An investigation into joint HIV and TB epidemics in South Africa

Carel Diederik Pretorius

Dissertation presented for the degree of
Doctor of Philosophy

Department of Physics
Faculty of Science

Supervisor: Prof. Kristian Müller-Nedebock
Co-supervisor: Dr. Alex Welte

December 2009

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 24 August 2009

Abstract

This dissertation investigates certain key aspects of mathematical modeling of HIV and TB epidemics in South Africa with particular emphasis on data from a single well-studied community. Data collected over a period of 15 years (1994 to 2009) in Masiphumelele, a township near Cape Town, South Africa are used to develop a community-level mathematical model of the local HIV-TB epidemic. The population is divided into six compartments and a system of differential equations is derived to describe the spread of the dual epidemic. Our numerical results suggest that increased access to antiretroviral therapy (ART) could decrease not only the HIV prevalence, but also the TB notification rate. We present a modeling framework for studying the statistical properties of fluctuations in models of any population of a similar size. Viewing the epidemic as a jump process, the method entails an expansion of a master equation in a small parameter; in this case in inverse powers of the square root of the population size. We derive two-time correlation functions to study the correlation between different types of active TB events, and show how a temporal element could be added to the definition of TB clusters, which are currently defined solely by DNA type. We add age structure to the HIV-TB model in order to investigate the demographical impact of HIV-TB epidemics. Our analysis suggests that, contrary to general belief, HIV-positive cases are not making a substantial contribution to the spread of TB in Masiphumelele. We develop an age-structured model of the HIV-TB epidemic at a national level in order to study the potential impact of a proposed universal test and treat program for HIV on dual HIV-TB epidemics. Our simulations show that generalized ART could significantly reduce the TB notification rate and the TB-related mortality rate in the short term. The timescale of the impact of ART on HIV prevalence is likely to be longer. We study the potential impact of more conventional control measures against HIV. Guidance for possible future and/or additional interventions emerge naturally from the results. We advocate a reduction in intergenerational sex, based on our finding that 1.5-2.5 standard deviation in the age difference between sexual partners is necessary to create and sustain a major HIV epidemic. A simulation framework is developed to help quantify variance in age-structured epidemic models. The expansion technique is generalized to derive a Fokker-Planck equation. Directions for future work, particularly in terms of developing methods to model fluctuations and validate mixing assumptions in epidemiological models, are identified.

Opsomming

Hierdie proefskrif ondersoek aspekte van die wiskundige modelering van HIV en TB epidemies in Suid Afrika en fokus ook op 'n spesifieke gemeenskap. Data wat oor 'n periode van 15 jaar ingesamel is (1994 tot 2009) in Masiphumelele, 'n woonbuurt naby Kaapstad, Suid Afrika word gebruik om 'n wiskundige model te skep wat HIV-TB in die gemeenskap modeleer. Die populasie word in ses kompartemente verdeel en 'n stel differensiaal vergelykings word afgelei om die verspreiding van dié epidemies te ondersoek. Ons numeriese resultate toon aan dat verhoogde toegang tot antiretrovirale behandeling (ARB) die potensiaal het om HIV prevalensie die TB koers beduidend te laat daal. Ons ontwikkel 'n raamwerk waarmee die statistiese eienskappe van fluktuasies ondersoek kan word in enige populasie van dieselfde grootte. Die metode ontwikkel 'n meester vergelyking vir die onderliggende geboorte-dood stogastiese proses en brei dit uit in terme van 'n klein parameter; in dié geval in inverse magte van die vierkantswortel van die populasie grootte. Die twee-tyd korrelasie funksies word afgelei, en word gebruik om die korrelasie tussen verskillende tipes van TB episodes te bestudeer, asook om te wys hoe 'n tydselement aan die definisie van TB groeperings gegee kan word. Dié word tans slegs d.m.v DNA tipe geklassifiseer. Ouderdomstruktuur word aan die model toegevoeg om die demografiese impak van HIV-TB epidemies te bestudeer. Ons analise toon aan dat, anders as wat algemeen aanvaar word, maak HIV-positiewe gevalle nie 'n groot bydrae tot die verspreiding van TB in Masiphumelele nie. Ons ontwikkel 'n ouderdom-gestruktureerde model van HIV-TB op nasionale vlak en gebruik die model om die potensiële impak van 'n universele toets- en behandel strategie op die HIV-TB epidemies te ondersoek. Ons simulasies toon aan dat algemene ARB waarskynlik 'n groot impak op die TB aanmeldings koers asook die TB-verwante mortaliteits koers kan hê binne 'n relatiewe kort tydperk. Die impak op HIV prevalensie sal eers oor 'n veel langer periode duidelik word. Ons ondersoek ook die moontlikheid van meer konvensionele beheermaatreels. Ons ontmoedig tussengenerasie seksuale omgang, gegrond op ons bevinding dat 'n standaard afwyking van 1.5-2.5 in die ouderdoms verskil tussen seksuele vennote, nodig is om 'n HIV epidemie van stapel te stuur en te onderhou. Ons ontwikkel 'n simulasie raamwerk om variansie in ouderdomgestruktureerde modelle te benader. Die uitbreidingstegniek word veralgemeen om 'n Fokker-Planck vergelyking af te lei. Ons identifiseer probleme in die ontwikkeling van metodes om interaksie patrone en fluktuasies te modeleer in epidemiologiese modelle as opgawe vir toekomstige werk.

Acknowledgements

I am indebted to many people for contributing to this dissertation. Firstly, I would like to thank my supervisor, Prof. Kristian Müller-Nedebock, for offering mentorship, encouragement and expertise in the application of ideas from statistical physics to the field of HIV-TB epidemiology. His editing and restructuring of the dissertation was especially valuable. I also want to thank Alex Welte, who co-supervised this work. His advice and mentorship over the last few years have been valuable to me.

I met Nicolas Bacaër from the Institut de Recherche pour le Développement (IRD) in Paris three years ago, at the time when he was making the early strides in our modeling of dual HIV-TB epidemics. His modeling experience and generosity in sharing his ideas had a profound influence on my work. I would also like to thank him for his hospitality when I visited the IRD in 2008.

I want to thank Brian Williams, who is a great source of inspiration to me. Throughout this project he made many valuable contributions to the modeling challenges I faced.

SACEMA (The South African Centre for Epidemiological Modelling and Analysis), in Stellenbosch, provided not only an intellectually stimulating environment but a group of great friends and collaborators. I climbed many mountains around Stellenbosch with my director Prof. John Hargrove and value his support and advice immensely. My good friend and collaborator Rachid Ouifki, a researcher at SACEMA, is a constant source of encouragement to me. An ongoing collaboration with my friend Wim Delva has benefitted my understanding of the public health perspective of sexually transmitted disease dynamics. My other friends and colleagues at or linked to SACEMA: Jeremy Lauer, Jonathan Dushoff, Ekkehard Kopp (the list goes on) have all been generous in their friendship, hospitality and sharing of ideas. Jeremy Lauer and Wim Delva have been especially helpful in proofreading parts of the manuscript, and made valuable suggestions to improve its presentation. Discussions with Stephane Verguent and Séverin-Guy Mahiane, who visited SACEMA at the beginning of 2009, benefited some of my modeling work.

I want to thank Robin Wood, who heads the Desmond Tutu HIV Centre at UCT, for sharing both data collected in this community and his vast experience in the clinical aspects of HIV-TB epidemiology.

I cannot thank my family and my girlfriend Monica Guy enough for their unwavering support and inspiration.

Contents

List of Figures

List of Tables

Chapter 1

Introduction to HIV and TB epidemics

Communicable diseases provide fascinating material for mathematical modeling, and in return, modeling can offer important insights into disease control. Seen as stochastic processes, these diseases exhibit a great deal of randomness in the timing and manner of their invasion, persistence and ultimate disappearance from victim populations. They can attain alarmingly high levels of infection relatively quickly (e.g. HIV in Sub-Saharan Africa), or remain limited to periodic outbreaks from fairly low small endemic levels (e.g. chickenpox or mumps). Their normal trajectory runs from individuals into the wider population. They diffuse through social networks. Sometimes they interact with other diseases along the way. There is an urgent societal need for a better understanding of the macro characteristics of disease progression. Mathematical epidemiology has thus evolved into an active field of modeling research, which makes valuable contributions to disease control and public health decision making.

Few epidemiological scenarios are as gloomy as the one presented by HIV and *Mycobacterium tuberculosis* (MTB) in certain regions of the world [17]. A few simple facts suffice to paint the picture: in 2005 nearly 40 million people were living with HIV and 3 million died of AIDS [8]. Nearly 2 billion were latently infected with MTB [30]. There were 8.8 million new cases of active tuberculosis (TB). A total of 1.6 million people died of TB in 2005, including 195,000 people who were HIV_+ [116]. The numbers increase every year. According to a 2009 report [11] by the World Health Organization (WHO) there were 9.24 million new cases of TB (all forms) in 2006, and 9.27 million in 2007, of which 44% were smear positive.

One of the hardest hit regions in the world is Southern Africa, and South Africa in particular, where HIV and TB are a leading cause of death. In 2005, 5.5 million South Africans were living with HIV (12% of the country's population) and 285,000 developed active TB. Among these TB cases, up to 60% were HIV_+, mostly due to the fact that HIV/MTB co-infected people exhibit an increased probability of developing active TB [96,

97]. The incidence of TB in South Africa reached 600 per 100,000 per year – one of the highest in the world. The above mentioned report [11, SA profile] estimates that the TB burden continues to be exceptionally high in South Africa. For example, the Western Cape region registered almost 700 per 100,000 cases of active TB in 2007.

What is going wrong? If current disease control strategies are failing to meet their targets (which they are) there must be a problem either with the scientific basis of these strategies or with their application – or both. One example is the Directly Observed Treatment, Short-course (DOTS) strategy for TB, recommended by the WHO. Despite widespread implementation of DOTS, TB notifications have exploded in recent years in many communities across South Africa. There are ongoing debates as to whether DOTS is being applied correctly or whether it is in any case doomed to failure in a setting of very high HIV prevalence. The situation is similar for HIV. Clinicians are now recognizing the value of modeling in helping to identify more effective disease control strategies.

This dissertation aims to explore and unify the techniques of mathematical modeling as applied to HIV and TB. Working in collaboration with clinicians [1] in the Western Cape region, and with access to their insights and data, the project explores some of the boundaries of applied mathematical modeling of these diseases in Masiphumelele, a township near Cape Town with high HIV and TB prevalence, as well as in South Africa as whole.

1.1 HIV and TB epidemics in a peri-urban community, Masiphumele

Masiphumelele (meaning 'We will succeed') is an informal settlement area in the Western Cape. Steady immigration from the Eastern Cape since its proclamation in 1992 has led to severe overcrowding, because the town cannot physically grow in any direction. A population of 13,000 is now living in an area of about 1 km^2 which is geographically isolated from other communities. Housing is informal and most people work in the informal sector of the greater Cape Town region.

Research programs of the Desmond Tutu HIV Foundation at Masiphumelele continue to generate data on the epidemiology of HIV and TB [69, 123]. Using their data, we can get a picture of how HIV has 'driven' TB over the last two decades. TB notifications continue to escalate in Masiphumelele despite implementation of DOTS. This occurs in the presence of a growing HIV epidemic. The situation does not bode well for intervention in similar townships and South Africa as a whole, because the capacity to intervene at Masiphumelele is much greater than nationally. Treatment programs here are reaching a much greater proportion of those in need than the national average, and yet they are still not working satisfactorily. In Ch. 2 we develop a simple compartmental model to study dual HIV-TB epidemics in this community. Despite, or perhaps because of its simplicity, the model can

[1] Prof. Robin Wood leads a group of clinicians working in a community clinic built in 2000 by the Desmond Tutu HIV Foundation

help us to understand the impact of various control measures in dealing with the extremely high rates of both HIV and TB.

Referring to the problem experienced by the DOTS strategy in controlling TB, as mentioned in Sect. 1, a recent article has reported a very high prevalence of infection with MTB among children aged 5 to 17 in Masiphumele [80]. The TB notification rate in the adult population has increased by a factor of over 5 the last 15 years, yet the annual risk of infection (ARI) was reported to be roughly 4% over this period. In Ch. 4 we develop a simple stochastic model to study whether a sharp increase in notified active TB cases has led to an increase in the annual risk of MTB infection among schoolchildren.

Masiphumelele's complex demographics also play a role in the spread of HIV and TB in its population – providing yet another interesting twist to our model. The population has grown considerably over the past decade. The age pyramid is skewed inasmuch as there are relatively more young adults than children and older people, which may be the result of immigration. Clinical data show that TB and HIV have become progressively more concentrated among young adults during the last few years. We build a continuous-time age-structured model of HIV and TB dynamics to study the shift of the HIV and TB burden to younger age groups.

1.2 Modeling HIV and TB epidemics in South Africa

Epidemiological modeling is not just interesting mathematically; it plays a crucial role in guiding interventions which are being discussed and implemented at the time of writing. Recent modeling work [47] suggested that a very active program of HIV testing, with all detected HIV_+ people immediately receiving ART, could be an efficient way of not only controlling but even eradicating the HIV epidemic. With the exception of male circumcision (an HIV control strategy which is currently taking off in South Africa following the work of Auvert et al. [14]) HIV control measures have thus far failed. HIV prevalence continues to grow, with high-risk groups in many communities facing unprecedented incidence of infection. We develop an age-structured model to investigate the potential impact of a universal test and treat strategy (UTTS) on HIV-TB dual epidemics in South Africa.

The key motivation for using ART as a prevention tool against HIV is that it reduces viral load and infectiousness. ART could reduce HIV at a community level and possibly eradicate it, if each infected case treated with ART caused fewer secondary infective cases. We model different assumptions about the degree to which ART reduces infectiousness. In Ch. 5. we study an optimistic scenario, where HIV_+ cases on ART are not infectious, to investigate the impact of UTTS on the prevalence of HIV.

Our model also explores the likely impact of UTTS on TB, which is strongly linked to HIV progression. The degree of this impact will depend on how much ART will reconstitute the immune system of an HIV_+ individual. In Ch. 5. we make the optimistic assumption that ART will revert the 'TB parameters' of an HIV_+ individual to that of someone HIV_-,

in order to investigate if the TB epidemic will revert back to its pre-HIV level.

