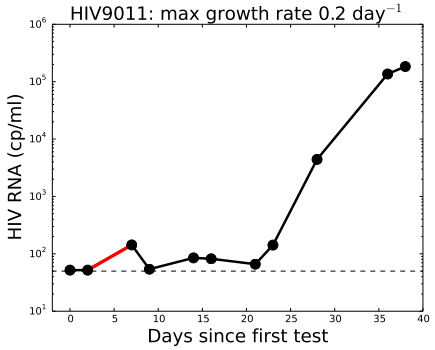
1006: Since the left values of the steepest slope is undetectable, yet preceded by a positive viral load measurement (110 RNA copies/mL), we suspect the left value of the steepest slope to be a false negative. Hence the actual steepest slope should be much more shallow.



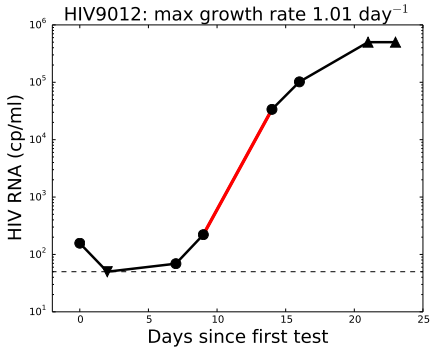
63521: Since the left values of the steepest slope is undetectable, yet preceded by a positive viral load measurement (63 RNA copies/mL), we suspect the left value of the steepest slope to be a false negative. Hence the actual steepest slope should be much more shallow.



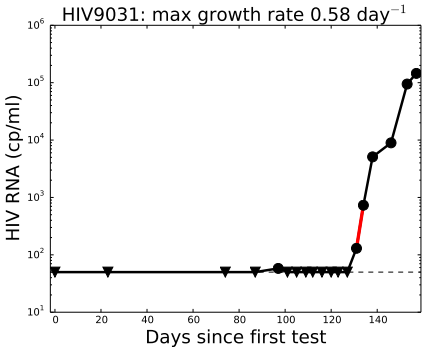
HIV9011: The bDNA test shows positive values immediately, which are just above the limit of detection. Due to the number of positive test results it is unlikely, but not impossible, that these are all false positive tests, and that the real viral increase is only observed after day 20. If this is not the case we have most likely missed the initial rate of viral increase.



HIV9012: The bDNA test shows positive values immediately, which are just above the limit of detection. Due to the number of positive test results it is unlikely, but not impossible, that these are all false positive tests, and that the real viral increase is only observed after day 5. If this is not the case we have most likely missed the initial rate of viral increase.

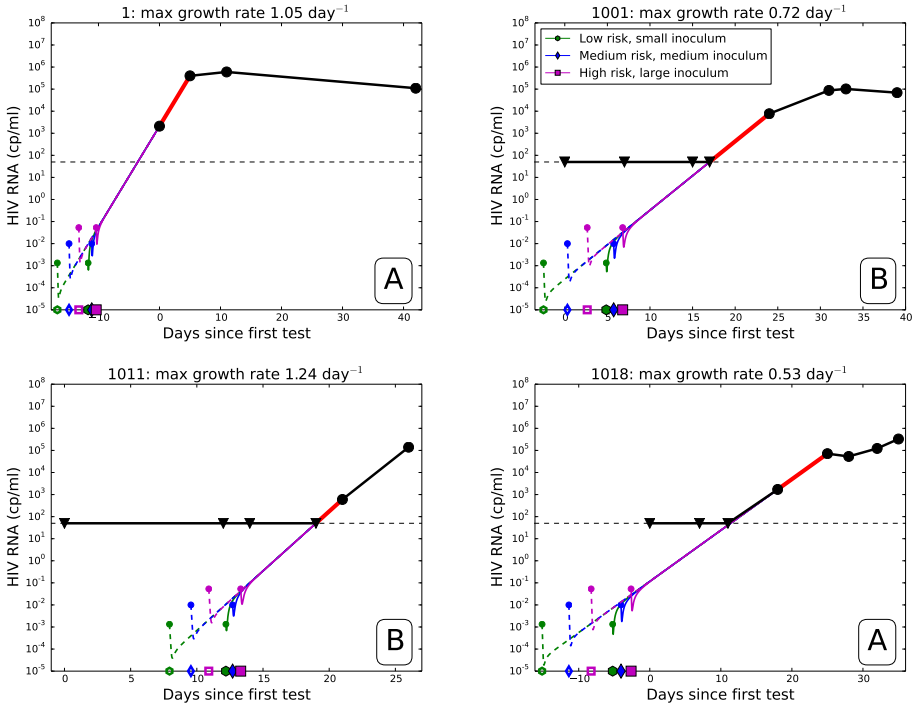


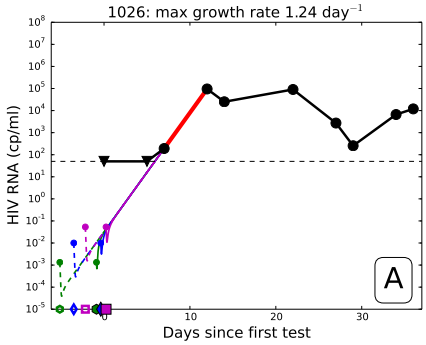
Finally, patient HIV9031 was also considered for exclusion because of the early detectable viral load. However, we decided to keep HIV9031 in the dataset because the single positive bDNA test is followed by eight negative test results, which makes us believe that the first positive test was a false positive. Further, this time series is present in another dataset that we use, and there the RNA test found no HIV RNA in the sample.

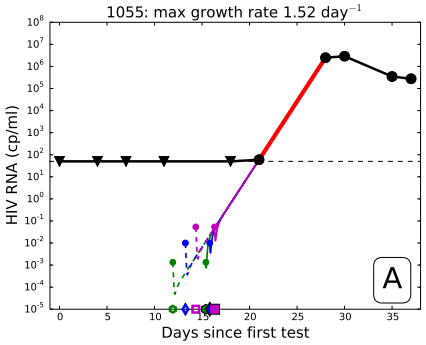


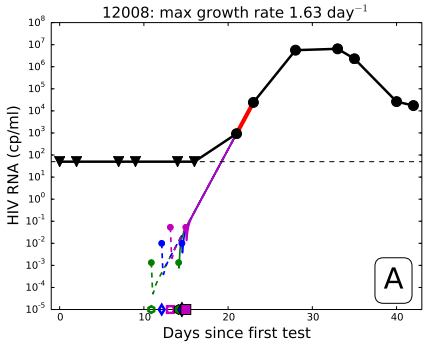
B.2 Patient data

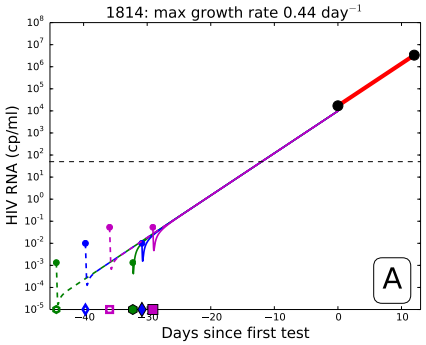
Here we plot the data of all 78 patients that were included in our analysis. We also show the fit for each patient, for all risk categories (see main paper).