Along with UTTS, we also study the potential impact of other control measures. The role of intergenerational sex in creating and maintaining a persistent HIV epidemic is still not completely understood. In Ch. 6, we develop a model with random mixing and explicit relationship dynamics, which allows us to investigate some of the subtle consequences of variance in the age difference between sexual partners.

1.3 Outline of this work

Chapter 2 introduces a model for TB and HIV epidemics. Similar models until now have focused more on HIV than TB; our model has a better balance of complexity and fit to available data. The model is used as a platform for combining various sources of data on the HIV and TB epidemics in Masiphumelele. Sensitivity analysis of the model is performed with respect to key parameters. It shows that HIV_+ individuals have a short infectious period relative to HIV_- cases. The next-generation matrix is derived and the basic reproductive number (R_0) is computed for the two epidemics. The model is used to study the impact of different interventions against HIV and TB, of which increased condom use, TB detection rates and isoniazid preventative therapy have been shown to have a clear impact. The model shows that the impact of increased ART on the TB burden in this community is uncertain. Before we can tout the expected benefits of UTTS we need to achieve a greater understanding of how much it reduces TB notification rates and HIV-related mortality, TB reactivation and HIV transmission.

Chapter 3 adapts a technique from statistical physics to model the steady-state fluctuations in a model with TB only. Continuous differential equations are often applied to small populations such as Masiphumelele, with little time spent on understanding the uncertainty associated with deterministic models, brought about by small-population effects. The Fokker-Planck equation for the fluctuations in the system is used to quantify this uncertainty. It is also used to develop a method to characterize the temporal aspects of the 'clustering' of active TB events.

Chapter 4 investigates the curious finding of a recent tuberculin skin test (TST) study among schoolchildren in Masiphumelele, that the annual risk of MTB infection (ARI) has remained relatively constant over the last 15 years. This is despite the dramatic increase in notified TB cases, in part due to an escalating HIV epidemic. We use a binomial-chain type model to compute the likelihood of the TST data-set. Using age-structured TB and HIV data from Masiphumelele, and an age-structured model, we investigate whether trends in the TB epidemic are linked to similar trends in the HIV epidemic. The reproductive value is used to establish exactly who is making the biggest contribution to the TB epidemic.

Chapter 5 builds on the models of Chapters **2** and **4**, and develops an age-structured model for HIV and TB epidemics in South Africa as a whole. The motivation for the model is to inform a proposed universal test and treat strategy, aimed at not only treating individuals already suffering from HIV, but also at using ART as a prevention tool by reducing the infectiousness of individuals and therefore the spread of HIV. Central to these impact studies is the next-generation matrix and its largest eigenvalue (R_0), which is widely used to estimate the impact of interventions on disease control. We generalize this method to study age-structured populations and the infectious histories of individuals through realistic life events.

Chapter 6 investigates the idea that variance in age separation between sexual partners is needed for HIV to spread. If everybody always has partners of exactly the same age as themselves, the epidemic dies out [118, p.562]. Using a hypothetical community, based on but not a true reflection of the HIV epidemic in Masiphumelele, a two-sex model is used to investigate this critical aspect of the spread of HIV. The model is structured according to age and gender, and allows individuals to transition between being young, 'eligible for relationships' and 'entering relationships' through an age-dependent partner choice. Using a semi-Markov process, we show how R_0 for HIV, i.e. its invasion criterion, varies as a function of age separation between partners. A method is also developed to approximate the effect of concurrent casual relationships, focussing on the impact it has on the critical role played by age separation between partners.

Chapter 7 studies individual-based models using both analytical and stochastic simulation techniques. It shows how a linear transition matrix of Markov processes can be used to model a 'sum over all histories' of possible individual transitions. The analogous Gillespie stochastic simulation technique allows us to sum over all possibilities while simulating the evolution of an interacting system. A platform is designed to study partnering models and to record the relationship network formed. It is also used to estimate some properties of the model developed in Ch. 6, such as the number of partners infected during a lifetime. The 'master equation expansion' method is used to study an age-structured model with two genders and mass-action interaction between them. A Fokker-Planck equation is then derived to study fluctuations in the system and is used to estimate the variance associated with macroscopic values at equilibrium.

Appendices A list of abbreviations used in this dissertation in tabled in appendix A. The next-generation matrix approach to modeling structured population is discussed in appendix B.1 and appendix B.2. Many of the principal ideas of applied demography are based on the so-called characteristic equation. This equation is studied in appendix B.3.

1.4 Publications

This dissertation was built around the following papers and presentations at conferences:

Chapter 2. *Modeling the joint epidemics of TB and HIV in a South African Township.* Bacaër N., Ouifki R., Pretorius C.D., Wood R., Williams B. Journal of Mathematical Biology. 2008 Oct;57(4):557-93.

The results of this paper were presented by C.D. Pretorius at the 38th Union World Conference on Lung Health 8-12 November 2007, Cape Town, South Africa in a talk titled: *Modelling the joint epidemics of TB and HIV in a peri-urban community in South Africa: What are the prospects for control?*

Chapter 4. *On the relationship between age, annual rate of infection, and prevalence of Mycobacterium Tuberculosis in a South African Township.* Pretorius C.D., Bacaër. N., Williams B., Wood R., Ouifki R. Clinical Infectious Diseases. 2009;48(7):994-6.

This paper was completed while C.D. Pretorius was visiting the Institut de Recherche pour le Developpement (IRD) in Bondy, France. The method used in this paper for computing the annual risk of MTB infection was presented by C.D. Pretorius at an invited talk at WHO (Geneva), 25 September 2008.

Chapter 5. *Modeling the potential impact on HIV and tuberculosis of a generalized access to antiretrovirals in South Africa.* Pretorius C.D., Bacaër. N., Submitted to Bulletin of Mathematical Biology.

This paper was presented by C.D. Pretorius at a conference titled: *Can we treat our way out of the HIV epidemic.* 6-8 May 2009, Stellenbosch, South Africa. The conference was jointly hosted by VLIR, UGhent, SACEMA, UWC and AIMS.

Chapter 6. The model developed in this chapter was presented by C.D. Pretorius at the 4th South African AIDS conference, 8-12 April 2009, Durban, South Africa, titled: *Intergenerational sex and the epidemic of HIV.*

Other publications. The following papers are not directly related to this dissertation. The first two were influential in developing ideas for modeling relevant heterogeneity in epidemic models. They resulted from a ground-breaking male circumcision clinical trial in Orange Farm, near Johannesburg, South Africa, performed by B. Auvert, et al. [14].

The third is an opinion piece written for the South African Medical Journal, following the 4th South African AIDS conference, 2009. In this paper the authors question whether scaling up the current response to the HIV epidemic will eventually limit the spread of and eradicate HIV in Southern Africa. The argument for an early and universal test and treat strategy for HIV is discussed.

- *The effect of heterogeneity on HIV prevention trials.* Auvert B., Sitta R., Zarca K., Mahiane G., Pretorius C., Lissouba P. Under preparation for: Clinical Trials.
- *Mathematical Models for the co-infection by two Sexually Transmitted Agents: the HIV/HSV-2 case.* Mahiane S.G, Ndong-Nguema E.P, Auvert B., Pretorius C. Under preparation for: Journal of the Royal Statistical Society, Series C.
- *Is scaling up enough to curb the HIV epidemic in southern Africa?* Delva W., Pretorius C., Temmerman M. SAMJ. In press.

Chapter 2

Modeling joint HIV and TB epidemics in a South African township

We present a simple mathematical model with six compartments for the interaction between HIV and TB epidemics. Using data from a township near Cape Town, South Africa, where the prevalence of HIV is above 20% and where the TB notification rate is close to 2,000 per 100,000 per year, we estimate some of the model parameters and study how various control measures might change the course of these epidemics. Condom promotion, increased TB detection and TB preventative therapy have a clear positive effect, but there are some difficulties in predicting the effect of ART at the population level. ART reduces the risk of co-infected individuals on ART by up to 80%, but their life expectancy and infectious period is also greatly increased. As a result, ART may increase TB transmission.

Detailed studies of these epidemics in a township near Cape Town have been published recently [18, 68, 69, 71, 73, 74, 75, 123]. Estimates of the TB notification rate (based on the yearly number of TB notifications, on two population censes conducted in 1996 and in 2004, and assuming a linear population increase in between) and of the prevalence of HIV (estimated using data from an antenatal clinic) are shown in Tab. 2.1.

Table 2.1: TB notifications per 100,000 per year and HIV prevalence (%). Data from [69, Tab. 1].

Year	1996	1997	1998	1999	2000	2001	2002	2003	2004
TB	580	653	913	897	982	1,410	1,366	1,472	1,468
HIV	6.3	8.9	11.6	14.2	16.5	18.4	19.9	21.1	21.9

For the year 2005, 259 TB cases were reported among adults (aged$\geq$ 15) [123]; 66% of those who were tested for HIV were HIV$_+$. The adult population was then estimated to be 10,400 and the total population 13,000. The TB notification rate in the whole population was

therefore over $259/13,000 \simeq 1,992$ per 100,000 per year. Moreover, in a sample population of 762 adults, 12 had undiagnosed TB (3 HIV_- and 9 HIV_+). Around 23% (174/762) of the sample population was HIV_+. More than 80% of smear-positive TB cases receiving treatment were cured.

In Bacaër, et al. (Tab. 2[17]) we present an extensive survey of models developed for HIV-TB dynamics. These were not very effective in understanding HIV-TB dynamics in real communities. The models have been of essentially two different types: either computer simulation studies focusing on transient behavior (usually in response to intervention) of realistic but complex models, or mathematical studies of simpler but less realistic models focusing on steady states and their stability. All of these models not only contain many unknown parameters but also rely on little data. The model we develop strikes a balance between complexity and the amount of data available [17]. This chapter highlights some of its key findings.

2.1 A model for HIV-TB epidemics

The compartmental structure of our model combines two states for HIV (HIV_- and HIV_+) with three states for TB (susceptible, latent TB and active TB as in [84, 83, 100]). The notations for the resulting six compartments are shown in Tab. 2.2. The subscript 1 always refers to HIV_- individuals and the subscript 2 to HIV_+ individuals. Compartments E_1, E_2, I_1 and I_2 represent those infected with MTB.

Table 2.2: The six compartments of the model and some notations.

S_1	number of HIV_- individuals who are not infected with MTB
S_2	number of HIV_+ individuals who are not infected with MTB
E_1	number of HIV_- individuals with latent TB
E_2	number of HIV_+ individuals with latent TB
I_1	number of HIV_- individuals with active TB
I_2	number of HIV_+ individuals with active TB
P	total population: $P = S_1 + E_1 + I_1 + S_2 + E_2 + I_2$
H	HIV prevalence: $H = (S_2 + E_2 + I_2)/P$

The parameters of the model are shown in Tab. 2.3. The "physiological" parameters are more or less the same for people throughout the world or at least for people living in sub-Saharan Africa: the death rates μ_1 and μ_2, the TB parameters p_1, p_2, q_1, q_2, a_1, a_2, m_1 and m_2. On the contrary, the "social" parameters depend on the area under study, in particular on population density and living conditions (the transmission rates k_1 and k_2), access to TB clinics (the detection rates γ_1 and γ_2), quality of treatment (ε_1 and ε_2), sexual habits and local cofactors for the transmission of HIV such as other sexually transmitted diseases and male circumcision (d), speed at which information on HIV diffuses (λ) or epidemic history (t_0). Estimates for most physiological parameters can be found in the medical literature.

All "social" parameters have to be estimated from local data.

Table 2.3: The 22 parameters of the model and some extra notations (subscript 1 for HIV_- individuals, subscript 2 for HIV_+ individuals).

B	birth rate
μ_1, μ_2	death rate of individuals who do not have active TB
k_1, k_2	maximum transmission rate of MTB
p_1, p_2	proportion of new infections with fast progression to TB
q_1, q_2	proportion of reinfections with fast progression to TB
a_1, a_2	progression rate from latent TB to active TB
β_1, β_2	recovery rate from active TB without treatment
γ_1, γ_2	detection rate of active TB cases
ε_1, ε_2	probability of successful treatment for detected active TB cases
m_1, m_2	death rate for active TB cases
d	maximum transmission rate of HIV
λ	parameter representing behavior change
t_0	time of introduction of HIV
p'_1, p'_2	proportion with slow progression to TB: $p'_1 = 1 - p_1$, $p'_2 = 1 - p_2$
b_1, b_2	recovery rate from TB: $b_1 = \beta_1 + \gamma_1\,\varepsilon_1$, $b_2 = \beta_2 + \gamma_2\,\varepsilon_2$
$f(H)$	reduced transmission rate of HIV: $f(H) = d\,e^{-\lambda H}$

The equations of our model are:

$$\frac{dS_1}{dt} = B - S_1\,(k_1\,I_1 + k_2\,I_2)/P - \mu_1\,S_1 - f(H)\,H\,S_1\,, \tag{2.1}$$

$$\frac{dE_1}{dt} = (p'_1\,S_1 - q_1\,E_1)(k_1\,I_1 + k_2\,I_2)/P - (a_1 + \mu_1)\,E_1 + b_1\,I_1 - f(H)\,H\,E_1\,, \tag{2.2}$$

$$\frac{dI_1}{dt} = (p_1\,S_1 + q_1\,E_1)(k_1\,I_1 + k_2\,I_2)/P - (b_1 + m_1)\,I_1 + a_1\,E_1 - f(H)\,H\,I_1\,, \tag{2.3}$$

for HIV_- individuals and

$$\frac{dS_2}{dt} = -S_2\,(k_1\,I_1 + k_2\,I_2)/P - \mu_2\,S_2 + f(H)\,H\,S_1\,, \tag{2.4}$$

$$\frac{dE_2}{dt} = (p'_2\,S_2 - q_2\,E_2)(k_1\,I_1 + k_2\,I_2)/P - (a_2 + \mu_2)\,E_2 + b_2\,I_2 + f(H)\,H\,E_1\,, \tag{2.5}$$

$$\frac{dI_2}{dt} = (p_2\,S_2 + q_2\,E_2)(k_1\,I_1 + k_2\,I_2)/P - (b_2 + m_2)I_2 + a_2\,E_2 + f(H)\,H\,I_1\,, \tag{2.6}$$

for HIV_+ individuals. The flows between the different compartments are shown in Fig. 2.1.

Tab. 2.4 shows the correspondence we will use between some medical vocabulary and our model. The TB notification rate is the rate at which individuals in compartments I_1 and I_2 are detected (only a fraction ε_1 or ε_2 of these really move back to the latent compartments E_1 and E_2). The TB incidence rate is the rate at which individuals enter the compartments I_1 and I_2 divided by the total population usually given "per 100,000 population per year". The MTB infection rate (the continuous-time analogue of the annual risk of infection) is the

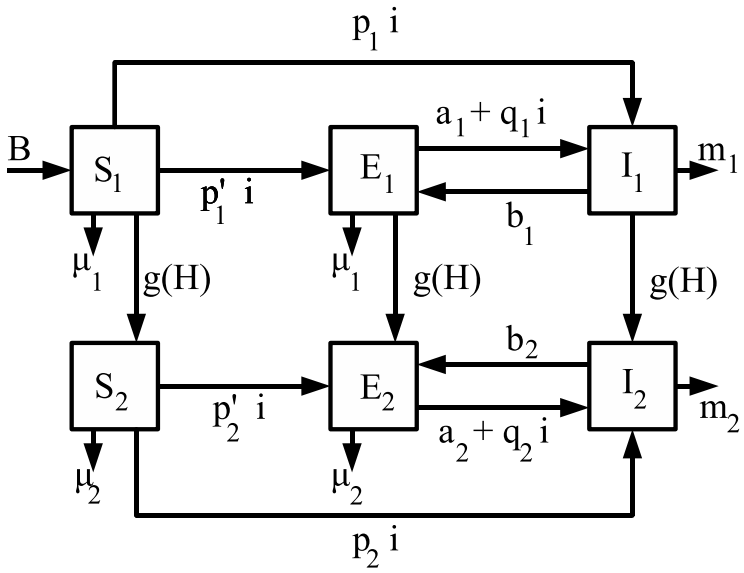


Figure 2.1: Flows between the compartments of the model. Here, $i = (k_1 I_1 + k_2 I_2)/P$ and $g(H) = f(H)\, H$.

rate at which individuals in compartments S_1 (resp. S_2) move to compartments E_1 or I_1 (resp. E_2 or I_2). MTB prevalence is the proportion of the total population in compartments E_1, I_1, E_2 or I_2. TB prevalence is the proportion of the total population in compartments I_1 or I_2. It includes active TB cases, i.e., either undiagnosed TB cases or TB cases that have been detected but that are unsuccessfully treated.

Table 2.4: Correspondence between some medical vocabulary and the model.

TB notification rate	$(\gamma_1 I_1 + \gamma_2 I_2)/P$
MTB infection rate	$(k_1 I_1 + k_2 I_2)/P$
"total" TB incidence rate	$T = a_1 E_1 + a_2 E_2 +$ $(p_1 S_1 + p_2 S_2 + q_1 E_1 + q_2 E_2)(k_1 I_1 + k_2 I_2)$
TB incidence rate	T/P
MTB prevalence	$(E_1 + I_1 + E_2 + I_2)/P$
TB prevalence	$(I_1 + I_2)/P$
% endogenous reactivation	$(a_1 E_1 + a_2 E_2)/T$
% exogenous reinfection	$(q_1 E_1 + q_2 E_2)(k_1 I_1 + k_2 I_2)/T$
% primary disease	$(p_1 S_1 + p_2 S_2)\,(k_1 I_1 + k_2 I_2)/T$

A number of key points should be borne in mind:

- At time t_0, we assume that one HIV$_+$ person is introduced in an HIV-free steady population where TB is endemic. We chose this first HIV case to be in state S_2. The formulae for S_1, E_1 and I_1 at the endemic TB steady state will be given in Sect. 2.2.

- Age and sex are not taken into account. In particular, the model cannot distinguish different routes of transmission of HIV, such as sexual transmission and mother-to-

child transmission. We did not distinguish pulmonary from extra-pulmonary TB, smear-positive (infectious) TB from smear-negative (non-infectious) TB in order to reduce the number of compartments to a minimum.

- Drug-resistant TB is still very limited in the South African township under study. The efficiency of BCG vaccination is also unclear. We have not included these aspects in our model.

- In Eq. (2.1), the birth rate is assumed to be a constant independent of the number of individuals who die of HIV and/or TB. Therefore, our model considers the evolution of cohorts with a fixed size at birth. If on the other hand we assumed that births are proportional to the population, then a steady-state analysis would become impossible. The demography of the township is in fact quite complex. The population has grown considerably over the past decade. The age pyramid is skewed with more young adults and few children and elderly people (see Sect. 4.2.2).

- In Eqs. (2.1) and (2.4), we chose the "standard form" for TB infection and reinfection as in [40, 98, 100], and not the "mass action" form used e.g. in [46, 84, 83]. With a constant birth rate, the total population decreases as the HIV epidemic develops. If we used the "mass action" form for TB transmission, the transmission rate would also decrease and this would artificially slow down the TB epidemic.

- In Eqs. (2.1)-(2.3), we also chose the "standard form" for the transmission of HIV as e.g. in [98].

- We note how the equations model individuals that are unsuccessfully treated for TB. They are counted in the TB notification rate $\gamma_1 I_1 + \gamma_2 I_2$, and induce lower recovery rates $b_1 = \beta_1 + \gamma_1 \varepsilon_1$ and $b_2 = \beta_2 + \gamma_2 \varepsilon_2$ among active TB cases. However, they are not counted in a separate compartment.

2.2 Mathematical analysis

We discuss some of the highlights of the mathematical analysis of the HIV-TB model given Eqs. (2.1)-(2.6). A more detailed discussion is available in Sect. 4 [17]. Notable is the derivation and analysis of a quadratic equation for the TB-only steady state, which shows that a "transcritical bifurcation" in the TB-only steady state is possible only when $q_1 > p_1$. However, realistic values for q_1 are always less than those of p_1, because a particular episode of TB infection offers a degree of protection against future episodes (see [17, Sect. 5.3] for a detailed discussion). This finding suggests that the parameter region with a backward bifurcation is a mathematical curiosity that does not occur in practice, confirming the remarks in [76] and the conclusion suggested by [100].

The disease-free steady state with no TB and no HIV is given by $S_1^0 = B/\mu_1$ and $E_1 = I_1 = S_2 = E_2 = I_2 = 0$. Numerically, $S_1^0 = 10,000$.

TB only

Background. The model with TB but no HIV consists only of three compartments (S_1, E_1, I_1) satisfying Eqs. (2.1)-(2.3) with $I_2 = 0$, $H = 0$, and $P = S_1 + E_1 + I_1$:

$$\frac{dS_1}{dt} = B - k_1 S_1 I_1/P - \mu_1 S_1 \,, \tag{2.7}$$

$$\frac{dE_1}{dt} = (p_1' S_1 - q_1 E_1)\, k_1 I_1/P - (a_1 + \mu_1)\, E_1 + b_1 I_1 \,, \tag{2.8}$$

$$\frac{dI_1}{dt} = (p_1 S_1 + q_1 E_1)\, k_1 I_1/P - (b_1 + m_1)\, I_1 + a_1 E_1 \,. \tag{2.9}$$

Analysis. Linearizing system given by Eqs. (2.7)-(2.9) near the disease-free steady state when $S_2 = E_2 = I_2 = 0$, we obtain:

$$\frac{dE_1}{dt} \simeq k_1 p_1' I_1 - (a_1 + \mu_1)\, E_1 + b_1 I_1 \,,$$

$$\frac{dI_1}{dt} \simeq k_1 p_1 I_1 - (b_1 + m_1)\, I_1 + a_1 E_1 \,.$$

It follows that the basic reproduction number R_0^{TB} for TB, as defined in [36], is the spectral radius of the matrix

$$\begin{pmatrix} 0 & k_1 p_1' \\ 0 & k_1 p_1 \end{pmatrix} \begin{pmatrix} a_1 + \mu_1 & -b_1 \\ -a_1 & b_1 + m_1 \end{pmatrix}^{-1} ,$$

which does not depend on the reinfection parameter q_1 and can easily be computed:

$$R_0^{\mathrm{TB}} = \frac{k_1 (a_1 + p_1\, \mu_1)}{a_1 m_1 + m_1 \mu_1 + \mu_1 b_1} \,. \tag{2.10}$$

Because this formula does not depend on the reinfection parameter q_1, it is the same as [86, Eq. (10)]. When $b_1 = 0$ and $p_1 = 0$, it is the same as the formula given in [40, §1].

HIV only

When there is no TB, system given by Eqs. (2.1)-(2.6) reduces to

$$\frac{dS_1}{dt} = B - \mu_1\, S_1 - f(H)\, H\, S_1 \,, \quad \frac{dS_2}{dt} = -\mu_2\, S_2 + f(H)\, H\, S_1 \tag{2.11}$$

with $H = S_2/(S_1 + S_2)$. Similar epidemic models with a contact rate depending nonlinearly on the number of infected individuals have been studied for example in [53, 112]. A more complicated model for HIV transmission with a contact rate depending nonlinearly on the prevalence was used in [122]. First, linearize the second equation in Eq. (2.11) near the disease-free steady state $S_1 = S_1^0$ and $S_2 = 0$:

$$\frac{dS_2}{dt} \simeq -\mu_2\, S_2 + f(0)\, S_2 \,.$$

Hence, the basic reproduction number for HIV is given by

$$R_0^{\rm HIV} = f(0)/\mu_2\,.$$

HIV and TB

The endemic TB steady state can be invaded by HIV. Linearizing system given by Eqs. (2.4)-(2.6) near this steady state and setting

$$P^* = S_1^* + E_1^* + I_1^*\,, \quad s_1^* = S_1^*/P^*\,, \quad e_1^* = E_1^*/P^*\,, \quad i_1^* = I_1^*/P^*\,, \tag{2.12}$$

we obtain

$$\begin{aligned}
\frac{dS_2}{dt} &\simeq -k_1 S_2\, i_1^* - \mu_2\, S_2 \, + f(0)\, s_1^*\,(S_2 + E_2 + I_2)\,,\\
\frac{dE_2}{dt} &\simeq k_1\,(p_2'\, S_2 - q_2\, E_2)\, i_1^* - (a_2 + \mu_2)\, E_2 + b_2\, I_2 + f(0)\, e_1^*\,(S_2 + E_2 + I_2)\,,\\
\frac{dI_2}{dt} &\simeq k_1\,(p_2\, S_2 + q_2\, E_2)\, i_1^* - (b_2 + m_2)\, I_2 + a_2\, E_2 + f(0)\, i_1^*\,(S_2 + E_2 + I_2)\,.
\end{aligned}$$

Therefore, the basic reproduction number $r_0^{\rm HIV}$ for HIV when introduced in a population at the TB endemic steady state (note that $r_0^{\rm HIV}$ is different from $R_0^{\rm HIV}$) is the spectral radius of the matrix:

$$f(0)\begin{pmatrix} s_1^* & s_1^* & s_1^* \\ e_1^* & e_1^* & e_1^* \\ i_1^* & i_1^* & i_1^* \end{pmatrix}\begin{pmatrix} k_1 i_1^* + \mu_2 & 0 & 0 \\ -k_1 p_2'\, i_1^* & k_1 q_2\, i_1^* + a_2 + \mu_2 & -b_2 \\ -k_1 p_2\, i_1^* & -k_1 q_2\, i_1^* - a_2 & b_2 + m_2 \end{pmatrix}^{-1}. \tag{2.13}$$

We note that this matrix is of rank 1 so the spectral radius is equal to the trace. Hence, one gets

$$r_0^{\rm HIV} = f(0)\,(s_1^*\,\tau_{S_2} + e_1^*\,\tau_{E_2} + i_1^*\,\tau_{I_2})\,,$$

where τ_{S_2}, τ_{E_2} and τ_{I_2} are complex expressions with a simple interpretation. For example, τ_{S_2} is the life expectation of a person from the moment he/she enters state S_2 (in the linearized model). In particular, τ_{S_2}, τ_{E_2} and τ_{I_2} are all strictly less than $1/\mu_2$ if $m_2 > \mu_2$ (as should be). Therefore,

$$r_0^{\rm HIV} < R_0^{\rm HIV}.$$

Not surprisingly, the expected number of secondary HIV-cases produced by an "average" HIV$_+$ person in a population with endemic TB is less then in a population with no TB since active TB may shorten the life of such a person.

Similarly, the endemic steady state with HIV can be invaded by TB. Linearizing Eqs. (2.2)-(2.3)-(2.5)-(2.6) near $(\widehat{S}_1, 0, 0, \widehat{S}_2, 0, 0)$ and setting

$$\widehat{P} = \widehat{S}_1 + \widehat{S}_2\,, \quad \widehat{s}_1 = \widehat{S}_1/\widehat{P} = 1 - \widehat{H}\,, \quad \widehat{s}_2 = \widehat{S}_2/\widehat{P} = \widehat{H}\,,$$

we obtain

$$\begin{aligned}
\frac{dE_1}{dt} &\simeq p_1' \widehat{s_1}(k_1 I_1 + k_2 I_2) - (a_1 + \mu_1) E_1 + b_1 I_1 - f(\widehat{H})\, \widehat{H}\, E_1 \,, \\
\frac{dI_1}{dt} &\simeq p_1 \widehat{s_1}(k_1 I_1 + k_2 I_2) - (b_1 + m_1) I_1 + a_1 E_1 - f(\widehat{H})\, \widehat{H}\, I_1 \,, \\
\frac{dE_2}{dt} &\simeq p_2' \widehat{s_2}\,(k_1 I_1 + k_2 I_2) - (a_2 + \mu_2) E_2 + b_2 I_2 + f(\widehat{H})\, \widehat{H}\, E_1 \,, \\
\frac{dI_2}{dt} &\simeq p_2 \widehat{s_2}\,(k_1 I_1 + k_2 I_2) - (b_2 + m_2) I_2 + a_2 E_2 + f(\widehat{H})\, \widehat{H}\, I_1 \,.
\end{aligned}$$

It follows that the basic reproduction number r_0^{TB} for TB when introduced in a population at the HIV endemic steady state is the spectral radius of the matrix $M\,N^{-1}$, where

$$M = \begin{pmatrix} 0 & p_1' k_1 \widehat{s_1} & 0 & p_1' k_2\, \widehat{s_1} \\ 0 & p_1 k_1 \widehat{s_1} & 0 & p_1 k_2\, \widehat{s_1} \\ 0 & p_2'\, k_1 \widehat{s_2} & 0 & p_2' k_2\, \widehat{s_2} \\ 0 & p_2\, k_1 \widehat{s_2} & 0 & p_2\, k_2\, \widehat{s_2} \end{pmatrix} \tag{2.14}$$

and

$$N = \begin{pmatrix} a_1 + \mu_1 + f(\widehat{H})\, \widehat{H} & -b_1 & 0 & 0 \\ -a_1 & b_1 + m_1 + f(\widehat{H})\, \widehat{H} & 0 & 0 \\ -f(\widehat{H})\, \widehat{H} & 0 & a_2 + \mu_2 & -b_2 \\ 0 & -f(\widehat{H})\, \widehat{H} & -a_2 & b_2 + m_2 \end{pmatrix}.$$

The matrix M is the number of infections per unit time caused by different TB types (only active cases are infective). N is a state transition matrix, with N^{-1} representing the average time spent in each state. It is clear that the total number of infections caused is captured by MN^{-1}. This matrix is generalized into an age-structured operator in Ch. 4, and is used to study the relative impact of different active TB types, in settings with high HIV prevalence.