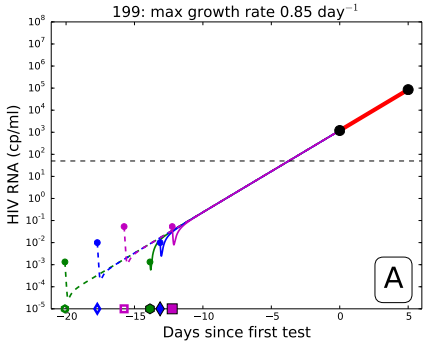


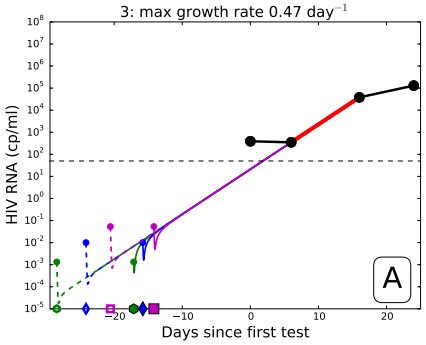

1026: max growth rate 1.24 day^{-1}
HIV RNA (cp/ml)
Days since first test
A

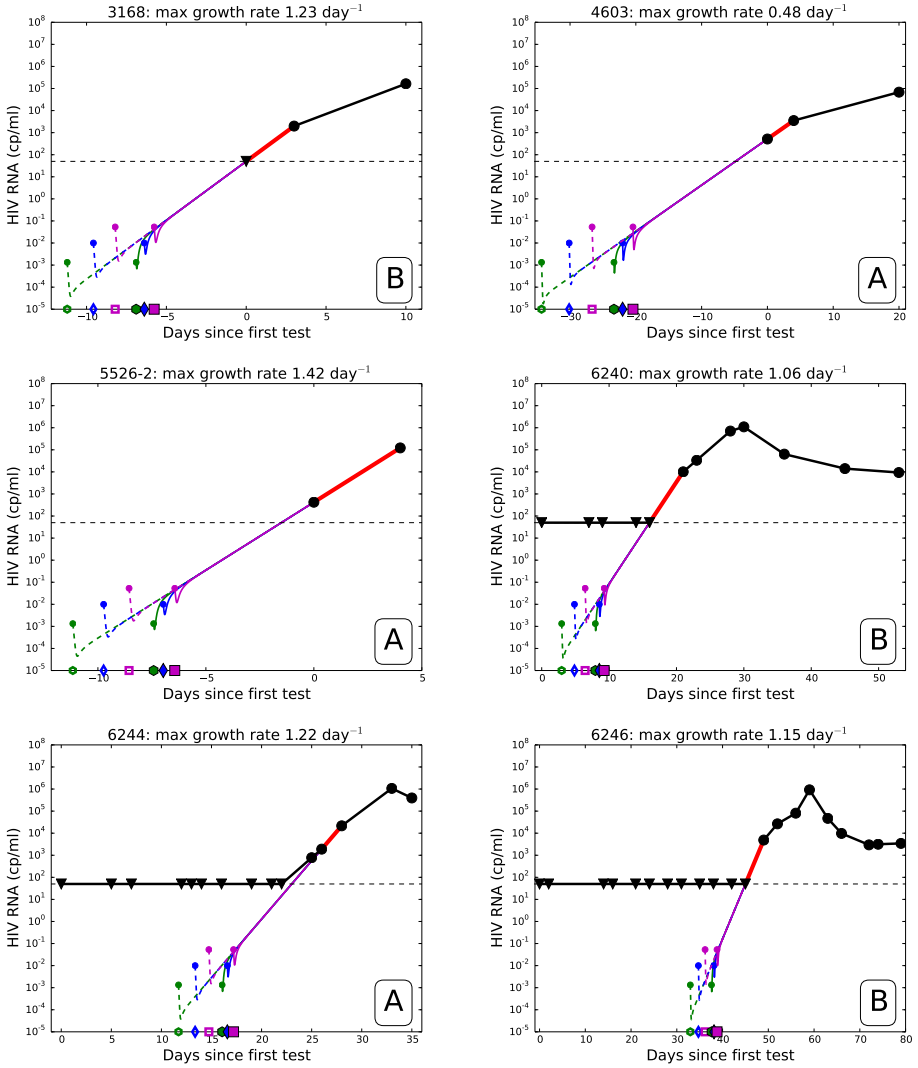

1055: max growth rate 1.52 day^{-1}
HIV RNA (cp/ml)
Days since first test
A

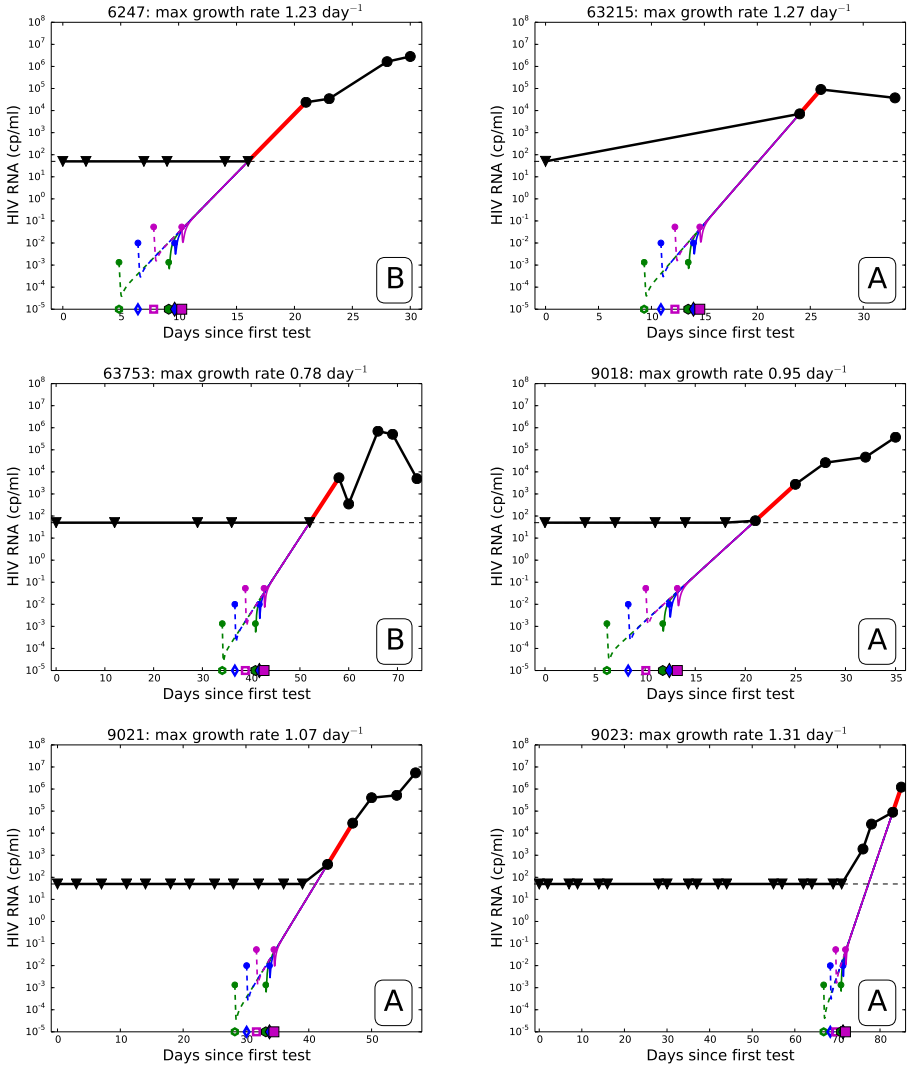

12008: max growth rate 1.63 day^{-1}
HIV RNA (cp/ml)
Days since first test
A

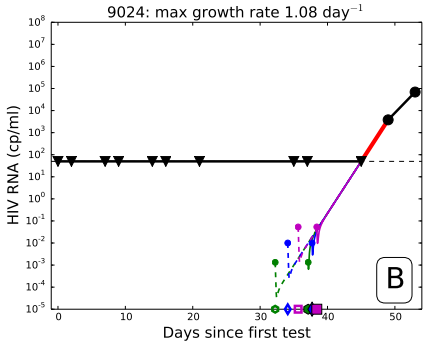

1814: max growth rate 0.44 day^{-1}
HIV RNA (cp/ml)
Days since first test
A

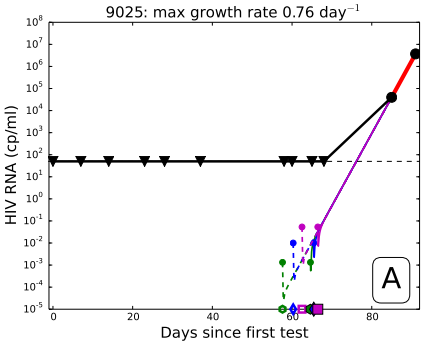

199: max growth rate 0.85 day^{-1}
HIV RNA (cp/ml)
Days since first test
A

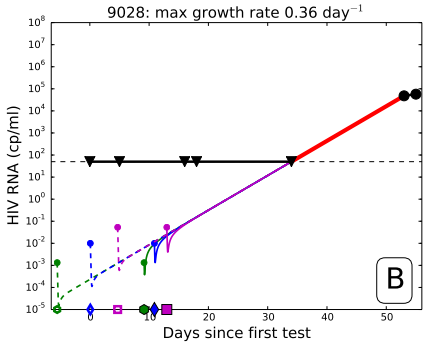

3: max growth rate 0.47 day^{-1}
HIV RNA (cp/ml)
Days since first test
A

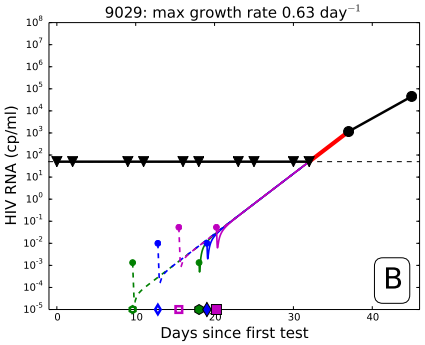

3168: max growth rate 1.23 day^{-1}
4603: max growth rate 0.48 day^{-1}
5526-2: max growth rate 1.42 day^{-1}
6240: max growth rate 1.06 day^{-1}
6244: max growth rate 1.22 day^{-1}
6246: max growth rate 1.15 day^{-1}
HIV RNA (cp/ml)
Days since first test
A
B

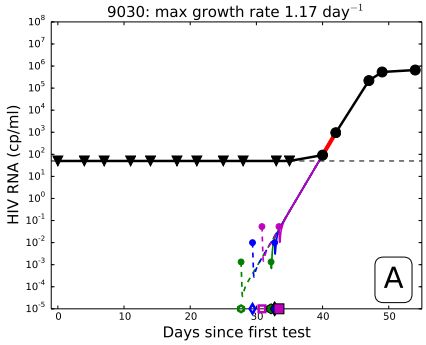

6247: max growth rate 1.23 day^{-1}
63215: max growth rate 1.27 day^{-1}
63753: max growth rate 0.78 day^{-1}
9018: max growth rate 0.95 day^{-1}
9021: max growth rate 1.07 day^{-1}
9023: max growth rate 1.31 day^{-1}
HIV RNA (cp/ml)
Days since first test
A
B