Whether r_0^{TB} is bigger or smaller than R_0^{TB} seems to depend on the parameter values chosen. Assuming realistically that $q_1 \leq p_1$ (so that there is no backward bifurcation for the model with TB but no HIV), this linear stability analysis suggests the following conjecture:

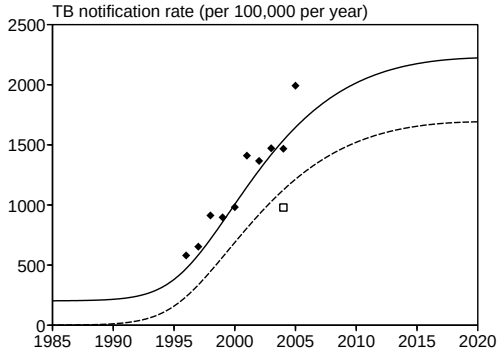
- when $R_0^{\mathrm{HIV}} < 1$ and $R_0^{\mathrm{TB}} < 1$, the disease-free steady state is a global attractor of system (2.1)-(2.6);
- when $R_0^{\mathrm{HIV}} > 1$ and $r_0^{\mathrm{TB}} < 1$, the HIV-endemic steady state is a global attractor;
- when $R_0^{\mathrm{TB}} > 1$ and $r_0^{\mathrm{HIV}} < 1$, the TB-endemic steady state is a global attractor;
- in all other cases, there is an endemic steady state with both HIV and TB, which has to be computed numerically, and which is a global attractor.

Since $R_0^{\mathrm{HIV}} > r_0^{\mathrm{HIV}}$, the fourth case contains in fact only two subcases:

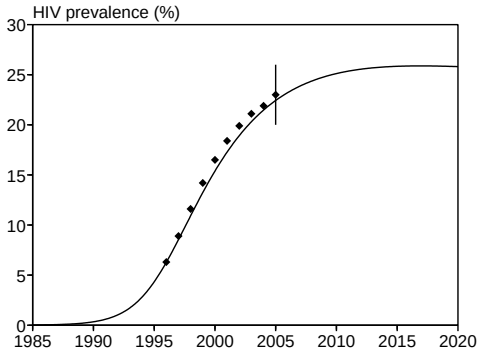
- $R_0^{\mathrm{HIV}} > 1$, $r_0^{\mathrm{TB}} > 1$, $R_0^{\mathrm{TB}} > 1$ and $r_0^{\mathrm{HIV}} > 1$. Both the HIV-endemic and the TB-endemic steady states exist but they are saddle points.

- $R_0^{\mathrm{HIV}} > 1$, $r_0^{\mathrm{TB}} > 1$, and $R_0^{\mathrm{TB}} < 1$. The HIV-endemic steady state exists but it is a saddle point. There is no TB-endemic steady state.

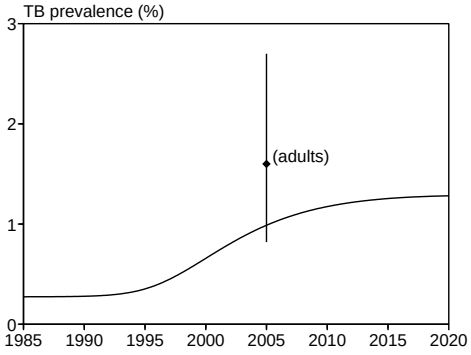
2.3 Simulation and parameter estimation



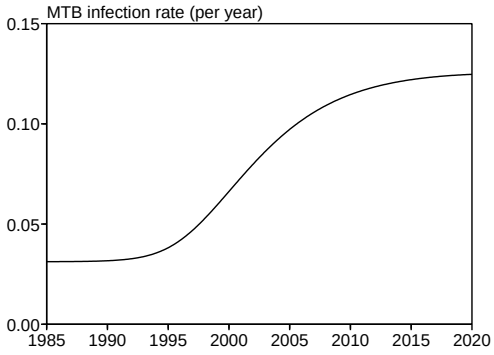
(a) TB notifications per 100,000



(b) HIV prevalence



(c) TB prevalence



(d) annual risk of MTB infection

Figure 2.2: (a) Data and simulation curve for the TB notification rate. The dashed curve shows the contribution of HIV+ individuals (only one data point). (b) Data and simulation curve for HIV prevalence. (c) Simulation curve for the prevalence of active TB. The data point with 95%CI corresponds to the prevalence of undiagnosed TB among adults. (d) MTB infection rate. The model parameters can be adjusted to give an MTB infection rate of 4% as suggested by a skin test survey in Masimpumelele performed in 2007 (see 4.1).

In Bacaër, et al. [17] we present a detailed discussion of how the parameters of our model were either fixed from literature or estimated by fitting data from Masiphumelele. This task was laborious since reliable data on both HIV and TB are still rare. We refer to Bacaër, et al. [Sect. 5–6][17] for a detailed discussion of the parameter estimation approach we took, and mention here parameters relevant to fitting to data in Masiphumelele.

The model was numerically solved with an ODE solver (Fig. 2.2). Note that in the simulation the peak for the prevalence of HIV (Fig. 2.2(b)) occurs at about the same time

Table 2.5: Numerical values for the parameters of the model.

	HIV$_-$			HIV$_+$		
mortality	μ_1	0.02/yr	[28]	μ_2	0.1/yr	[28]
TB mortality	m_1	0.25/yr	[30]	m_2	1.6/yr	[30]
MTB infections	k_1	11.4/yr	fit	k_2	$k_1 \times 2/3$	[28]
fast route	p_1	11%	[110]	p_2	30%	fit
slow route	a_1	0.0003/yr	[110]	a_2	0.08/yr	[18, 96, 97]
reinfection	q_1	$0.7\,p_1$	[110]	q_2	$0.75\,p_2$	[28]
recovery	β_1	0.25/yr	[30]	β_2	0.4/yr	[30]
detection	γ_1	0.74/yr	[29, 123]	γ_2	3.0/yr	[29, 123]
treatment	ε_1	80%	[123]	ε_2	80%	[123]
births	B	200/yr	[69]			
contact rate	d	0.7/yr	fit			
prevention	λ	5.9	fit			
initial year	t_0	1984	fit			

as the peak for the TB notification rate (Fig. 2.2(a)). This does not seem incompatible with data from Kenya [31, Fig. 1], which suggests a delay of several years between the rise of HIV and the rise of TB. One reason for such a delay may be that active TB tends to appear with a higher frequency in the late stages of HIV infection. We note, however, that the data from Masiphumelele does not show any clear delay.

Fig. 2.2(a) shows the contribution of HIV$_+$ cases to the TB notification rate. Together with the prevalence of TB at one point in time (see Fig. 2.2(c)), these two extra constraints should make our parameter estimates robust. However, we note that the annual risk of MTB infection (ARI), depicted in Fig. 2.2(d) is unrealistically high. The model parameters are slightly adjusted in Sect. 4.1 following the results (presented in 2008, after completing this work [17]) of a tuberculin skin test survey in 2007 among schoolchildren in the community, which suggest that the ARI has remained relatively constant over the last 15 years.

Detection rates γ_1 and γ_2. Wood et al. [123] reported 259 TB notifications among adults (age$\geq$ 15) in 2005; 66% of those who were tested for HIV were HIV$_+$. The adult population in that year was estimated to be 10,400. Moreover, in a sample population of 762 adults, 12 had undiagnosed TB (3 HIV$_-$ and 9 HIV$_+$). Therefore, we expect the following equations to hold:

$$\gamma_1\, I_1^{\text{adult}} \simeq 34\% \times 259, \quad I_1^{\text{adult}} \simeq 10,400 \times 3/762\,, \tag{2.15}$$

$$\gamma_2\, I_2^{\text{adult}} \simeq 66\% \times 259\,, \quad I_2^{\text{adult}} \simeq 10,400 \times 9/762\,. \tag{2.16}$$

This gives the estimates $\gamma_1 \simeq 2.2$ per year and $\gamma_2 \simeq 1.4$ per year. Note however that since the ratios 3/762 and 9/762 are small, the uncertainty is large: the 95% binomial confidence interval for the ratios 3/762 and 9/762 are $(0.08\%, 1.15\%)$ and $(0.54\%, 2.23\%)$ respectively. Using (2.15)-(2.16), the corresponding interval for γ_1 is $(0.74, 10.6)$ per year, and the one for γ_2 is $(0.74, 3.0)$ per year. Corbett et al. [29] suggest that γ_2 may be larger than γ_1. For our model, we chose the lower bound of the confidence interval for γ_1 ($\gamma_1 = 0.74$ per year) and the upper bound of the confidence interval for γ_2 ($\gamma_2 = 3.0$ per year). One motivation was that recent unpublished data shows that the MTB infection rate in the past few years have remained relatively constant – see Sect. 4.1. In our simulations, we found that this was only possible with values of γ_2 that are several times higher than γ_1. Indeed, the great increase in TB notifications has to be compensated by a shorter infectious period to keep the MTB infection rate at a relatively low level.

With these choices, we obtain $b_1 = \beta_1 + \gamma_1\varepsilon_1 \simeq 0.84$ per year and $b_2 = \beta_2 + \gamma_2\varepsilon_2 \simeq 2.8$ per year. For comparison, the values used for the whole of Uganda in [48] for b_1 and b_2 were both equal to 0.3 per year, but case detection is probably not as good as in Masiphumelele..

We note that the probabilities for TB to be detected are given by:

$$\frac{\gamma_1}{m_1 + \beta_1 + \gamma_1} \simeq 60\%, \quad \frac{\gamma_2}{m_2 + \beta_2 + \gamma_2} \simeq 60\%\,.$$

Despite the high death rate m_2, the detection probability for HIV_+ TB cases is the same as for HIV_- because of the high value of γ_2 used here. Recall that the target set by the World Health Organization for case detection is 70%. The average durations of disease are:

$$\frac{1}{b_1 + m_1} \simeq 0.92 \text{ year}\,, \quad \frac{1}{b_2 + m_2} \simeq 0.23 \text{ year}\,.$$

As a comparison, Corbett et al. [29] estimated the duration of (smear-positive) disease before diagnosis to be 1.15 year and 0.17 year for HIV_- and HIV_+ South African gold miners, respectively.

MTB transmission rate k_1. The average TB notification rate in the decade before 1995 in South Africa, i.e. before the rise of HIV, was about 200 per 100,000 per year (see [119] and [9, p. 184]). This is also a reasonable estimate for the township under study given the data from Tab. 2.1. We take $k_1 = 11.4$ per year, which corresponds to a TB notification rate of 203 per 100,000 per year.

HIV parameters d, λ and t_0. Summing the three equations (2.1)-(2.3) for HIV_- individuals and the three equations (2.4)-(2.6) for HIV_+ individuals, setting $X_1 = S_1 + E_1 + I_1$ and $X_2 = S_2 + E_2 + I_2$, and noticing that the prevalence of HIV is $H = X_2/(X_1 + X_2)$, we

obtain the system

$$\frac{dX_1}{dt} = B - \mu_1 X_1 - f(H) H X_1 + (\mu_1 - m_1) I_1 \ , \tag{2.17}$$

$$\frac{dX_2}{dt} = -\mu_2 X_2 + f(H) H X_1 + (\mu_2 - m_2) I_2 \ . \tag{2.18}$$

To get a first estimation of d, λ and t_0, we neglect the terms involving I_1 and I_2 (active TB cases form a very small proportion of the population). The resulting system involves only X_1 and X_2, and it is formally the same as system (2.11) for HIV without TB. Taking $X_1(t_0) = B/\mu_1$ and $X_2(t_0) = 1$, a good fit to HIV prevalence data from Tab. 2.1 is obtained with the parameters $d = 0.7$/year, $\lambda = 5.9$, and the year $t_0 = 1984$ for the beginning of the HIV epidemic. Three parameters are necessary and usually sufficient to fit any set of increasing numbers resembling the logistic curve, as is the case here.

The parameter p_2 for fast progression to TB among HIV$_+$ individuals. Di Perri et al. [35] studied an outbreak of TB among HIV$_+$ individuals: after the index case, eight people developed TB rapidly and six had a newly positive tuberculin skin test, suggesting that $8/14 \simeq 57\%$ of newly infected HIV$_+$ individuals develop primary TB disease. Daley et al. [32] studied a similar outbreak and found a proportion equal to $11/15 \simeq 73\%$. However, it is possible that only large outbreaks are studied, and that outbreaks with less cases of primary TB disease either less notable or are not a good subject for publication. A similar bias would occur if we based our estimate for the probability of fast progression to TB among HIV$_-$ individuals on reports of TB outbreaks such as the one investigated in [67], during which 14 out of 41 newly infected individuals (34%) developed primary disease. As a result, we vary p_2 in order to fit the data concerning the TB notification rate from Tab. 2.1. For this purpose, we simulated system (2.1)-(2.6) starting from the initial condition

$$S_1(t_0) = S_1^* \, , \quad E_1(t_0) = E_1^* \, , \quad I_1(t_0) = I_1^* \, , \quad S_2(t_0) = 1 \, , \quad E_2(t_0) = 0 \, , \quad I_2(t_0) = 0 \, .$$

At this point all the parameters in Tab. 2.5 have already been fixed except p_2. A relatively good fit was obtained with $p_2 = 30\%$ (plain line in Fig. 2.2(a)), i.e., nearly 3 times the value p_1 for HIV$_-$ individuals. Note that this value for p_2 is still lower than the ones obtained by studying TB outbreaks among HIV$_+$ individuals [32, 35]. Given the mortality μ_2 previously chosen for HIV$_+$ individuals, the estimates for a_2 and p_2 correspond to a probability $a_2/(a_2 + \mu_2) \simeq 44\%$ of progressing slowly from latent to active TB and to a probability $p_2 + a_2/(a_2 + \mu_2) \simeq 74\%$ of developing active TB after infection by MTB.