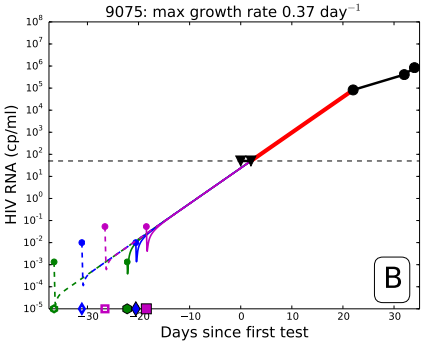

9024: max growth rate 1.08 day^{-1}
HIV RNA (cp/ml)
Days since first test
B

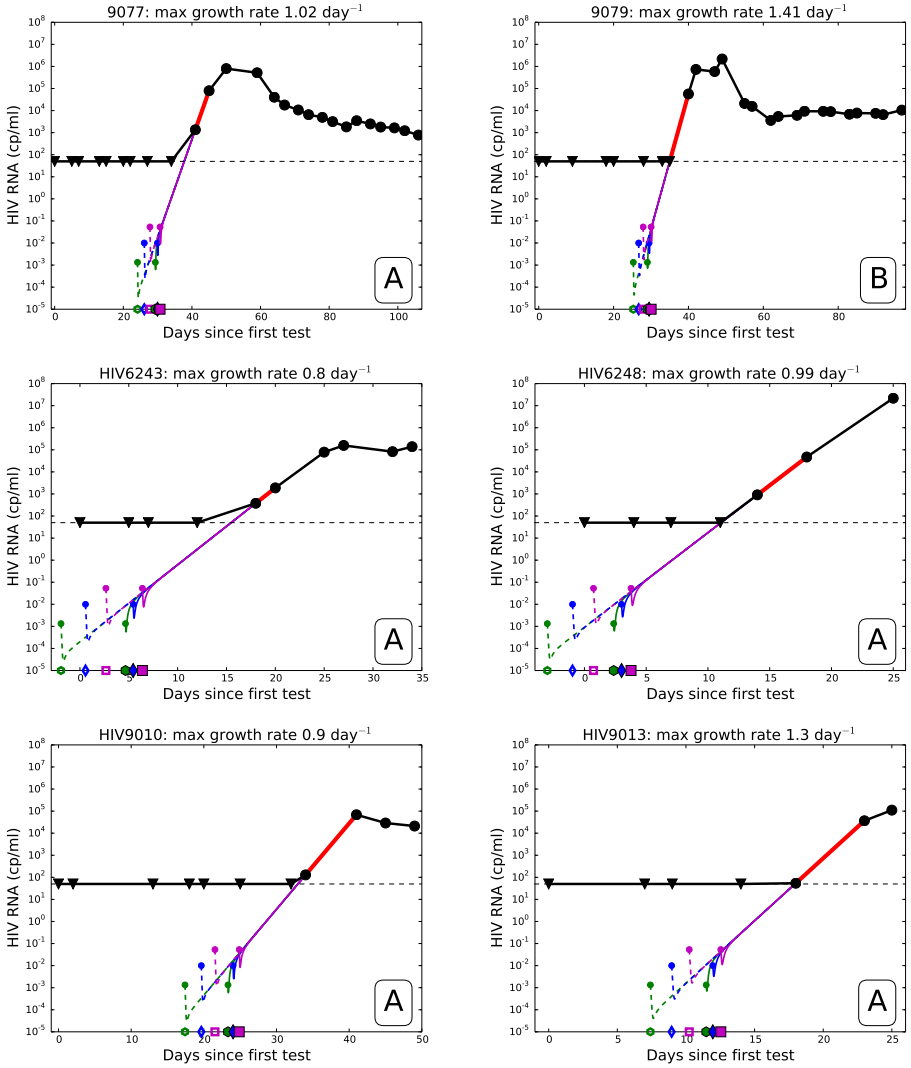

9025: max growth rate 0.76 day^{-1}
HIV RNA (cp/ml)
Days since first test
A

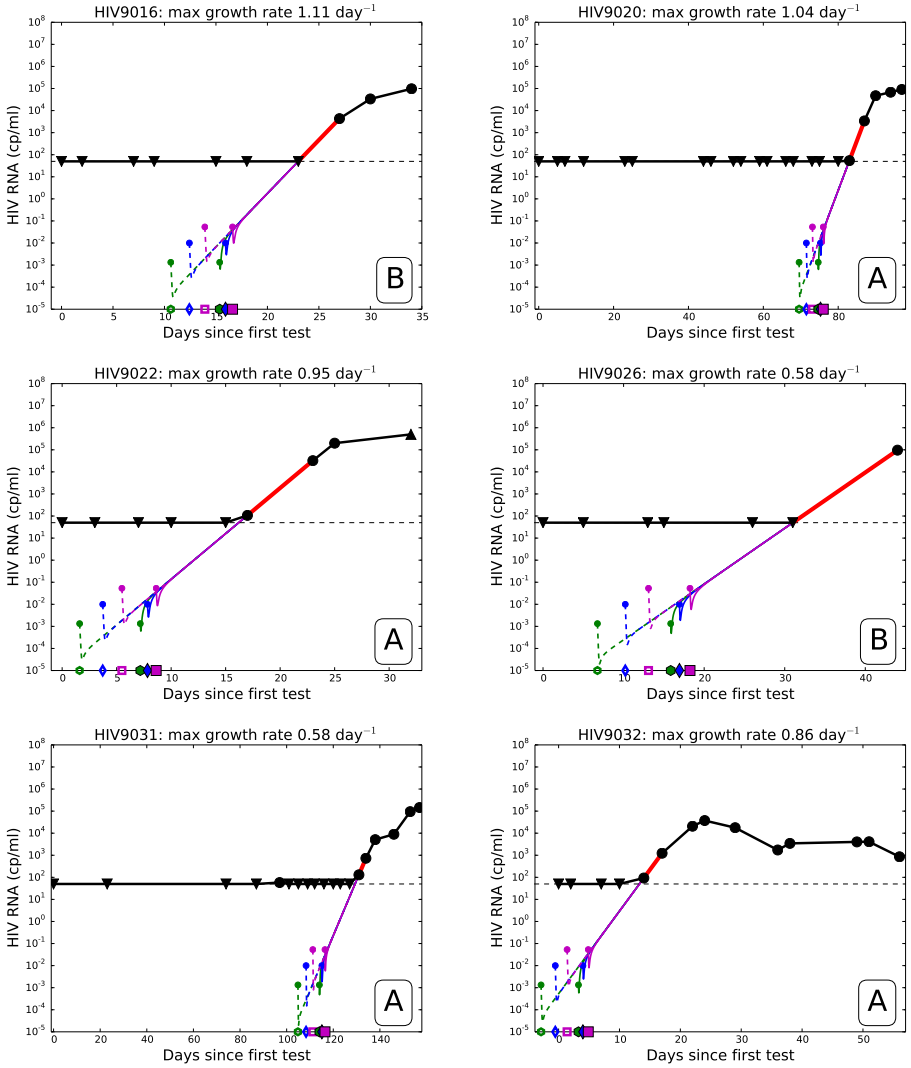

9028: max growth rate 0.36 day^{-1}
HIV RNA (cp/ml)
Days since first test
B

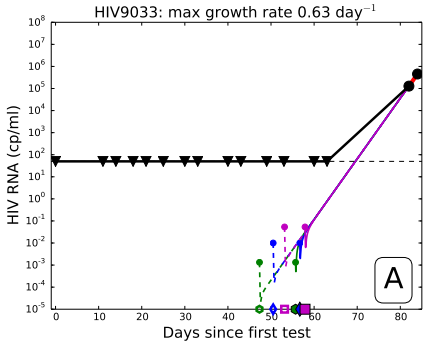

9029: max growth rate 0.63 day^{-1}
HIV RNA (cp/ml)
Days since first test
B

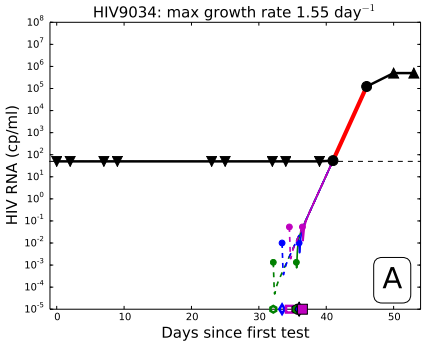

9030: max growth rate 1.17 day^{-1}
HIV RNA (cp/ml)
Days since first test
A

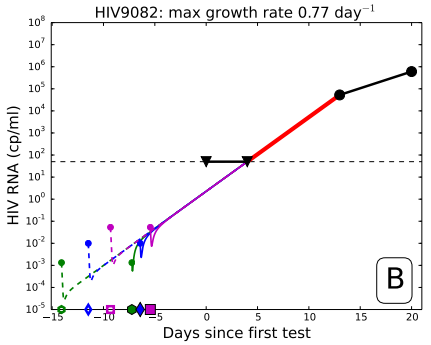

9075: max growth rate 0.37 day^{-1}
HIV RNA (cp/ml)
Days since first test
B

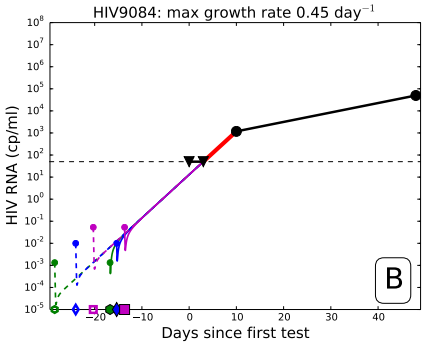

9077: max growth rate 1.02 day^{-1}
9079: max growth rate 1.41 day^{-1}
HIV6243: max growth rate 0.8 day^{-1}
HIV6248: max growth rate 0.99 day^{-1}
HIV9010: max growth rate 0.9 day^{-1}
HIV9013: max growth rate 1.3 day^{-1}
HIV RNA (cp/ml)
Days since first test
A
B

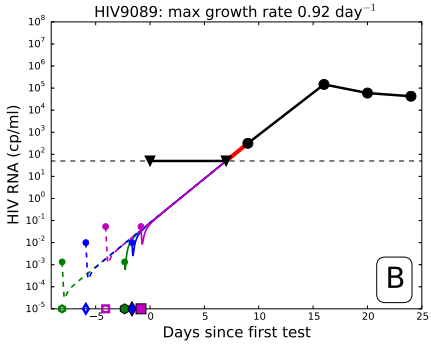

HIV9016: max growth rate 1.11 day^{-1}
HIV9020: max growth rate 1.04 day^{-1}
HIV9022: max growth rate 0.95 day^{-1}
HIV9026: max growth rate 0.58 day^{-1}
HIV9031: max growth rate 0.58 day^{-1}
HIV9032: max growth rate 0.86 day^{-1}
HIV RNA (cp/ml)
Days since first test
A
B