2.4 Sensitivity of steady states with respect to changes in parameter values

To study the sensitivity of the model with respect to parameters, we used numerical solutions of the mathematical formulae of Sect. 2.2 for the steady states. First, the disease-free steady state with no HIV and no TB is $S_1^0 = 10,000$. We also obtain

$$R_0^{\mathrm{TB}} \simeq 1.3\,, \quad R_0^{\mathrm{HIV}} \simeq 7.0\,, \quad r_0^{\mathrm{TB}} \simeq 1.7\,, \quad r_0^{\mathrm{HIV}} \simeq 5.8\,.$$

The estimate $R_0^{\mathrm{TB}} \simeq 1.3$ is close to the range 0.6–1.2 mentioned in the review [87]. Using national HIV prevalence data from antenatal clinics, Williams et al. [121, 122] found a similar result for R_0^{HIV}, namely 6.4 ± 1.6. Furthermore, note that $r_0^{\mathrm{TB}} > R_0^{\mathrm{TB}}$: an "average" person newly infected with MTB will produce more secondary cases if introduced in a TB-free population where HIV is endemic than if introduced in a completely disease-free population. This is mainly because this "average" person is likely to be HIV_+, so its probability of progressing to active TB and of infecting other people is high (this depends on the numerical values of several parameters, including a_2, but not on the structure of the model). Finally, r_0^{HIV} is less than R_0^{HIV} as explained in §2.2. In some sense, TB slows down the HIV epidemic.

Fig. 2.3 shows a bifurcation diagram of the steady states in the (k_1, d) parameter space using the numerical values from Tab. 2.5 except of course for k_1 and d and assuming that the ratio k_2/k_1 is fixed. The black dot near the 2,000 per 100,000 per year level curve for the TB notification rate corresponds to the values of k_1 and d in Tab. 2.5. The boundaries between the four domains of the bifurcation diagram ("disease-free", "HIV", "TB", and "HIV+TB") are obtained by the solving the four equations $R_0^{\mathrm{HIV}} = 1$, $r_0^{\mathrm{HIV}} = 1$, $R_0^{\mathrm{TB}} = 1$ and $r_0^{\mathrm{TB}} = 1$ with respect to k_1 and d. Since R_0^{HIV} does not depend on k_1 and R_0^{TB} does not depend on d, the line $R_0^{\mathrm{HIV}} = 1$ is horizontal and the line $R_0^{\mathrm{TB}} = 1$ is vertical. The line $r_0^{\mathrm{HIV}} = 1$ separates "TB" from "HIV+TB". The line $r_0^{\mathrm{TB}} = 1$ separates "HIV" from "HIV+TB".

Note in Fig. 2.3 how the level curves for the TB notification rate are distorted as they cross the line $r_0^{\mathrm{HIV}} = 1$ from the area labeled "TB" to the area labeled "HIV+TB". Notification rates near the "reinfection threshold" mentioned in Sect. 2.2 (for example the 1,000 and 2,000 level curves), which seemed totally unrealistic in the absence of HIV, occur now for smaller values of the transmission rate k_1 if HIV prevalence is high enough. With $k_1 = 11.4$ per year as in Tab. 2.5, the steady state TB notification rate increases from 200 to 2,000 per 100,000 per year as HIV prevalence increases from 0 to about 25%.

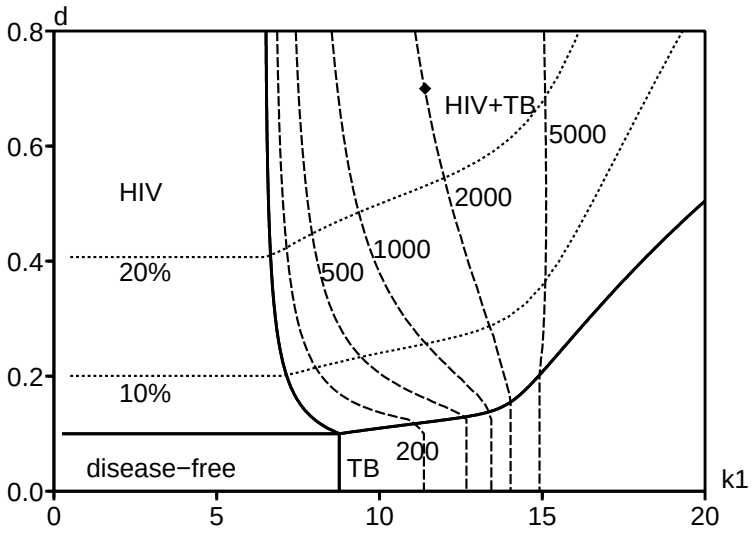


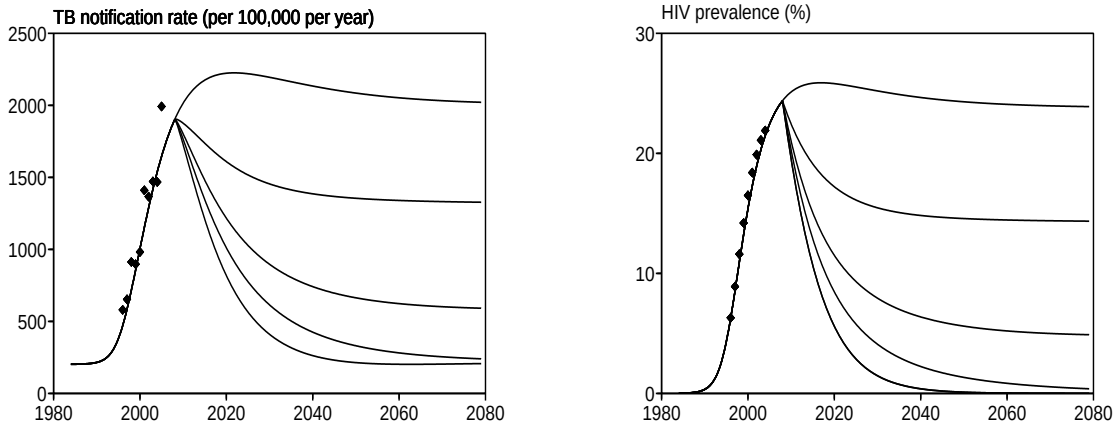
Figure 2.3: Bifurcation diagram in the (k_1, d) phase plane and level curves of the steady state TB notification rate (dashed lines, 500 stands for 500 per 100,000 per year) and of the steady state prevalence of HIV (dotted lines).

2.5 Impact of control measures

Increasing condom use

We note from Eqs. (2.13)-(6.9) that r_0^{HIV} is proportional to $f(0) = d$ (the maximum transmission rate of HIV) and that r_0^{TB} is proportional to k_1 (the maximum transmission rate of TB), the ratio k_2/k_1 being fixed. So if d is divided by at least r_0^{HIV} (the other parameters being kept constant), the new r_0^{HIV} will be less than 1 and HIV will disappear in the long run. Similarly, if k_1 is divided by at least r_0^{TB}, the new r_0^{TB} will be less than 1 and TB will disappear in the long run. In Fig. 2.3, starting from the black dot representing the real situation, one can check that if k_1 is divided by $r_0^{\mathrm{TB}} \simeq 1.7$, we move from the area labeled "HIV+TB" to the area with HIV only. If d is divided by $r_0^{\mathrm{HIV}} \simeq 5.8$, we move from the area "HIV+TB" to the area with TB only. To decrease the parameter k_1, living conditions should be changed. The parameter d decreases if more condoms are used.

Fig. 2.4 shows the impact of a sudden decrease of the HIV transmission rate d, from an initial value d to a new value d', on the prevalence of HIV (Fig. 2.4(b)) and also indirectly on the TB notification rate (Fig. 2.4(a)). The impact is obviously a monotonic function of d', as one would expect. We can check on these simulations that HIV disappears in the long run only if $d' < d/r_0^{\mathrm{HIV}} \simeq d/5.8$ (that is in the two simulations $d' = d/8$ and $d' = 0$ but not when $d' = d$, $d' = d/2$ or $d' = d/4$). If so, the TB notification rate returns finally to its level of the beginning of the 1980's, before HIV was introduced. The asymptotic TB notification rate and prevalence of HIV can also be read directly from the level curves in Fig. 2.3, but the speed at which these steady states are reached can only be seen in Fig. 2.4.



(a) reduction in TB notification rate (b) reduction in HIV prevalence

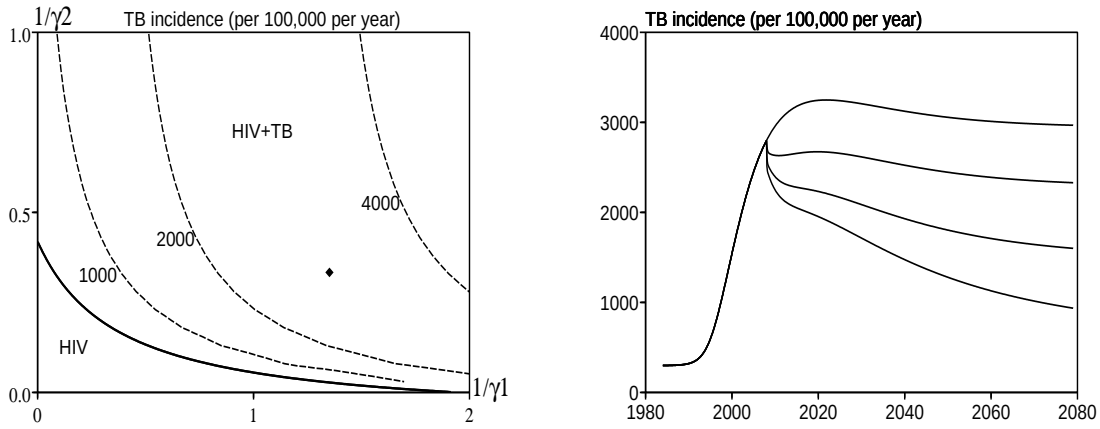
Figure 2.4: Assuming that a sudden increase in condom use occurs in the year 2008 (the maximum transmission rate d becomes d'). The different curves correspond from top to bottom to $d' = d$, $d' = d/2$, $d' = d/4$, $d' = d/8$ and $d' = 0$. (a) TB notification rate. (b) Prevalence of HIV.

Increasing TB detection

Now we consider the possibility of increasing the TB detection rates γ_1 and γ_2 and increasing the probabilities ε_1 and ε_2 of successful treatment. For the township, this could be achieved by actively searching for TB cases instead of waiting for them to come to the TB clinic. We note that the four parameters above enter the system of differential equations (2.1)-(2.6) only through the combination $b_1 = \beta_1 + \gamma_1\,\varepsilon_1$ and $b_2 = \beta_2 + \gamma_2\,\varepsilon_2$. However, γ_1 and γ_2 enter in the expression of the TB notification rate (through $\gamma_1\,I_1 + \gamma_2\,I_2$). If γ_1 or γ_2 increase, the steady state TB notification rate may increase and will start decreasing only if γ_1 or γ_2 are high enough. It is therefore not suitable to use the TB notification rate as a measure of the severity of the situation when the detection rate changes. Instead, we will use the TB incidence rate.

Fig. 2.5(a) shows the bifurcation diagram and the level curves of the steady state TB incidence rate in the parameter space $(1/\gamma_1, 1/\gamma_2)$, using the numerical values from Tab. 2.5 for the other parameters. Since γ_1 and γ_2 do not enter in the formula for R_0^{HIV}, the HIV-endemic steady state is always there. The question is: when can it be invaded by TB? This is given by the equation $r_0^{\mathrm{TB}} = 1$, an implicit equation for γ_1 and γ_2 shown by the thick black line separating "HIV" from "HIV+TB" in the bottom left corner of Fig. 2.5(a). The values for γ_1 and γ_2 in Tab. 2.5 correspond to the black dot shown in the figure.

Fig. 2.5(b) shows the impact of a sudden increase in the TB detection rate γ_2 for HIV_+ individuals. This has almost no impact on the curve for the prevalence of HIV so we do not show it. Of course, the TB incidence decreases monotonically as the detection rate increases.



(a) Bifurcation diagram

(b) TB incidence rate

Figure 2.5: Increasing the TB detection rate: (a) Bifurcation diagram in the phase plane $(1/\gamma_1, 1/\gamma_2)$ and level curves of the TB incidence rate. (b) TB incidence rate as a function of time, assuming that a sudden increase in the TB detection rate for HIV_+ individuals occurs in the year 2008. The parameter γ_2 is replaced from top to bottom by γ_2, $2\gamma_2$, $4\gamma_2$ or $8\gamma_2$.