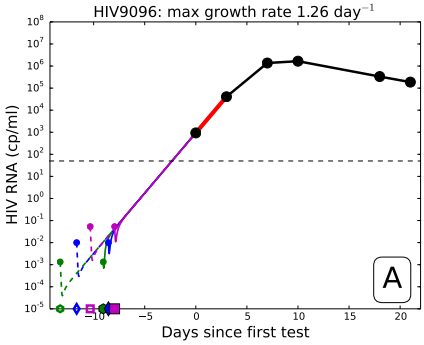

HIV9033: max growth rate 0.63 day^{-1}
HIV RNA (cp/ml)
Days since first test
A

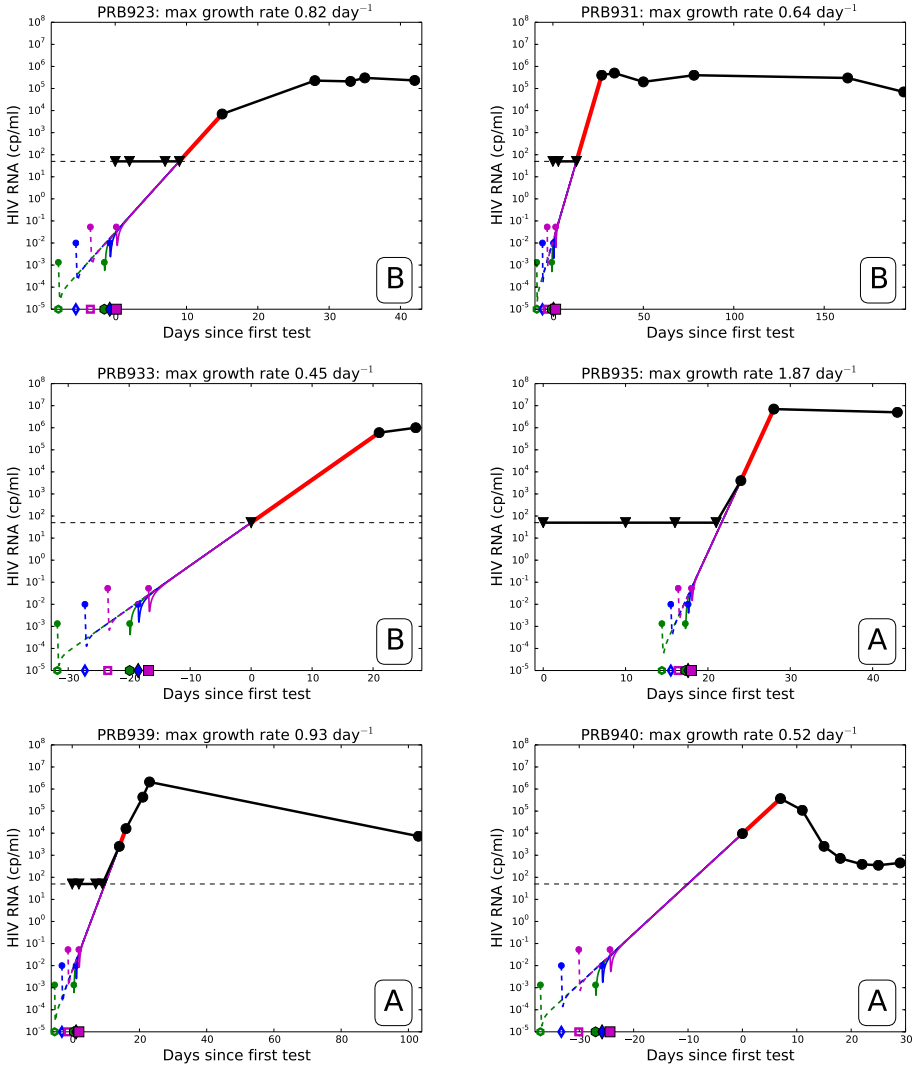

HIV9034: max growth rate 1.55 day^{-1}
HIV RNA (cp/ml)
Days since first test
A

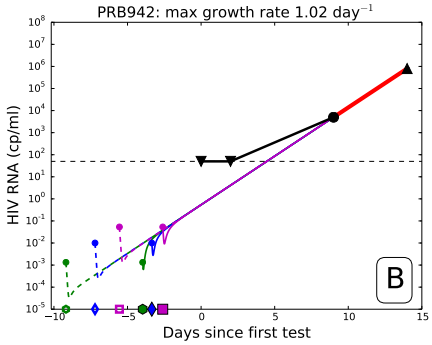

HIV9082: max growth rate 0.77 day^{-1}
HIV RNA (cp/ml)
Days since first test
B

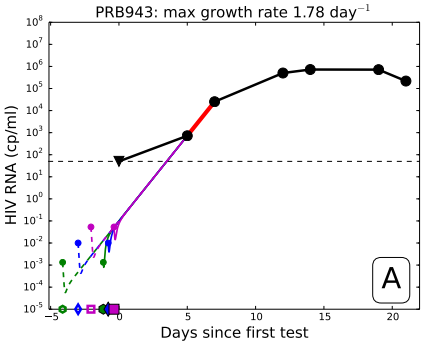

HIV9084: max growth rate 0.45 day^{-1}
HIV RNA (cp/ml)
Days since first test
B

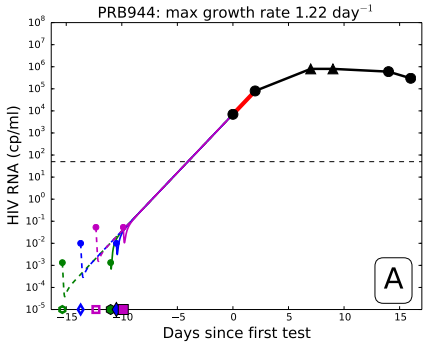

HIV9089: max growth rate 0.92 day^{-1}
HIV RNA (cp/ml)
Days since first test
B

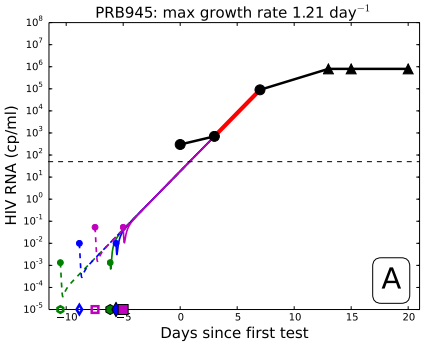

HIV9096: max growth rate 1.26 day^{-1}
HIV RNA (cp/ml)
Days since first test
A

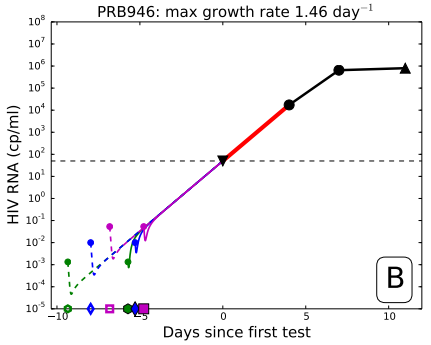

PRB923: max growth rate 0.82 day^{-1}
PRB931: max growth rate 0.64 day^{-1}
PRB933: max growth rate 0.45 day^{-1}
PRB935: max growth rate 1.87 day^{-1}
PRB939: max growth rate 0.93 day^{-1}
PRB940: max growth rate 0.52 day^{-1}
HIV RNA (cp/ml)
Days since first test
B
B
B
A
A
A

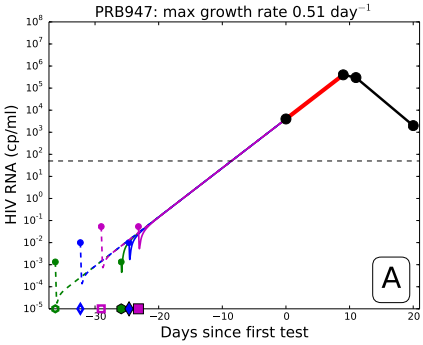

PRB942: max growth rate 1.02 day^{-1}
HIV RNA (cp/ml)
Days since first test
B

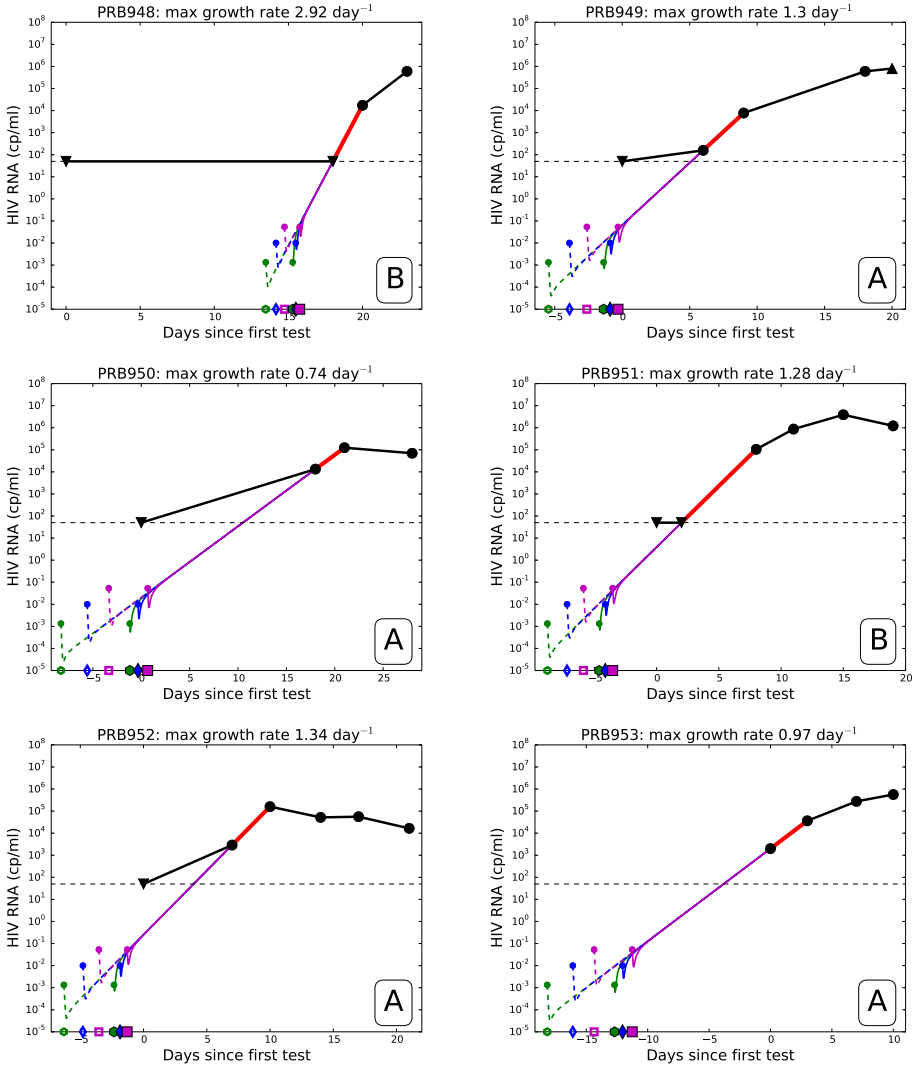

PRB943: max growth rate 1.78 day^{-1}
HIV RNA (cp/ml)
Days since first test
A

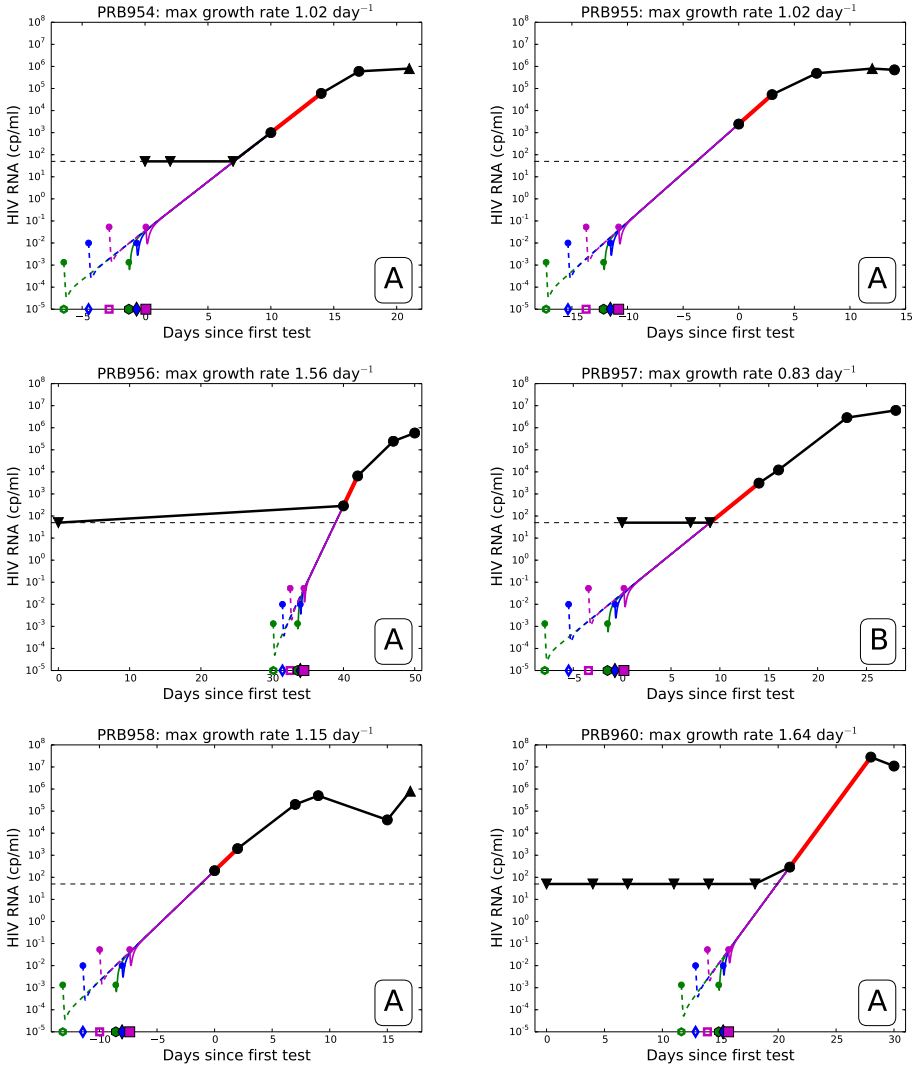

PRB944: max growth rate 1.22 day^{-1}
HIV RNA (cp/ml)
Days since first test
A

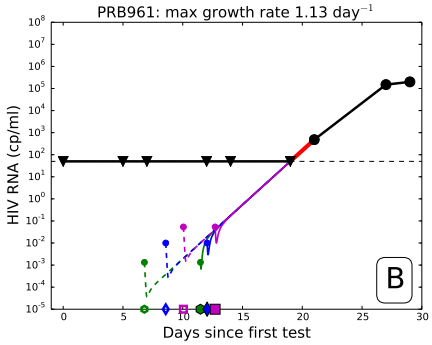

PRB945: max growth rate 1.21 day^{-1}
HIV RNA (cp/ml)
Days since first test
A

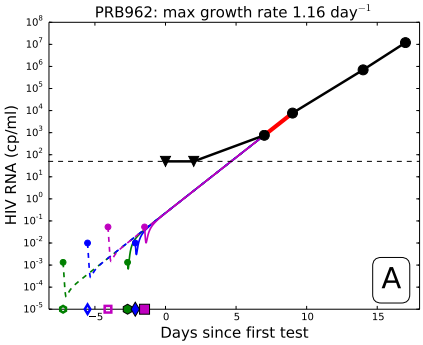

PRB946: max growth rate 1.46 day^{-1}
HIV RNA (cp/ml)
Days since first test
B

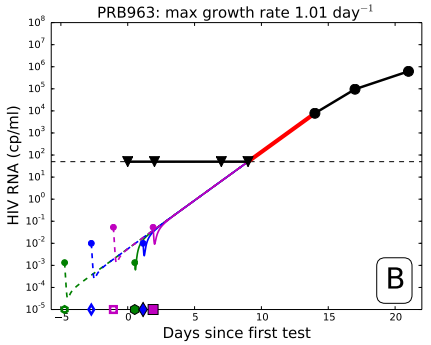

PRB947: max growth rate 0.51 day^{-1}
HIV RNA (cp/ml)
Days since first test
A

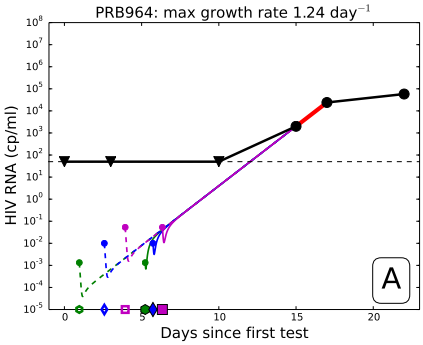

PRB948: max growth rate 2.92 day^{-1}
PRB949: max growth rate 1.3 day^{-1}
PRB950: max growth rate 0.74 day^{-1}
PRB951: max growth rate 1.28 day^{-1}
PRB952: max growth rate 1.34 day^{-1}
PRB953: max growth rate 0.97 day^{-1}
HIV RNA (cp/ml)
Days since first test
A
B

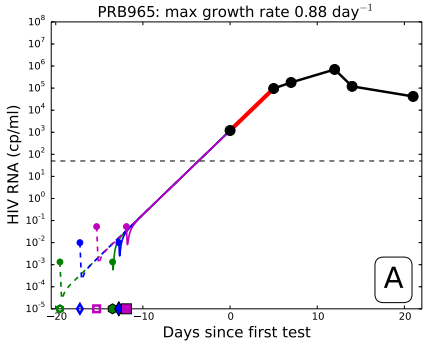

PRB954: max growth rate 1.02 day^{-1}
PRB955: max growth rate 1.02 day^{-1}
PRB956: max growth rate 1.56 day^{-1}
PRB957: max growth rate 0.83 day^{-1}
PRB958: max growth rate 1.15 day^{-1}
PRB960: max growth rate 1.64 day^{-1}
HIV RNA (cp/ml)
Days since first test
A
A
A
B
A
A

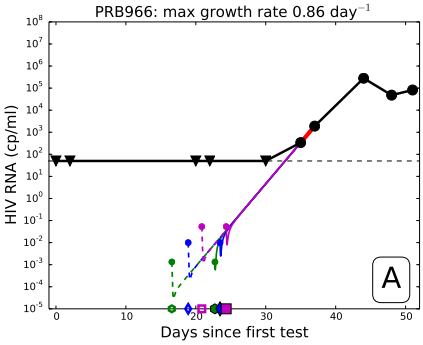

PRB961: max growth rate 1.13 day^{-1}
HIV RNA (cp/ml)
Days since first test
B

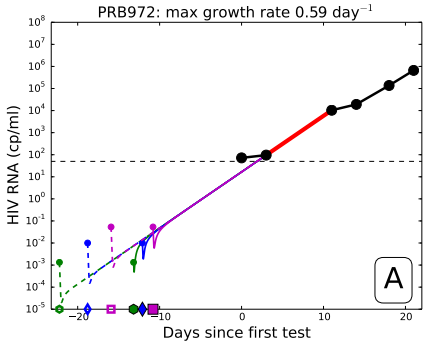

PRB962: max growth rate 1.16 day^{-1}
HIV RNA (cp/ml)
Days since first test
A

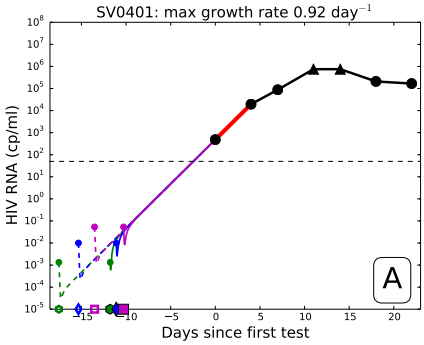

PRB963: max growth rate 1.01 day^{-1}
HIV RNA (cp/ml)
Days since first test
B


PRB964: max growth rate 1.24 day^{-1}
HIV RNA (cp/ml)
Days since first test
A


PRB965: max growth rate 0.88 day^{-1}
HIV RNA (cp/ml)
Days since first test
A


PRB966: max growth rate 0.86 day^{-1}
HIV RNA (cp/ml)
Days since first test
A


PRB972: max growth rate 0.59 day^{-1}
HIV RNA (cp/ml)
Days since first test
A


SV0401: max growth rate 0.92 day^{-1}
HIV RNA (cp/ml)
Days since first test
A

Appendix C

Additional material for malaria

The content of this appendix and the corresponding Chapter 5 stems from a course project in *MATH 561: Mathematics of Infectious Diseases and Immunology* taught 2011 at UBC, which was published in *Malaria Journal* [66]. I am the lead author of the publication, which is joint work with Michael Lindstrom, Anja Gumpinger, Jielin Zhu and Daniel Coombs. The journal article is distributed under the terms of the Creative Commons Attribution License (CC-BY) and has been modified slightly to fit the style and context of this thesis.

C.1 Proof of theorems

In this appendix we present the lengthy but straightforward proofs of some mathematical results used in Chapter 5.

C.1 Theorem (Local Stability).

1. *The trivial equilibrium* $(h,S,I,F) = (0,0,0,0)$ *is locally asymptotically stable if and only if* $\kappa < \frac{(\mu+\alpha)(\mu+\sigma)}{\mu+\alpha+\sigma}$.

2. *The malaria-free equilibrium* $(h,S,I,F) = (0,S_m,0,F_m)$ *exists in the positive plane and is locally asymptotically stable if and only if both of the following conditions are fulfilled:*

$$\kappa > \frac{(\mu+\alpha)(\mu+\sigma)}{\mu+\alpha+\sigma} \quad \textit{and} \quad R_0 = \beta\sqrt{\frac{\gamma S_m}{\rho(\mu+\alpha)H}} < 1.$$

Where $S_m = \frac{P(\mu+\sigma)}{\mu+\alpha+\sigma}(1-\frac{(\mu+\alpha)(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)})$ *and* $F_m = \frac{\alpha}{\mu+\sigma}S_m$.

3. *The endemic equilibrium* $(h,S,I,F) = (h_e,S_e,I_e,F_e)$ *exists in the positive plane and is locally asymptotically stable if and only if both of the following conditions are fulfilled:*

$$\kappa > \frac{(\mu+\alpha)(\mu+\sigma)}{\mu+\alpha+\sigma} \quad \textit{and} \quad R_0 = \beta\sqrt{\frac{\gamma S_m}{\rho(\mu+\alpha)H}} > 1.$$

Where $h_e = \frac{R_0^2-1}{R_0^2+\beta/(\mu+\alpha)}$, $S_e = (1-\frac{\beta}{\mu+\alpha+\beta}\frac{R_0^2-1}{R_0^2})S_m$, $I_e = \frac{\beta}{\mu+\alpha+\beta}\frac{R_0^2-1}{R_0^2}S_m$, $F_e = F_m$.

Proof: We first show how to get to the steady states and that the system indeed has exactly these three steady states. The trivial steady state $(h,S,I,F) = (0,0,0,0)$ is straightforward.

In the malaria-free steady state neither infected humans ($h_m = 0$), nor infected mosquitoes ($I_m = 0$) are present, but susceptible mosquitoes S_m and mosquitoes with fungus F_m are around. The equations reduce to

$$\begin{aligned}\frac{dS}{dt} &= \kappa(S+F)\left(1-\frac{S+F}{P}\right)-(\mu+\alpha)S\\ \frac{dF}{dt} &= \alpha S-(\mu+\sigma)F\end{aligned}$$

Setting the rate of change $\frac{dF}{dt}$ to zero tells us $F_m = \frac{\alpha}{\mu+\sigma}S_m$. Plugging this into the former equation, after we set $\frac{dS}{dt}$ to zero yields (note that $S_m + F_m = \frac{\mu+\alpha+\sigma}{\mu+\sigma}S_m$)

$$\begin{aligned}0 &= \kappa S_m\frac{\mu+\alpha+\sigma}{\mu+\sigma}\left(1-\frac{\mu+\alpha+\sigma}{\mu+\sigma}\frac{S_m}{P}\right)-(\mu+\alpha)S_m\\ 0 &= \frac{\kappa(\mu+\alpha+\sigma)}{\mu+\sigma}-\frac{\kappa(\mu+\alpha+\sigma)^2}{P(\mu+\sigma)^2}S_m-\frac{(\mu+\alpha)(\mu+\sigma)}{\mu+\sigma}\\ S_m &= \frac{P(\mu+\sigma)(\kappa(\mu+\alpha+\sigma)-(\mu+\alpha)(\mu+\sigma))}{\kappa(\mu+\alpha+\sigma)^2}\\ &= \frac{P(\mu+\sigma)}{\mu+\alpha+\sigma}\left(1-\frac{(\mu+\alpha)(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)}\right),\end{aligned}$$

where we divided by S_m since we assumed $S_m \neq 0$. If $S_m = 0$ we reach the trivial steady state. Therefore

$$F_m = \frac{\alpha}{\mu+\sigma} S_m = \frac{\alpha P}{\mu+\alpha+\sigma}\left(1 - \frac{(\mu+\alpha)(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)}\right).$$

This brings us to the endemic steady state, where h_e, S_e, I_e and F_e are all distinct from zero. (Note that $S_e = 0$ or $F_e = 0$ would lead us to the trivial steady state, since populations can not be negative.)