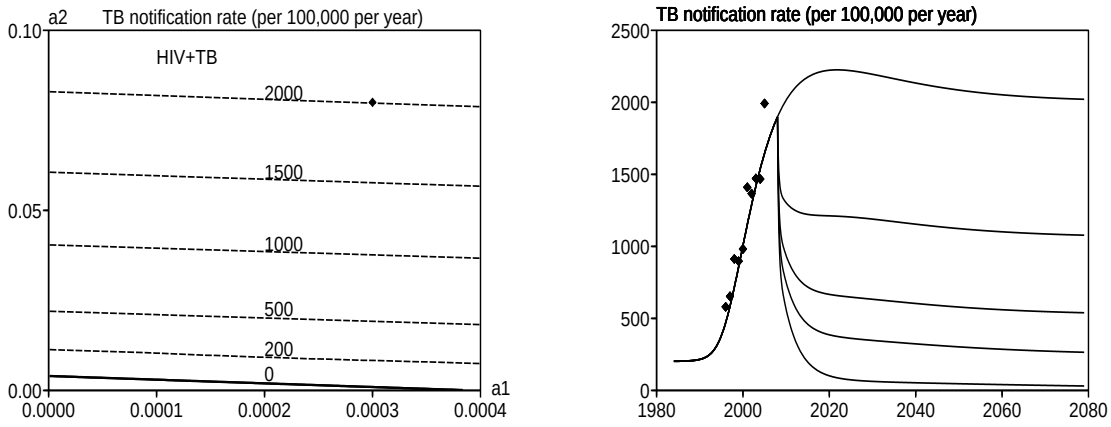
Isoniazid preventive therapy

This control measure reduces the parameter a_1 if used for HIV_- individuals and the parameter a_2 if used for HIV_+ individuals. These parameters do not enter in the formula for R_0^{HIV}, so HIV is always present and the question is whether TB can be stopped in the presence of HIV: the threshold is given by $r_0^{\text{TB}} = 1$ (the corresponding curve appears in the bottom of Fig. 2.6(a) as the level set 0). The level curves of the TB notification rate in the diagram (a_1, a_2) are almost horizontal (Fig. 2.6(a)). Therefore preventive therapy used for HIV_+ individuals (reducing a_2) has a much greater impact on the TB notification rate than if used for HIV_- individuals (reducing a_1). The values for a_1 and a_2 in Tab. 2.5 correspond to the black dot in Fig. 2.6(a) close to the 2,000 per 100,000 per year level curve.

Fig. 2.6(b) shows the impact of a sudden decrease of the progression rate a_2 for HIV_+ individuals due to isoniazid preventive therapy. Since this has almost no impact on the curve for the prevalence of HIV, we do not show it. The steady state TB notification rate decreases monotonically as a_2 decreases.

Antiretroviral treatment (ART)

ART reduces viral load and therefore also the transmission parameter d for HIV. On the other hand, ART also increases the life expectancy of HIV_+ individuals by decreasing μ_2 and m_2 (of course not below the natural mortality μ_1), which increases the number of people living with HIV and enhances further transmission of HIV. These two effects are antagonistic. The impact on HIV at the population level is not obvious and depends in detail on how much each of the three parameters involved changes with ART. Besides, ART



(a) bifurcation diagram (b) TB notification rate

Figure 2.6: Isoniazid preventive therapy for HIV$_+$ individuals (decreasing a_2): (a) Bifurcation diagram in the phase plane (a_1, a_2) and level curves of the steady state TB notification rate. (b) TB notification rate as a function of time. Assumption: starting in 2008, a_2 is replaced from top to bottom by a_2, $a_2/2$, $a_2/4$, $a_2/8$ or 0.

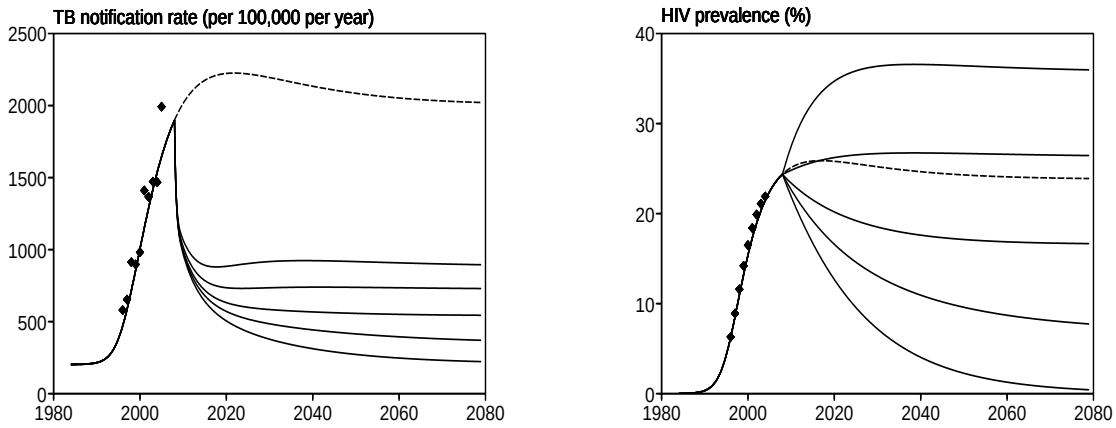
reduces the average rate a_2 at which co-infected individuals develop active TB, though not to the same level a_1 as HIV$-$ MTB-infected individuals [18, 68, 70, 74], and even if "immune reconstitution disease" may on the contrary increase a_2 during the first few months of ART treatment [72]. Again, the effect of ART on TB is uncertain because HIV$_+$ individuals under ART live longer. Quantitatively, ART was shown in studies in South Africa [18, 70] and Brazil [81] to reduce a_2 by 80% , i.e., to divide a_2 by 5. With $a_2 = 0.08$ per year without ART, this gives $a_2 = 0.016$ per year under ART. This is still 50 times higher than the parameter $a_1 = 0.0003$ per year for HIV$_-$ individuals. Another report [71] mentioned a risk 5 to 10 times higher after three years of ART compared to HIV$_-$ individuals. We assume furthermore that:

- μ_2 is divided by 2 under ART, giving $\mu_2 = 0.05$ per year instead of 0.1 per year, still higher than the natural mortality $\mu_1 = 0.02$ per year; the new life expectancy for HIV$_+$ individuals under ART is 20 years;
- m_2 is divided by 2 under ART (the new m_2 is 0.8 per year, compared to $m_1 = 0.25$ per year).

We determined what would happen under various assumptions for the HIV transmission parameter d (Fig. 2.7), assuming that 100% of HIV$_+$ individuals are immediately put on ART starting in 2008, independently of their CD4 cell count (a variable which is not included in our model anyway). This hypothesis is of course quite optimistic and would require the entire adult population of the township to be tested for HIV. Furthermore, note that in practice and in more realistic models, some factors may favour a delayed initiation of ART

[75]. With our choice of parameter values, we find a decrease for the TB notification rate even in the extreme case where ART would have no influence on the parameter d (Fig. 2.7(a), top plain curve), a case which would lead to an increase in HIV prevalence (Fig. 2.7(b), top plain curve).

The cases where $d' = d/2$ and $d' = d/4$ are probably more realistic, since we expect HIV transmission to decrease if everybody knows his/her HIV status. In such cases (and assuming that the other parameters values have been correctly chosen), HIV prevalence would decrease for $d' = d/4$ but not for $d' = d/2$ (Fig. 2.7(b), second and third plain curves from the top). So the future of HIV prevalence under ART is uncertain. However, with a progression rate a_2 reduced by 80% and a life expectancy $1/\mu_2$ multiplied by 2, it seems that ART would dramatically decrease the TB notification rate even though the new reactivation rate for HIV_+ individuals would still be several times higher than the one for HIV_- individuals.



(a) reduction in TB notification rate

(b) reduction in HIV prevalence

Figure 2.7: ART. (a) TB notification rate as a function of time. (b) HIV prevalence as a function of time. Assumption: 100% of HIV_+ individuals are put on ART starting in 2008. The parameter μ_2 is replaced by $\mu_2/2$, the parameter m_2 by $m_2/2$, the parameter a_2 by $a_2/5$, while the parameter d is replaced either by d, $d/2$, $d/4$, $d/8$, or 0 (from top to bottom). The dashed line shows the case without intervention.

2.6 Conclusions

This work provide a solid basis for modeling the simultaneous HIV and TB epidemics in a township near Cape Town, South Africa, for which a considerable amount of data is available. Deliberately keeping the number of parameters as small as possible enabled us to provide a fairly complete mathematical picture of the model with HIV or TB only. We used numerical methods to study a model of dual HIV and TB infection.

Before using the model to study the impact of interventions, we investigated the sensitivity of steady-state values for TB notification rates and HIV prevalence with respect to

key parameters. We used a bifurcation diagram of the steady states in the (TB contact rate, HIV contact rate) space. The diagram has four domains, delineating 'disease-free', 'HIV-only', 'TB-only', and 'HIV+TB' epidemics. The diagram shows how steady-state TB notification increases sharply as the prevalence of HIV increases.

Among the control measures studied, most have an obvious positive impact in controlling the HIV or TB epidemics. This is the case for condom use, increased TB detection and preventative treatment. The situation for ART is more complex and the impact depends in detail on how much it reduces mortality and TB activation rates, because these rates in turn determine the duration of an active TB episode. However, although the future for the prevalence of HIV in Masiphumelele is uncertain, it seems that a generalized access to ART might lead to a significant decrease of the TB notification rate.

If HIV_+ individuals under ART live approximately 2 times longer than the average 10 year survival time of HIV_+ individuals with no access to ART, then one could expect the TB incidence to be reduced by 60%. We investigated what would happen if ART were to increase the prevalence of HIV and indirectly the incidence of TB. Even in the worst case scenario we considered, where HIV prevalence increased as a result of ART, the TB notification rate decreased considerably. Our understanding of the impact of ART on both the HIV and TB epidemics will improve, as ART become more available in the Masiphumelele.

Chapter 3

Fluctuations and correlations in a model with TB only

The system of ODEs used in Sect. 2.1 models the time evolution of the macroscopic variables relating to TB and HIV interaction. It is an accurate macroscopic description when the population is 'large', mixes homogenously and fluctuations due to individual stochasticity are small. However, the population modeled here consists of only 10,000 individuals and small-population effects may limit the validity of the macroscopic equations. The method presented here can be used to help quantify uncertainty associated with macroscopic equations. It models fluctuations as a multivariate Gaussian distribution, which can be incorporated into a Bayesian model fitting framework.

We use a 'system size expansion' technique developed by Van Kampen [107, Ch. 10] to study the stochastic process underlying Eq. (2.1)-(2.6), and to gain insight into the statistical properties of the system's fluctuations. A Fokker-Planck approximation for the master equation of the system is derived. Using this equation we are able derive equations for the variances and co-variances of the fluctuations. We also study the decay of fluctuations near the equilibrium state. This method thus gives a handle on both the deterministic and stochastic descriptions of the system [89, Ch. 8].

However, it must be noted that the Fokker-Planck equation (FPE) is a valid approximation to the master equation of the system in the sense of a large population size expansion [43, Ch. 7]. Given that the population we model is not large, what we obtain is not a definitive description of the statistical properties of fluctuations but an insight into how these depend on population size. The approximation becomes more valid as population size increases. The method can be used to study fluctuations and correlations of Markov processes in a systematic way. Birth-death processes, such as the one explored in this chapter, are an important example in population dynamics.

Recent DNA fingerprinting work in Masiphumele allows us to classify TB strains, sampled from active TB cases, by their DNA type. Within the scope of commonly used epidemiological modeling tools, there is currently no method available for classifying temporal

clusters of TB events. What is the timescale beyond which two events are not statistically correlated, when they are of the same DNA type? We use point process theory to explore this question. Probability density functions for TB events are derived directly from the underlying dynamical model, for events that lead to active TB episodes. Using these density functions, two-time correlation functions can be calculated.

3.1 The master equation of a TB-only model

Consider again the compartmental model of Sect. 2.1, with S, E and I representing susceptible, latent and active TB cases respectively, ignoring the HIV-related part of this model. That is, we ignore all terms with subscript 2 in Eqs. (2.1)-(2.6), and ignore subscript 1 in the remaining terms:

$$\begin{aligned} \frac{dS}{dt} &= B - kS_1 I/P - \mu S \\ \frac{dE}{dt} &= (p'\,S - q\,E)\,kI/P - (a+\mu)\,E + bI \\ \frac{dI}{dt} &= (p\,S + q\,E)\,kI/P - (b+m)\,I + aE\,. \end{aligned} \tag{3.1}$$

To study this system microscopically we have to derive a 'master equation' for the system. This is an equation for the time-dependent transition probability of individuals who experience a finite (and small) number of events. The following equations state these events (e_i, $i = 1...9$) and the probability that they occur in a small time interval Δt:

$$\begin{aligned} P\{\mathrm{e}_1\} &= \frac{p'k}{\Omega}\,SI\,\Delta t + \mathcal{O}(\Delta t) \\ P\{\mathrm{e}_2\} &= \frac{pk}{\Omega}\,SI\,\Delta t + \mathcal{O}(\Delta t) \\ P\{\mathrm{e}_3\} &= a\,E\Delta t + \mathcal{O}(\Delta t) \\ P\{\mathrm{e}_4\} &= \frac{qk}{\Omega}\,EI\Delta t + \mathcal{O}(\Delta t) \\ P\{\mathrm{e}_5\} &= b\,I\Delta t + \mathcal{O}(\Delta t) \\ P\{\mathrm{e}_6\} &= \mu S\,\Delta t + \mathcal{O}(\Delta t) \\ P\{\mathrm{e}_7\} &= \mu E\,\Delta t + \mathcal{O}(\Delta t) \\ P\{\mathrm{e}_8\} &= mI\,\Delta t + \mathcal{O}(\Delta t) \\ P\{\mathrm{e}_9\} &= B\Delta t + \mathcal{O}(\Delta t)\,, \end{aligned}$$

where

- We have used the same notation of the simple model presented in Sect. 2.1, dropping the subscript 1. The total population size is now given by Ω, with P used for the probability of finding the population in a certain state.

- e_1- a susceptible becomes latently infected due to contact with an infective.
- e_2- a susceptible becomes actively infected due to contact with an infective.
- e_3- a latently infective becomes actively infected due to natural activation.
- e_4- a latently infective becomes actively infected due to exogenous reinfection.
- e_5- an infective becomes susceptible through recovery.
- e_6- a susceptible is removed due to death.
- e_7- a latently infective is removed due to death.
- e_8- an actively infected is removed due to death.
- e_9- a new susceptible enters the population.
- The time interval $\Delta\, t$ is small enough for only one event to occur in each time interval of this duration.
- Each individual of each category S(susceptible), E(latently infected), I(actively infected) is equally likely to experience events for that category.