First, note that we can divide by $1-h_e$, since $h_e \neq 1$ due to the first equation in (1). Setting $\frac{dh}{dt} = 0$ yields $I_e = \frac{\rho H h_e}{\beta\gamma(1-h_e)}$. Setting $\frac{dI}{dt} = 0$ yields $I_e = \frac{\beta S_e h_e}{\mu+\alpha}$. Combining those those two equations yields $S_e = \frac{\rho H(\mu+\alpha)}{\beta^2\gamma(1-h_e)}$. Next, setting $\frac{dF}{dt} = 0$ we obtain

$$F_e = \frac{\alpha}{\mu+\sigma}(S_e + I_e) = \frac{\alpha\rho H}{\beta^2\gamma(\mu+\sigma)(1-h_e)}(\beta h_e + \mu + \alpha).$$

Therefore,

$$\begin{aligned} S_e + I_e + F_e &= \frac{\rho H}{\beta\gamma(1-h_e)}\left(\frac{\mu+\alpha}{\beta} + h_e + \frac{\alpha}{\beta(\mu+\sigma)}(\beta h_e + \mu + \alpha)\right) \\ &= \frac{\rho H}{\beta^2\gamma(1-h_e)}\left(\mu+\alpha+\beta h_e + \frac{\alpha}{\mu+\sigma}(\beta h_e + \mu + \alpha)\right) \\ &= \frac{\rho H(\mu+\alpha+\sigma)}{\beta^2\gamma(\mu+\sigma)(1-h_e)}(\beta h_e + \mu + \alpha) \end{aligned}$$

Plugging this into the equation $\frac{dS}{dt} = 0$ will lead to h_e, but not before a lengthy calculation:

$$\begin{aligned} 0 = \kappa &\frac{\rho H(\mu+\alpha+\sigma)}{\beta^2\gamma(\mu+\sigma)(1-h_e)}(\beta h_e + \mu + \alpha) \\ &\times \left(1 - \frac{\rho H(\mu+\alpha+\sigma)}{P\beta^2\gamma(\mu+\sigma)(1-h_e)}(\beta h_e + \mu + \alpha)\right) \\ &- \beta h_e \frac{\rho H(\mu+\alpha)}{\beta^2\gamma(1-h_e)} - (\mu+\alpha)\frac{\rho H(\mu+\alpha)}{\beta^2\gamma(1-h_e)} \end{aligned}$$

multiplying by $\frac{\beta^2\gamma(1-h_e)}{\rho H}$ yields

$$0 = \frac{\kappa(\mu+\alpha+\sigma)}{\mu+\sigma}(\beta h_e+\mu+\alpha)\left(1-\frac{\rho H(\mu+\alpha+\sigma)}{P\beta^2\gamma(\mu+\sigma)(1-h_e)}(\beta h_e+\mu+\alpha)\right) - (\mu+\alpha)(\beta h_e+\mu+\alpha).$$

Multiplying by $\frac{\mu+\sigma}{\kappa(\mu+\alpha+\sigma)(\beta h_e+\mu+\alpha)}$, which is positive since h_e is positive, and rearranging we obtain

$$\frac{\rho H(\mu+\alpha+\sigma)(\beta h_e+\mu+\alpha)}{P\beta^2\gamma(\mu+\sigma)(1-h_e)} = 1-\frac{(\mu+\alpha)(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)}$$

To solve this for h_e we multiply by the denominator on the left hand side and expand the resulting terms:

$$\begin{aligned}&\rho H(\mu+\alpha+\sigma)\beta h_e+\rho H(\mu+\alpha+\sigma)(\mu+\alpha)\\ &\quad = P\beta^2\gamma(\mu+\sigma)-P\beta^2\gamma(\mu+\sigma)h_e\\ &\qquad -\frac{P\beta^2\gamma(\mu+\alpha)(\mu+\sigma)^2}{\kappa(\mu+\alpha+\sigma)}+\frac{P\beta^2\gamma(\mu+\alpha)(\mu+\sigma)^2}{\kappa(\mu+\alpha+\sigma)}h_e,\end{aligned}$$

which we rearrange to

$$\begin{aligned}&h_e\left[\rho\beta H(\mu+\alpha+\sigma)+P\beta^2\gamma(\mu+\sigma)\left(1-\frac{(\mu+\alpha)(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)}\right)\right]\\ &\qquad = P\beta^2\gamma(\mu+\sigma)\left(1-\frac{(\mu+\alpha)(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)}\right)-\rho H(\mu+\alpha+\sigma)(\mu+\alpha).\end{aligned}$$

We now divide by $\rho H(\mu+\alpha)(\mu+\alpha+\sigma)$ and recall that

$$R_0^2 = \frac{\beta^2\gamma S_m}{\rho(\mu+\alpha)H} = \frac{P\beta^2\gamma(\mu+\sigma)}{\rho H(\mu+\alpha)(\mu+\alpha+\sigma)}(1-\frac{(\mu+\alpha)(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)}).$$

This results in

$$h_e\left(R_0^2+\frac{\beta}{\mu+\alpha}\right) = R_0^2-1, \quad \text{or equivalently} \quad h_e = \frac{R_0^2-1}{R_0^2+\beta/(\mu+\alpha)}.$$

From here we can quickly find S_e, I_e and F_e:

$$S_e = \frac{\rho H(\mu+\alpha)}{\beta^2\gamma(1-h_e)} = \frac{S_m}{R_0^2(1-h_e)} = \frac{S_m}{R_0^2}\frac{R_0^2+\frac{\beta}{\mu+\alpha}}{1+\frac{\beta}{\mu+\alpha}} = S_m\left(1-\frac{\beta}{\beta+\mu+\alpha}\frac{R_0^2-1}{R_0^2}\right)$$

$$I_e = \frac{\rho H h_e}{\beta\gamma(1-h_e)} = \frac{\beta S_m}{R_0^2(\mu+\alpha)}\frac{h_e}{1-h_e} = \frac{\beta S_m}{R_0^2(\mu+\alpha)}\frac{R_0^2-1}{R_0^2+\frac{\beta}{\mu+\alpha}}\frac{R_0^2+\frac{\beta}{\mu+\alpha}}{1+\frac{\beta}{\mu+\alpha}}$$

$$= S_m\frac{\beta}{\beta+\mu+\alpha}\frac{R_0^2-1}{R_0^2}$$

$$F_e = \frac{\alpha}{\mu+\sigma}(S_e+I_e) = \frac{\alpha}{\mu+\sigma}S_m = F_m$$

To determine the local stability of the steady states we compute the characteristic polynomial of the linearized system around the equilibrium. To simplify the notation we abbreviate $\Upsilon = \kappa(1-2(S+I+F)/P)$.

$$\begin{vmatrix} -\beta\gamma\frac{I}{H}-\rho-\lambda & 0 & \beta\gamma\frac{1}{H}(1-h) & 0 \\ -\beta S & \Upsilon-\beta h-\mu-\alpha-\lambda & \Upsilon & \Upsilon \\ \beta S & \beta h & -(\mu+\alpha)-\lambda & 0 \\ 0 & \alpha & \alpha & -(\mu+\sigma)-\lambda \end{vmatrix}$$

$$= \begin{vmatrix} -\beta\gamma\frac{I}{H}-\rho-\lambda & 0 & \beta\gamma\frac{1}{H}(1-h) & 0 \\ 0 & \Upsilon-(\mu+\alpha)-\lambda & \Upsilon-(\mu+\alpha)-\lambda & \Upsilon \\ \beta S & \beta h & -(\mu+\alpha)-\lambda & 0 \\ 0 & \alpha & \alpha & -(\mu+\sigma)-\lambda \end{vmatrix}$$

$$= \begin{vmatrix} -\beta\gamma\frac{I}{H}-\rho-\lambda & 0 & \beta\gamma\frac{1}{H}(1-h) & 0 \\ 0 & \Upsilon-(\mu+\alpha)-\lambda & 0 & \Upsilon \\ \beta S & \beta h & -(\mu+\alpha)-\beta h-\lambda & 0 \\ 0 & \alpha & 0 & -(\mu+\sigma)-\lambda \end{vmatrix}$$

$$= \underbrace{\begin{vmatrix} -\beta\gamma_H^I - \rho - \lambda & \beta\gamma_H^1(1-h) \\ \beta S & -(\mu+\alpha) - \beta h - \lambda \end{vmatrix}}_{=:A} \times \underbrace{\begin{vmatrix} \kappa(1-\Upsilon) - (\mu+\alpha) - \lambda & \kappa(1-\Upsilon) \\ \alpha & -(\mu+\sigma) - \lambda \end{vmatrix}}_{=:B}$$

The fourth-order polynomial factors nicely into two second-order polynomials. This is to be expected: Polynomial A says if malaria can invade the population, polynomial B decides if mosquitoes can invade the environment.

We now split the analysis into the three equilibria:

1. **Trivial equilibrium** $(h,S,I,F) = (0,0,0,0)$**.** Then A becomes

$$A = \det \begin{vmatrix} -\rho - \lambda & \beta\gamma_H^1 \\ 0 & -(\mu+\alpha) - \lambda \end{vmatrix} = (\rho+\lambda)(\mu+\alpha+\lambda),$$

which clearly has the two negative roots $\lambda = -\rho$ and $\lambda = -(\mu+\alpha)$. Hence, malaria can not invade a void population and the stability of the trivial steady state is solely determined by B.

B becomes

$$\begin{aligned} B &= \begin{vmatrix} \kappa - (\mu+\alpha) - \lambda & \kappa \\ \alpha & -(\mu+\sigma) - \lambda \end{vmatrix} \\ &= ((\kappa - (\mu+\alpha)) - \lambda)((-(\mu+\sigma)) - \lambda) - \alpha\kappa \\ &= \lambda^2 + [(\mu+\sigma) + (\mu+\alpha) - \kappa]\lambda + (\mu+\alpha)(\mu+\sigma) - \kappa(\alpha+\mu+\sigma) \end{aligned}$$

The constant term is positive if and only if

$$\kappa < \frac{(\mu+\alpha)(\mu+\sigma)}{\mu+\alpha+\sigma} = (\mu+\alpha)\frac{\mu+\sigma}{\mu+\alpha+\sigma} \tag{C.1}$$

If this condition is not fulfilled, then there is a positive root and hence the steady state is not stable. If this condition is fulfilled, then $\kappa < (\mu+\alpha) < (\mu+\alpha)+(\mu+\sigma)$ and hence the linear term is also positive. Thus only

negative roots exist and the steady state is table. Hence the trivial steady state is stable if and only if the condition (C.1) on κ holds.

2. **Malaria-free equilibrium** $(h,S,I,F) = (0,S_m,0,F_m)$**.**

   In the malaria-free equilibrium A becomes

$$A = \det \begin{vmatrix} -\rho-\lambda & \beta\gamma\frac{1}{H} \\ \beta S_m & -(\mu+\alpha)-\lambda \end{vmatrix} = (-\rho-\lambda)(-(\mu+\alpha)-\lambda) - \beta^2\gamma\frac{S_m}{H}$$
$$= \lambda^2 + (\rho+\mu+\alpha)\lambda + \rho(\mu+\alpha) - \beta^2\gamma\frac{S_m}{H}.$$

   This polynomial has only negative roots if and only if the constant term is positive, i.e. if and only if

$$\rho(\mu+\alpha) - \beta^2\gamma\frac{S_m}{H} > 0,$$

   which is equivalent to $R_0 = \beta\sqrt{\frac{\gamma S_m}{\rho(\mu+\alpha)H}} < 1$.

   The polynomial B becomes

$$B = \det \begin{vmatrix} \kappa\left(1-\frac{2(S_m+F_m)}{P}\right) - (\mu+\alpha) - \lambda & \kappa\left(1-\frac{2(S_m+F_m)}{P}\right) \\ \alpha & -(\mu+\sigma)-\lambda \end{vmatrix}$$

   Before we go about calculating this, let us first simplify the first row a little bit:

$$\begin{aligned} S_m + F_m &= S_m + \frac{\alpha}{\mu+\sigma}S_m = \frac{\mu+\alpha+\sigma}{\mu+\sigma}S_m \\ &= P\frac{\kappa(\mu+\alpha+\sigma) - (\mu+\alpha)(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)} \\ &= P\left(1 - \frac{(\mu+\alpha)(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)}\right) \end{aligned}$$

   With that we obtain

$$\kappa\left(1-\frac{2(S_m+F_m)}{P}\right) = \kappa\left(1-2\left(1-\frac{(\mu+\alpha)(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)}\right)\right)$$

$$= 2\frac{(\mu+\alpha)(\mu+\sigma)}{\mu+\alpha+\sigma} - \kappa.$$

Therefore

$$\begin{aligned}
B &= \det \begin{vmatrix} \kappa\left(1-\frac{2(S_m+F_m)}{P}\right) - (\mu+\alpha) - \lambda & \kappa\left(1-\frac{2(S_m+F_m)}{P}\right) \\ \alpha & -(\mu+\sigma)-\lambda \end{vmatrix} \\
&= \det \begin{vmatrix} 2\frac{(\mu+\alpha)(\mu+\sigma)}{\mu+\alpha+\sigma} - \kappa - (\mu+\alpha) - \lambda & 2\frac{(\mu+\alpha)(\mu+\sigma)}{\mu+\alpha+\sigma} - \kappa \\ \alpha & -(\mu+\sigma)-\lambda \end{vmatrix} \\
&= \left(2\frac{(\mu+\alpha)(\mu+\sigma)}{\mu+\alpha+\sigma} - \kappa - (\mu+\alpha) - \lambda\right)(-(\mu+\sigma)-\lambda) \\
&\quad - \alpha\left(2\frac{(\mu+\alpha)(\mu+\sigma)}{\mu+\alpha+\sigma} - \kappa\right) \\
&= \lambda^2 + \lambda\left[\kappa + (\mu+\alpha) + (\mu+\sigma) - 2\frac{(\mu+\alpha)(\mu+\sigma)}{\mu+\alpha+\sigma}\right] \\
&\quad + \kappa(\mu+\alpha+\sigma) + (\mu+\alpha)(\mu+\sigma) - 2\frac{(\mu+\alpha)(\mu+\sigma)}{\mu+\alpha+\sigma}(\mu+\alpha+\sigma) \\
&= \lambda^2 + \lambda\left[\kappa + (\mu+\alpha)\left(1-\frac{\mu+\sigma}{\mu+\alpha+\sigma}\right) + (\mu+\sigma)\left(1-\frac{\mu+\alpha}{\mu+\alpha+\sigma}\right)\right] \\
&\quad + \kappa(\mu+\alpha+\sigma) - (\mu+\alpha)(\mu+\sigma).
\end{aligned}$$

It is now obvious that this polynomial has two roots with negative real part, if only if the constant term is positive, i.e. if

$$\kappa > \frac{(\mu+\alpha)(\mu+\sigma)}{\mu+\alpha+\sigma}.$$

3. **Endemic equilibrium** $(h,S,I,F) = (h_e,S_e,I_e,F_e)$**.**

   In the endemic steady state the polynomial A becomes

$$\begin{aligned}
A &= \begin{vmatrix} -\beta\gamma\frac{I_e}{H} - \rho - \lambda & \beta\gamma\frac{1}{H}(1-h_e) \\ \beta S_e & -(\mu+\alpha) - \beta h_e - \lambda \end{vmatrix} \\
&= \lambda^2 + \lambda\left[\beta\gamma\frac{I_e}{H} + \rho + \mu + \alpha + \beta h_e\right] \\
&\quad + \beta\gamma\frac{I_e}{H}(\mu+\alpha) + \beta^2\gamma\frac{I_e}{H}h_e + \rho(\mu+\alpha) + \beta\rho h_e - \beta^2\gamma\frac{S_e}{H} + \beta^2\gamma\frac{S_e}{H}h_e
\end{aligned}$$

$$= \lambda^2 + \lambda \left[\beta\gamma \frac{I_e}{H} + \rho + \mu + \alpha + \beta h_e \right] + \beta\gamma \frac{I_e}{H}(\mu+\alpha) + \rho(\mu+\alpha) + \beta^2\gamma \frac{I_e}{H} - \beta^2\gamma \frac{S_e}{H}(1-h_e),$$

where we used that $\rho h_e = \beta\gamma \frac{I_e}{H}(1-h_e)$. In the constant term we now substitute S_e and I_e with the expressions we just found above, and also use the definition of R_0. Factoring the result yields

$$\begin{aligned} A &= \lambda^2 + \lambda \left[\beta\gamma \frac{I_e}{H} + \rho + \mu + \alpha + \beta h_e \right] \\ &\quad + \frac{\beta^2\gamma S_m}{H} \left(\frac{\mu+\alpha}{\beta+\mu+\alpha} \frac{R_0^2-1}{R_0^2} + \frac{1}{R_0^2} + \frac{\beta}{\beta+\mu+\alpha} \frac{R_0^2-1}{R_0^2} \right. \\ &\qquad \left. - (1-h_e) \left(1 - \frac{\beta}{\beta+\mu+\alpha} \frac{R_0^2-1}{R_0^2} \right) \right) \\ &= \lambda^2 + \lambda \left[\beta\gamma \frac{I_e}{H} + \rho + \mu + \alpha + \beta h_e \right] \\ &\quad + R_0^2 \rho(\mu+\alpha) \left(h_e + (1-h_e) \frac{\beta}{\beta+\mu+\alpha} \frac{R_0^2-1}{R_0^2} \right), \end{aligned}$$

where we again used the definition of R_0. We now see that, if $R_0 > 1$, then all roots of A have negative real part.

For B, note that $S_e + I_e + F_e = S_m + F_m$, and hence B is the same as in the malaria-free steady state.

C.2 Theorem (Fungus exposure rate).

The malaria prevalence in humans $h_e = h_e(\alpha, \sigma)$ decreases when the fungus-exposure rate α increases. The total number of mosquitoes $S_e + I_e + F_e = (S_e + I_e + F_e)(\alpha, \sigma)$ also decreases when α increases, unless the fungus-induced death rate σ is zero, in which case the total number of mosquitoes is unaffected by α.

Proof: Before we prove the claim, we first show that $S_m = S_m(\alpha, \sigma)$ is a decreasing function of α is, i.e.

$$\frac{\partial S_m}{\partial \alpha}(\alpha, \sigma) < 0$$

for all $\alpha, \sigma \geq 0$. A straightforward calculation yields

$$\begin{aligned}
\frac{\partial S_m}{\partial \alpha}(\alpha,\sigma) &= \frac{\partial}{\partial \alpha}\left[\frac{P(\mu+\sigma)}{\mu+\alpha+\sigma}\left(1-\frac{(\mu+\alpha)(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)}\right)\right] \\
&= -\frac{P(\mu+\sigma)}{(\mu+\alpha+\sigma)^2}\left[\left(1-\frac{(\mu+\alpha)(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)}\right)\right. \\
&\qquad\qquad \left. + \frac{\kappa(\mu+\alpha+\sigma)(\mu+\sigma)-\kappa(\mu+\alpha)(\mu+\sigma)}{\kappa^2(\mu+\alpha+\sigma)}\right] \\
&= -\frac{P(\mu+\sigma)}{(\mu+\alpha+\sigma)^2}\left(1-2\frac{(\mu+\alpha)(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)}+\frac{\mu+\sigma}{\kappa}\right) \\
&= -\frac{P(\mu+\sigma)}{(\mu+\alpha+\sigma)^2}\left(\frac{\mu+\sigma}{\kappa}\left(1-\frac{\mu+\alpha}{\mu+\alpha+\sigma}\right)\right. \\
&\qquad\qquad \left. + \left(1-\frac{(\mu+\alpha)(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)}\right)\right) \\
&< 0,
\end{aligned}$$

since $\kappa > \frac{(\mu+\alpha)(\mu+\sigma)}{\mu+\alpha+\sigma}$ by the assumption that malaria is endemic.

Now, let $0 \leq \alpha_1 < \alpha_2$ and fix $\sigma \geq 0$. For notational convenience we omit the argument σ in the following calculation. From the calculation above we can further deduce that

$$S_m(\alpha_1) > S_m(\alpha_2) \Longrightarrow R_0^2(\alpha_1) = \frac{\beta^2\gamma S_m(\alpha_1)}{\rho(\mu+\alpha_1)H} > \frac{\beta^2\gamma S_m(\alpha_2)}{\rho(\mu+\alpha_2)H} = R_0^2(\alpha_2).$$

Finally, we are now in position to show the first claim:

$$\begin{aligned}
& h_e(\alpha_2) < h_e(\alpha_1) \\
\Leftrightarrow \quad & \frac{R_0^2(\alpha_2)-1}{R_0^2(\alpha_2)+\beta/(\mu+\alpha_2)} < \frac{R_0^2(\alpha_1)-1}{R_0^2(\alpha_1)+\beta/(\mu+\alpha_1)} \\
\Leftrightarrow \quad & 0 < (\mu+\alpha_1)(\mu+\alpha_2)(R_0^2(\alpha_1)-R_0^2(\alpha_2)) \\
& \qquad + \beta(R_0^2(\alpha_1)(\mu+\alpha_1)-R_0^2(\alpha_2)(\mu+\alpha_2))+\beta(\alpha_2-\alpha_1) \\
\Leftrightarrow \quad & 0 < (\mu+\alpha_1)(\mu+\alpha_2)(R_0^2(\alpha_1)-R_0^2(\alpha_2))
\end{aligned}$$

$$+\beta\frac{\beta^2\gamma}{\rho H}(S_m(\alpha_1)-S_m(\alpha_2))+\beta(\alpha_2-\alpha_1).$$

The former results show that the last line is true, hence the first claim follows. To see that the total number of mosquitoes $S_e+I_e+F_e$ also decreases with α, we calculate that

$$\begin{aligned}
\frac{\partial S_e+I_e+F_e}{\partial\alpha}(\alpha,\sigma) &= \frac{\partial}{\partial\alpha}\frac{\mu+\alpha+\sigma}{\mu+\sigma}S_m(\alpha,\sigma)\\
&= \frac{1}{\mu+\sigma}S_m(\alpha,\sigma)+\frac{\mu+\alpha+\sigma}{\mu+\sigma}\frac{\partial S_m}{\partial\alpha}(\alpha,\sigma)\\
&= \frac{P}{\mu+\alpha+\sigma}\left(1-\frac{(\mu+\alpha)(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)}\right)\\
&\qquad -\frac{P}{\mu+\alpha+\sigma}\left(\frac{\mu+\sigma}{\kappa}\left(1-\frac{\mu+\alpha}{\mu+\alpha+\sigma}\right)\right.\\
&\qquad\qquad \left.+\left(1-\frac{(\mu+\alpha)(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)}\right)\right)\\
&= -\frac{P\sigma(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)^2}\\
&\leq 0,
\end{aligned}$$

which only vanishes when $\sigma=0$ and hence proves the second claim.

C.3 Theorem (Fungus-induced death rate on mosquitoes).
For $\alpha>0$ the total number of mosquitoes, $S_e+I_e+F_e$, decreases if the fungus-induced death rate σ increases.

Proof: This is a straightforward calculation. Recall $S_e+I_e+F_e=S_m+F_m=S_m(1+\frac{\alpha}{\mu+\sigma})=P(1-\frac{(\mu+\alpha)(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)})$ and calculate

$$\begin{aligned}
\frac{\partial(S_e+I_e+F_e)}{\partial\sigma}(\alpha,\sigma) &= -P\frac{\kappa(\mu+\alpha+\sigma)(\mu+\alpha)-\kappa(\mu+\alpha)(\mu+\sigma)}{\kappa^2(\mu+\alpha+\sigma)^2}\\
&= -\frac{P\alpha(\mu+\alpha)}{\kappa(\mu+\alpha+\sigma)^2},
\end{aligned}$$

which is negative for all $\alpha>0$.