Using this notation we can write the balance equation for the probability of having S susceptibles, E latently infectives and I actively infectives as follows:

$$\begin{aligned}\frac{1}{\Delta t}\,P_{(S,E,I)} \;=\;& \frac{p'k}{\Omega}(S+1)I\,P_{(S+1,E-1,I)} + \frac{pk}{\Omega}(S+1)(I-1)\,P_{(S+1,E,I-1)} \\ &+\frac{qk}{\Omega}(E+1)(I-1)\,P_{(S,E+1,I-1)} + a(E+1)\,P_{(S,E+1,I-1)} \\ &+b(I+1)\,P_{(S,E-1,I+1)} + \mu(S+1)\,P_{(S+1,E,I)} + \mu(E+1)\,P_{(S,E+1,I)} \\ &+m(I+1)\,P_{(S,E,I+1)} + \Omega B\,P_{(S-1,E,I)} \\ &-\left(1-p\frac{k}{\Omega}SI - p_1'\frac{k}{\Omega}SI - aE - bI - \mu S - \mu S - mI - B\Omega\right)\,P_{(S,E,I)}\,.\end{aligned}$$

We now introduce a simplifying notation which allows us to write the master equation more compactly. Let $n_1 = S$, $n_2 = E$, $n_3 = I$, with $n = (n_1, n_2, n_3)$ representing the state of the population. We also introduce 'step operators' (or 'raising' and 'lowering' operators), $D_{n_i} : n_i \mapsto n_i + 1$ and $D_{n_i}^{-1} : n_i \mapsto n_i - 1$. Using this notation, the master equation can be written as:

$$\begin{aligned}\frac{d}{dt}P_n &= \frac{p'k}{\Omega}\big[D_{n_1}D_{n_2}^{-1}-1\big]n_1n_3\,P_n+\frac{pk}{\Omega}\big[D_{n_1}D_{n_3}^{-1}-1\big]n_1n_3\,P_n\\ &+\frac{qk}{\Omega}\big[D_{n_2}D_{n_3}^{-1}-1\big]n_2n_3\,P_n\\ &+a\big[D_{n_2}D_{n_3}^{-1}-1\big]n_2\,P_n+b\big[D_{n_2}^{-1}D_{n_3}-1\big]n_3\,P_n\\ &+\mu\big[D_{n_1}-1\big]n_1\,P_n+\mu\big[D_{n_2}-1\big]n_2\,P_n+\mu\big[D_{n_3}-1\big]n_3\,P_n\\ &+B\Omega\big[D_{n_1}^{-1}-1\big]\,P_n\end{aligned}$$

.

3.2 Expansion of the master equation

For a large population system we expect $P_{S,E,I}$ to peak around the macroscopic values given by the macroscopic Eq. (3.1). The size of fluctuations are expected to be of the order $\Omega^{1/2}$ and their effect on macroscopic equations are expected to be of the order $\Omega^{-1/2}$, where Ω is the size of the system. A systematic expansion of $P_{S,E,I}$ in powers of $\Omega^{-1/2}$ was developed by van Kampen [107, Ch. 10]. In this section we will apply this large Ω expansion technique, outlining briefly details which are elaborated in [107, Ch. 10]. An application of this technique to a simple SIR model can be found in [27] and [89, Ch. 8.5].

Working from the assumption that $P_{S,E,I}$ is peaked around macroscopic averages, with a width of the order $\Omega^{1/2}$, we transform the n_i to new random variables y_i:

$$\begin{aligned}n_1(t) &= \Omega\mathbb{S}(t)+\Omega^{\frac{1}{2}}y_1(t)\,,\\ n_2(t) &= \Omega\mathbb{E}(t)+\Omega^{\frac{1}{2}}y_2(t)\,,\\ n_3(t) &= \Omega\mathbb{I}(t)+\Omega^{\frac{1}{2}}y_3(t)\,.\end{aligned}$$

$\mathbb{S}(t),\mathbb{E}(t)$ and $\mathbb{I}(t)$ are not *a priori* identified with averages of n_1, n_2 and n_3 respectively. However, it will emerge that they are precisely these averages [79].

The operators $D_{n_i}, D_{n_i}^{-1}$ can be replaced by their Taylor expansions:

$$D_{n_i}=\exp\left(\frac{1}{\sqrt{\Omega}}\frac{\partial}{\partial y_i}\right)=1+\Omega^{-\frac{1}{2}}\frac{\partial}{\partial y_i}+\frac{1}{2!}\Omega^{-1}\frac{\partial^2}{\partial y_i}+\ldots \tag{3.2}$$

$$D_{n_i}^{-1}=\exp\left(-\frac{1}{\sqrt{\Omega}}\frac{\partial}{\partial y_i}\right)=1-\Omega^{-\frac{1}{2}}\frac{\partial}{\partial y_i}+\frac{1}{2!}\Omega^{-1}\frac{\partial^2}{\partial y_i}+\ldots \tag{3.3}$$

Using these expansions, the new probability is

$$P_n\left(\Omega\mathbb{S}(t)+\Omega^{\frac{1}{2}}y_1(t),\,\Omega\mathbb{E}(t)+\Omega^{\frac{1}{2}}y_2(t),\,\Omega\mathbb{I}(t)+\Omega^{\frac{1}{2}}y_3(t)\right)=\Psi(y,t)\,, \tag{3.4}$$

which is a time-dependent probability distribution for the fluctuations, has normalization condition:

$$1=\sum_{n_1,n_2,n_3}P\left(n_1,n_2,n_3,t\right)=\int\int\int\,\Psi(y_1,y_2,y_3,t)\,dy_1\,dy_2\,dy_3\,,$$

and obeys a master equation of the form

$$\frac{\partial\Psi}{\partial t}-\Omega^{1/2}\sum_{i=1:3}\frac{d\mathbb{S}}{dt}\frac{\partial\Psi}{\partial y_i}=\sum_{i=1}^{9}T_i+\mathcal{O}(\Omega^{-\frac{1}{2}})\,, \tag{3.5}$$

where $T_1,...,T_9$ are given by:

$$\begin{aligned}
T_1 &= \frac{p'k}{\Omega}\left\{\left[1+\Omega^{-1/2}\left(\frac{\partial}{\partial y_1}-\frac{\partial}{\partial y_2}\right)+\frac{1}{2}\Omega^{-1}\left(\frac{\partial}{\partial y_1}-\frac{\partial}{\partial y_2}\right)^2\right]-1\right\}\times\\
&\quad\left[\mathbb{S}\mathbb{I}\Omega^2+\left(\mathbb{S}y_3+\mathbb{I}y_1\right)\Omega^{3/2}+y_1y_3\Omega\right]\Psi(t)\\
T_2 &= \frac{pk}{\Omega}\left\{\left[1+\Omega^{-1/2}\left(\frac{\partial}{\partial y_1}-\frac{\partial}{\partial y_3}\right)+\frac{1}{2}\Omega^{-1}\left(\frac{\partial}{\partial y_1}-\frac{\partial}{\partial y_3}\right)^2\right]-1\right\}\times\\
&\quad\left[\mathbb{S}\mathbb{I}\Omega^2+\left(\mathbb{S}y_3+\mathbb{I}y_1\right)\Omega^{3/2}+y_1y_3\Omega\right]\Psi(t)\\
T_3 &= \frac{qk}{\Omega}\left\{\left[1+\Omega^{-1/2}\left(\frac{\partial}{\partial y_2}-\frac{\partial}{\partial y_3}\right)+\frac{1}{2}\Omega^{-1}\left(\frac{\partial}{\partial y_2}-\frac{\partial}{\partial y_3}\right)^2\right]-1\right\}\times\\
&\quad\left[\mathbb{E}\mathbb{I}\Omega^2+\left(\mathbb{E}y_3+\mathbb{I}y_2\right)\Omega^{3/2}+y_2y_3\Omega\right]\Psi(t)\\
T_4 &= a\left\{\left[1+\Omega^{-1/2}\left(\frac{\partial}{\partial y_2}-\frac{\partial}{\partial y_3}\right)+\frac{1}{2}\Omega^{-1}\left(\frac{\partial}{\partial y_2}-\frac{\partial}{\partial y_3}\right)^2\right]-1\right\}\times\\
&\quad\left[\Omega\mathbb{E}+\Omega^{1/2}y_2\right]\Psi(t)
\end{aligned}$$

$$
\begin{aligned}
T_5 &= b\left\{\left[1+\Omega^{-1/2}\left(\frac{\partial}{\partial y_3}-\frac{\partial}{\partial y_2}\right)+\frac{1}{2}\Omega^{-1}\left(\frac{\partial y}{\partial y_3}-\frac{\partial}{\partial y_2}\right)^2+...\right]-1\right\}\times \\
&\quad \left[\Omega\mathbb{I}+\Omega^{1/2}y_3\right]\Psi(t) \\
T_6 &= \mu\left\{\left[1+\Omega^{-1/2}\left(\frac{\partial}{\partial y_1}\right)+\frac{1}{2}\Omega^{-1}\left(\frac{\partial}{\partial y_1}\right)^2\right]-1\right\}\times \\
&\quad \left[\Omega\mathbb{S}+\Omega^{1/2}y_1\right]\Psi(t) \\
T_7 &= \mu\left\{\left[1+\Omega^{-1/2}\left(\frac{\partial}{\partial y_2}\right)+\frac{1}{2}\Omega^{-1}\left(\frac{\partial}{\partial y_2}\right)^2\right]-1\right\}\times \\
&\quad \left[\Omega\mathbb{E}+\Omega^{1/2}y_2\right]\Psi(t) \\
T_8 &= m\left\{\left[1+\Omega^{-1/2}\left(\frac{\partial}{\partial y_3}\right)+\frac{1}{2}\Omega^{-1}\left(\frac{\partial}{\partial y_3}\right)^2\right]-1\right\}\times \\
&\quad \left[\Omega\mathbb{I}+\Omega^{1/2}y_3\right]\Psi(t) \\
T_9 &= B\Omega\left\{\left[1-\Omega^{-1/2}\left(\frac{\partial}{\partial y_1}\right)+\frac{1}{2}\Omega^{-1}\left(\frac{\partial}{\partial y_1}\right)^2-1\right]\right\}\Psi(t)\,.
\end{aligned}
$$

The macroscopic equations of the system are given by equating terms of the order $\Omega^{1/2}$:

$$
\begin{aligned}
-\dot{\mathbb{S}} &= -B+k\mathbb{I}\mathbb{S}+\mu\mathbb{S}\,, \\
-\dot{\mathbb{E}} &= -kp'\mathbb{I}\mathbb{S}+q\mathbb{E}+a\mathbb{E}+\mu\mathbb{E}\,, \\
-\dot{\mathbb{I}} &= -kp\mathbb{I}\mathbb{S}-q\mathbb{E}+b\mathbb{I}+m\mathbb{I}\,.
\end{aligned}
$$

which are equivalent to the system of equations (3.1) we started from. The equations can be solved numerically using the initial conditions:

$$
\mathbb{S}(0)=\frac{S(0)}{\Omega},\ \mathbb{E}(0)=\frac{E(0)}{\Omega}\ \mathbb{I}(0)=\frac{I(0)}{\Omega}\,. \tag{3.6}
$$

Equating terms of the order Ω^0 gives:

$$
\begin{aligned}
\frac{\partial \Psi(y,t)}{\partial t} &= p'k\left(\frac{\partial}{\partial y_1}-\frac{\partial}{\partial y_2}\right)(\mathbb{S}y_3+\mathbb{I}y_1)\;\Psi+\frac{1}{2}p'k\left(\frac{\partial}{\partial y_1}-\frac{\partial}{\partial y_2}\right)^2\mathbb{S}\mathbb{I}\,\Psi \\
&+pk\left(\frac{\partial}{\partial y_1}-\frac{\partial}{\partial y_3}\right)(\mathbb{S}y_3+\mathbb{I}y_1)\;\Psi+\frac{1}{2}pk\left(\frac{\partial}{\partial y_1}-\frac{\partial}{\partial y_3}\right)^2\mathbb{S}\mathbb{I}\,\Psi \\
&+qk\left(\frac{\partial}{\partial y_2}-\frac{\partial}{\partial y_3}\right)(\mathbb{E}y_3+\mathbb{I}y_2)\;\Psi+\frac{1}{2}qk\left(\frac{\partial}{\partial y_2}-\frac{\partial}{\partial y_3}\right)^2\mathbb{E}\mathbb{I}\,\Psi \\
&+a\left(\frac{\partial}{\partial y_2}-\frac{\partial}{\partial y_3}\right)y_2\,\Psi+\frac{1}{2}a\left(\frac{\partial}{\partial y_2}-\frac{\partial}{\partial y_3}\right)^2\mathbb{E}\,\Psi \\
&+b\left(\frac{\partial}{\partial y_3}-\frac{\partial}{\partial y_2}\right)y_3\,\Psi+\frac{1}{2}b\left(\frac{\partial}{\partial y_3}-\frac{\partial}{\partial y_2}\right)^2\mathbb{I}\,\Psi \\
&+\mu\frac{\partial}{\partial y_1}y_1\,\Psi+\frac{1}{2}\,\mu\,\frac{\partial^2}{\partial y_1}\mathbb{S}\,\Psi+\mu\frac{\partial}{\partial y_2}y_2\,\Psi+\frac{1}{2}\,\mu\,\frac{\partial^2}{\partial y_1}\mathbb{E}\,\Psi \\
&+m\frac{\partial}{\partial y_3}y_3\,\Psi+\frac{1}{2}\,m\,\frac{\partial^2}{\partial y_3}\mathbb{I}\,\Psi+B\frac{1}{2}\frac{\partial}{\partial y_1}^{\,2}\Psi\,. \qquad (3.7)
\end{aligned}
$$