C.4 Theorem (Fungus-induced death rate on malaria prevalence in humans).
If $\alpha = 0$ (i.e. no fungus spraying) the human malaria incidence rate $h_e = h_e(0,\sigma)$ is constant in σ. If $\alpha > 0$ (i.e. the fungus is applied) the malaria prevalence in the human population $h_e(\alpha,\sigma)$, as well as the total number of infected mosquitoes $I_e(\alpha,\sigma)$, both have their maximum at

$$\sigma^* = \frac{(\kappa - 2\mu)(\mu+\alpha)}{(2\mu-\kappa)+2\alpha}.$$

Proof: Let $\alpha \geq 0$ be fixed. To find the local maximum of $h_e(\alpha,\sigma)$ we first calculate the derivative with respect to σ. This yields

$$\frac{\partial h_e}{\partial \sigma}(\alpha,\sigma) = \frac{1+\beta/(\mu+\alpha)}{(R_0^2(\alpha,\sigma)+\beta/(\mu+\alpha))^2}\frac{\partial R_0^2}{\partial \sigma}(\alpha,\sigma)$$

so that we see that $h_e(\alpha,\sigma)$ has a critical point whenever $R_0^2(\alpha,\sigma)$ has a critical point. However, $R_0^2(\alpha,\sigma)$ in turn a critical point whenever $S_m(\alpha,\sigma)$ has a critical point:

$$\frac{\partial R_0^2}{\partial \sigma}(\alpha,\sigma) = \frac{\beta^2\gamma}{\rho(\mu+\alpha)H}\frac{\partial S_m}{\partial \sigma}(\alpha,\sigma).$$

Thus we first study the critical points of $S_m(\alpha,\sigma)$.

$$\begin{aligned}\frac{\partial S_m}{\partial \sigma}(\alpha,\sigma) = P\Bigg[&\frac{\alpha}{(\mu+\alpha+\sigma)^2}\left(1-\frac{(\mu+\alpha)(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)}\right)\\ &-\frac{\mu+\sigma}{\mu+\alpha+\sigma}\frac{(\mu+\alpha+\sigma)(\mu+\alpha)-(\mu+\sigma)(\mu+\alpha)}{\kappa(\mu+\alpha+\sigma)^2}\Bigg]\\ &= \frac{P\alpha}{(\mu+\alpha+\sigma)^2}\left(1-2\frac{(\mu+\alpha)(\mu+\sigma)}{\kappa(\mu+\alpha+\sigma)}\right).\end{aligned}$$

Now we see that if $\alpha = 0$, then $S_m(0,\sigma)$, and hence $h_e(0,\sigma)$ is constant for all values of σ. So from now on let $\alpha > 0$. Then the value σ^* that makes $\frac{\partial}{\partial\sigma}S_m(\alpha,\sigma^*) = 0$ solves

$$1 = 2\frac{(\mu+\alpha)(\mu+\sigma^*)}{\kappa(\mu+\alpha+\sigma^*)}$$

$$\Leftrightarrow \qquad \sigma^*(\kappa - 2(\mu+\alpha)) = (\mu+\alpha)(2\mu-\kappa)$$

If $\kappa = 2(\mu+\alpha)$, then the equation above yields $0 = (\mu+\alpha)\alpha$, which is only true for $\alpha = 0$, since $\mu > 0$. As we have asserted before, h_2 is constant in this case. Therefore, if $\kappa = 2(\mu+\alpha)$ no maximum exists. If $\kappa \neq 2(\mu+\alpha)$ we obtain

$$\sigma^* = \frac{(\kappa-2\mu)(\mu+\alpha)}{(2\mu-\kappa)+2\alpha},$$

i.e. $\frac{\partial S_m}{\partial \sigma}(\alpha,\sigma^*) = 0$ and thus $\frac{\partial R_0^2}{\partial \sigma}(\alpha,\sigma^*) = 0 = \frac{\partial h_e}{\partial \sigma}(\alpha,\sigma^*)$. We use the second derivative test to quickly check that the value at (α,σ^*) is indeed a maximum:

$$\begin{aligned}
\frac{\partial^2 S_m}{\partial \sigma^2}(\alpha,\sigma^*) &= \frac{2P\alpha}{(\mu+\alpha+\sigma^*)^3}\Bigg(\overbrace{-1+2\frac{(\mu+\alpha)(\mu+\sigma^*)}{\kappa(\mu+\alpha+\sigma^*)}}^{=0} \\
&\qquad - \frac{(\mu+\alpha+\sigma^*)(\mu+\alpha)-(\mu+\sigma^*)(\mu+\alpha)}{\kappa(\mu+\alpha+\sigma^*)}\Bigg) \\
&= -2\frac{P\alpha(\mu+\alpha)}{\kappa(\mu+\alpha+\sigma^*)^3}\left(1-\frac{\mu+\sigma^*}{\mu+\alpha+\sigma^*}\right) \\
&< 0.
\end{aligned}$$

Since $\frac{\partial S_m}{\partial \sigma}(\alpha,\sigma)$ and and $\frac{\partial h_e}{\partial \sigma}(\alpha,\sigma)$ have the same sign everywhere this implies that h_e also has its maximum at (α,σ^*). Finally, I_e is an increasing function of h_e and hence is also maximized at (α,σ^*).

Parameters

We choose some reasonable parameter values to illustrate our analytical results. Consider a small village with constant population size of $H = 3000$ humans. It is fair to assume that female mosquitoes typically outnumber humans by a factor of three to ten (see [19, 101]), hence we choose the carrying capacity of mosquitoes to be $\tilde{P} = 10000$. About half of bites from infectious mosquitoes lead to an infection in a human [18, 19], i.e. $\gamma = 0.5$.

The biting rate β is difficult to measure as it depends on many factors, in partic-

ular (seasonal) human behavior. In our simple model we assume a constant mass-action rate. The value used in [101], $\beta\gamma\tilde{P}/H = 0.5$ would translates to $\beta = 0.3$ day^{-1}, which is about the maximal number of times one mosquito would want to bite humans per unit time, if humans were freely available, see [18]. Since it is reasonable to assume that humans are somewhat not always freely available in a village we choose $\beta = 0.15$ days^{-1}. This choice does not significantly affect the qualitative results of the study. Infected humans leave the infectious class by natural or malaria-induced death or when they develop (temporal) immunity. The latter is the dominating contributor, and usually happens after a few months to a year, hence we choose an average of 200 days, i.e. $\rho = 0.005$ days^{-1} (see [18, 64]). The mean life-time of an outgrown, biting, female mosquito is around ten days [64, 101] and hence the mosquito death rate is chosen to be $\mu = 0.1$ days^{-1}.

A typical mosquito population is not constant throughout a year but is heavily influenced by external factors such as rainfall and humidity. A small mosquito population at end of a dry season typically experiences a fast growth at the beginning of the rain season. Here is field data from [2], where the number of mosquitoes caught in a trap close to a village in Nigeria was measured between August 2005 and July 2006:

	Aug	Sep	Oct	Nov	Dec	Jan
Anopheles gambiae	23	24	25	21	9	5

	Feb	Mar	Apr	May	Jun	Jul
Anopheles gambiae	12	26	47	50	76	46

The growth phase from January until June can clearly be seen. We assume that, in this time period, the environmental conditions are so optimal for the mosquito population, that it experiences exponential growth. So we let $M = S + I$ be the total number of mosquitoes, set $\alpha = 0$, hence $F = 0$, since in the data no fungus was applied, and fit the data points above to the simplified model

$$\frac{dM}{dt} = (\kappa - \mu)M, \qquad M(0) = M_0.$$

Then we perform a least-square fit optimization for $(\kappa - \mu)$ and M_0 to obtain $(\kappa - \mu) \approx 0.38$ days^{-1}, or simply $\kappa = 0.48$ days^{-1}.

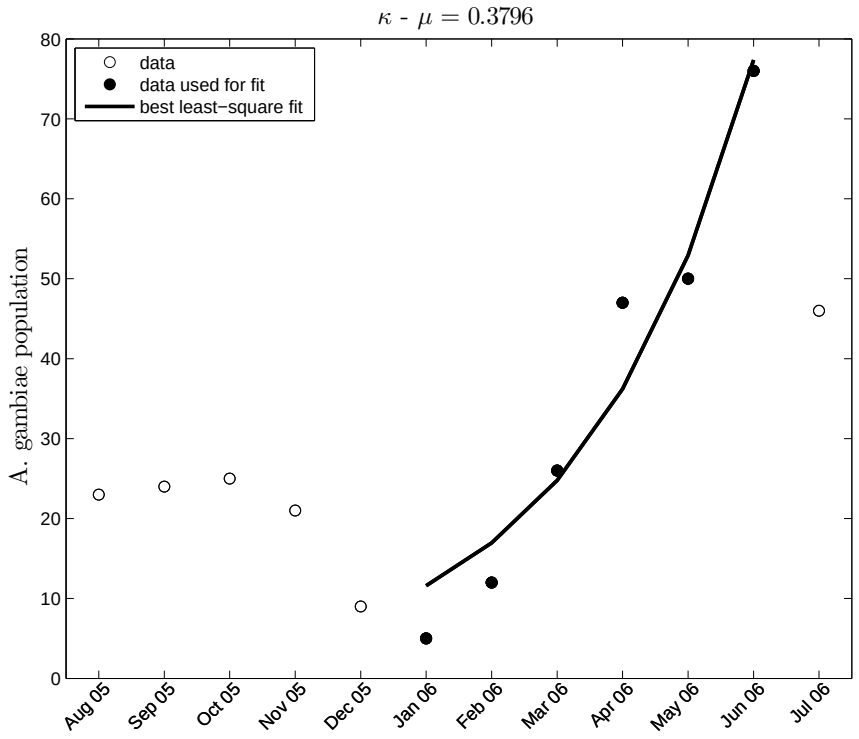


Figure C.1: Data fitting for κ. We fit an exponential function for six of the data points reported in [2] to find an approximation for $\kappa - \mu$.

In the life-stage-structured model we need to estimate the larval production rate κ_c and intra-specific larval competition parameter κ_L. For simplicity, we set $\kappa_c = \kappa$ as estimated above, and then choose κ_L so that the number of mosquitoes is equal to the carrying capacity in the absence of the fungus.

	Value	**Interpretation**	**Unit**	**Source**
H	3000	Total human population	humans	-
$\tilde{P}$	10000	Mosquito carrying capacity	mosquitoes	[18, 101]
γ	0.5	proportion of infectious bites from an infected mosquito	-	[18, 19]
β	0.15	per capita mosquito biting rate	days^{-1}	[18]
ρ	0.005	recovery rate of infected humans	days^{-1}	[18, 64]

	Value	Interpretation	Unit	Source
μ	0.1	Natural mosquito death rate	days^{-1}	[64, 101]
κ	0.48	Innate mosquito growth rate	days^{-1}	[2]
α	varied	Exposure rate of the fungus	days^{-1}	-
σ	varied	Fungus-induced death rate	days^{-1}	-
ξ	varied	Vertical transmissibility of fungus	days^{-1}	-
κ_c	0.48	Mosquito larvae birth rate	days^{-1}	see text
κ_L	3.8e-5	Larval competition parameter	days^{-1} larvae^{-1}	see text

Table C.1: Summary of estimates and references for the parameters used in the malaria models.