From this we derive a multivariate, linear and time-dependent FPE for Ψ:

$$
\frac{\partial \Psi(y,t)}{\partial t}=-\sum_{i,j}\mathbb{A}_{i,j}(t)\frac{\partial}{\partial y_i}\left(y_j\Psi\right)+\frac{1}{2}\sum_{i,j}\mathbb{B}_{i,j}(t)\left(\frac{\partial^2\Psi}{\partial y_i\partial y_j}\right), \qquad (3.8)
$$

where

$$
\begin{aligned}
\mathbb{A}(t) &= -\begin{pmatrix} k\mathbb{I}+\mu & 0 & k\mathbb{S} \\ -p'k\mathbb{I} & qk\mathbb{I}+a+\mu & -p'k\mathbb{S}+qk\mathbb{E}-b \\ -pk\mathbb{I} & -qk\mathbb{I}-a & -pk\mathbb{S}-qk\mathbb{E}+b+m \end{pmatrix}, \\
\mathbb{B}(t) &= \begin{pmatrix} B+k\mathbb{S}\mathbb{I}+\mu\mathbb{S} & -p'k\mathbb{S}\mathbb{I} & -pk\mathbb{S}\mathbb{I} \\ -p'k\mathbb{S}\mathbb{I} & a\mathbb{E}+b\mathbb{I}+p'k\mathbb{S}\mathbb{I}+q\mathbb{E}\mathbb{I}+\mu\mathbb{E} & -a\mathbb{E}-b\mathbb{I}-q\mathbb{E}\mathbb{I} \\ -pk\mathbb{S}\mathbb{I} & -a\mathbb{E}-b\mathbb{I}-q\mathbb{E}\mathbb{I} & a\mathbb{E}+b\mathbb{I}+m\mathbb{I}+pk\mathbb{S}\mathbb{I}+q\mathbb{E}\mathbb{I} \end{pmatrix}.
\end{aligned}
$$

Using the FPE (3.8) we can compute the moments of the fluctuations [107, Ch. 8, p. 211]:

$$
\begin{aligned}
\frac{d}{dt}\langle y_i\rangle &= \sum_j \mathbb{A}_{ij}\langle y_j\rangle\,, \\
\frac{d}{dt}\langle y_i y_j\rangle &= \sum_k \mathbb{A}_{ik}\langle y_k y_j\rangle+\sum_k \mathbb{A}_{jk}\langle y_i y_k\rangle+\mathbb{B}_{ij}\,. \qquad (3.9)
\end{aligned}
$$

By construction $Y_{ij}(t)=\langle y_i\, y_j\rangle$, being the variance of the multivariate Gaussian $\Psi(y,t)$, is symmetric and necessarily so is $\mathbb{B}$. Using these formulae we find the 'variational' equations:

$$
\begin{aligned}
\frac{d}{dt}\langle y_1\rangle &= (-k\mathbb{I}-\mu)\,\langle y_1\rangle - k\mathbb{S}\,\langle y_3\rangle + \mathcal{O}(\Omega^{-1})\,,\\
\frac{d}{dt}\langle y_2\rangle &= p'k\mathbb{I}\,\langle y_1\rangle + (-a-\mu-qk\mathbb{I})\,\langle y_2\rangle + \left(b+p'k\mathbb{S}-qk\mathbb{E}\right)\,\langle y_3\rangle + \mathcal{O}(\Omega^{-1})\,,\\
\frac{d}{dt}\langle y_3\rangle &= pk\mathbb{I}\,\langle y_1\rangle + (a+qk\mathbb{I})\,\langle y_2\rangle + (-b-m+pk\mathbb{S}+qk\mathbb{E})\,\langle y_3\rangle + \mathcal{O}(\Omega^{-1})\,. \quad (3.10)
\end{aligned}
$$

as well as equations for the variance and the co-variance of the fluctuations:

$$
\begin{aligned}
\frac{d}{dt}\langle y_1y_1\rangle &= B+\mu\mathbb{S}+pk\mathbb{S}\mathbb{I}+p'k\mathbb{S}\mathbb{I}+2\,(-k\mathbb{I}-\mu)\,\langle y_1y_1\rangle - k\mathbb{S}\,\langle y_1y_3\rangle - k\mathbb{S}\,\langle y_3y_1\rangle + \mathcal{O}(\Omega^{-\frac{1}{2}})\,,\\
\frac{d}{dt}\langle y_2y_2\rangle &= a\mathbb{E}+b\mathbb{I}+p'k\mathbb{S}\mathbb{I}+q\mathbb{E}\mathbb{I}+u\mathbb{E}+p'k\mathbb{I}\,\langle y_1y_2\rangle + p'k\mathbb{I}\,\langle y_2y_1\rangle + 2\,(-a-\mu-qk\mathbb{I})\,\langle y_2y_2\rangle\,,\\
&\quad + \left(b+p'k\mathbb{S}-qk\mathbb{E}\right)\,\langle y_2y_3\rangle + \left(b+p'k\mathbb{S}-qk\mathbb{E}\right)\,\langle y_3y_2\rangle + \mathcal{O}(\Omega^{-\frac{1}{2}})\,,\\
\frac{d}{dt}\langle y_3y_3\rangle &= a\mathbb{E}+b\mathbb{I}+m\mathbb{I}+pk\mathbb{S}\mathbb{I}+q\mathbb{E}\mathbb{I}+pk\mathbb{I}\,\langle y_1y_3\rangle + (a+qk\mathbb{I})\,\langle y_2y_3\rangle + pk\mathbb{I}\,\langle y_3y_1\rangle\,,\\
&\quad + (a+qk\mathbb{I})\,\langle y_3y_2\rangle + 2\,(-b-m+pk\mathbb{S}+qk\mathbb{E})\,\langle y_3y_3\rangle + \mathcal{O}(\Omega^{-\frac{1}{2}})\,,\\
\frac{d}{dt}\langle y_1y_2\rangle &= -p'k\mathbb{S}\mathbb{I}+p'k\mathbb{I}\,\langle y_1y_1\rangle + (-k\mathbb{I}-\mu)\,\langle y_1y_2\rangle + (-a-\mu-qk\mathbb{I})\,\langle y_1y_2\rangle\,,\\
&\quad + \left(b+p'k\mathbb{S}-qk\mathbb{E}\right)\,\langle y_1y_3\rangle - k\mathbb{S}\,\langle y_3y_2\rangle + \mathcal{O}(\Omega^{-\frac{1}{2}})\,,\\
\frac{d}{dt}\langle y_1y_3\rangle &= -pk\mathbb{S}\mathbb{I}+pk\mathbb{I}\,\langle y_1y_1\rangle + (a+qk\mathbb{I})\,\langle y_1y_2\rangle + (-k\mathbb{I}-\mu)\,\langle y_1y_3\rangle\,,\\
&\quad + (-b-m+pk\mathbb{S}+qk\mathbb{E})\,\langle y_1y_3\rangle - k\mathbb{S}\,\langle y_3y_3\rangle + \mathcal{O}(\Omega^{-\frac{1}{2}})\,,\\
\frac{d}{dt}\langle y_2y_3\rangle &= -a\mathbb{E}-b\mathbb{I}-q\mathbb{E}\mathbb{I}+p'k\mathbb{I}\,\langle y_1Y_3\rangle + pk\mathbb{I}\,\langle y_2y_1\rangle + (a+qk\mathbb{I})\,\langle y_2y_2\rangle\,,\\
&\quad + (-b-m+pk\mathbb{S}+qk\mathbb{E})\,\langle y_2y_3\rangle + (-a-\mu-qk\mathbb{I})\,\langle y_2y_3\rangle\,,\\
&\quad + \left(b+p'k\mathbb{S}-qk\mathbb{E}\right)\,\langle y_3y_3\rangle + \mathcal{O}(\Omega^{-\frac{1}{2}})\,. \quad (3.11)
\end{aligned}
$$

In this way fluctuations are modeled as a multivariate Gaussian distribution, the moments of which are given by Eqs. (3.10)-(3.11). An assumption, which is not verified here, was made that fluctuations are symmetric about the mean, and this may not be true for birth-death processes in general. The derivation also relies on the global stability of the stationary state. Note that the fluctuations are not correctly described by a Gaussian distribution during the exponential growth phase of the TB epidemic. In this region, fluctuations are no longer just of order $\Omega^{\frac{1}{2}}$ compared to macroscopic values. However, the stability of the macroscopic variables guarantees that fluctuations are bounded [107, Ch. 10.4].

Stochastic simulation

An equivalent description of the system can be given in terms of waiting times until competing 'events' that occur in the system. We identify the following time inhomogeneous Poisson processes, identifying also their intensities $a_i, i = 1...9$ at time t:

$$
\begin{aligned}
a_1 &= \frac{p'k}{\Omega} S(t)I(t)\,, \\
a_2 &= \frac{pk}{\Omega} S(t)I(t)\,, \\
a_3 &= a\,E(t)\,, \\
a_4 &= \frac{qk}{\Omega} E(t)I(t)\,, \\
a_5 &= b\,I(t)\,, \\
a_6 &= \mu S(t)\,, \\
a_7 &= \mu E(t)\,, \\
a_8 &= mI(t)\,, \\
a_9 &= B\,.
\end{aligned}
$$

Define a_0 as the intensity of the Poisson process corresponding to the first event:

$$
a_0 = \sum_{i=1,2,3,4,5,7,8,9} a_i\,.
$$

We can use a_0 to compute the average waiting time to the next event, and use it to decide which event it should be (on average). To this end, draw two numbers r_1 and r_2 from the unit uniform distribution. The time to the next event is given by sampling the inverse CDF of an exponential distribution with rate a_0:

$$
\tau = \frac{1}{a_0} * \log\left(\frac{1}{r_1}\right)\,. \tag{3.12}
$$

The event type is given by the smallest integer u for which:

$$
\sum_{v=1}^{u-1} a_v \;\leq\; r_2\, a_0 \leq \sum_{v=1}^{u} a_v\,. \tag{3.13}
$$

Having found τ and u by this method, the time evolution of the system is straightforward: total time is incremented by τ, the event is applied to update the relevant sub-population and an age of τ is added to each living population member. Suppose that an infection event ($u = 1$ or $u = 2$) is the next event. A susceptible person is selected at random, after which the population counts S, E and I are updated. This is the so-called Gillespie algorithm, widely used to simulate stochastic processes [45]. It can be shown that the Gillespie algorithm gives 'exact' realization of the stochastic process underlying systems given by a system of coupled ODEs. Sect. 7.2 uses this simulation technique in a more complex model.

The Gillespie algorithm can be used to study fluctuations of the system around equilib-

rium, using the stationary values for S, E, I in a model with TB only as initial condition. In Fig. 3.1(a,b,c) we show a stochastic realization of a TB epidemic (without HIV) in black. The stochastic simulation algorithm is started with the macroscopic steady state as initial condition. This avoids the problems of non-zero extinction probabilities of the disease free steady state, which leads to many stochastic realizations dying out. In red we show one standard deviation of the fluctuations. Note that the size of the fluctuations are large during the exponential growth phase, when the system undergoes a bifurcation.

It is clear that there are substantial small-population effects present in this model, and that fluctuations are large relative to macroscopic values. The question is: what population size will lead to relatively small fluctuations, and thus to an accurate system of macroscopic equations? In Fig. 3.1(d) we show the ratio of the size of the stationary fluctuations in I_1 with respect to the stationary size of I_1, plotted as a function of the total population size. To increase the population size we raise the birth rate from 200 per year to 900 per year. This leads to a four-fold increase in the population size. Fig. 3.1(d) illustrates the relationship between fluctuations and population size in a model where a sub-population is of particular interest. In this case, we focused on I_1, the category of active TB individuals.

Standard Deviation	Gaussian Approximation	Simulation
S	300	316
E	291	280
I	12	12

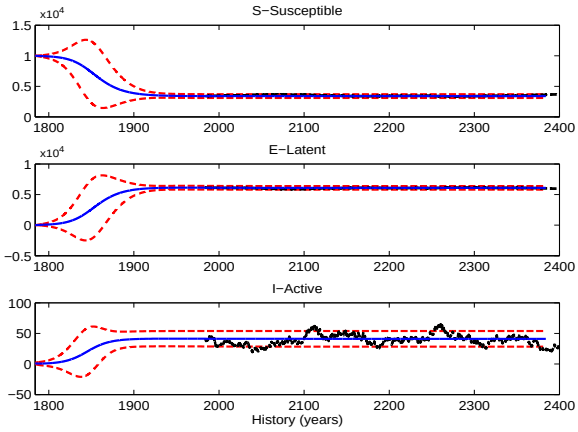
Table 3.1: Comparing calculation of standard deviation of fluctuations with estimation from simulation. The Gaussian approximation to fluctuations at equilibrium is reasonably accurate.

The expansion method used in Sect. 3 can be employed to study fluctuations in the HIV-TB compartmental of Sect. 2.1. However, the nonlinear function of HIV prevalence $f(H)$, which represents the sexual contact rate, with the parametric form

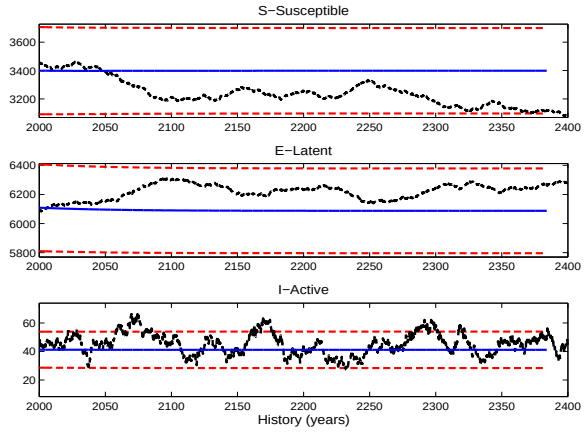
$$f(H) = d\,\mathrm{e}^{-\lambda H},$$

leads to tedious algebraic steps in the expansion. It is possible to go through a similar derivation to obtain a linear FPE with 6 by 6 matrices $\mathbb{A}$ and $\mathbb{B}$. If a standard form for HIV infection were used the algebraic steps would be easier, but then modeling the prevalence of HIV would become challenging.

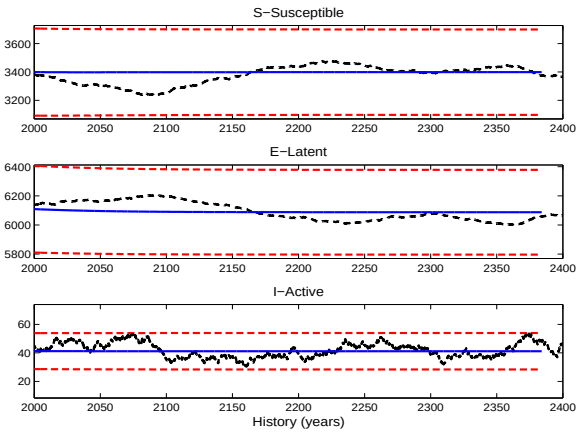
According to [45] the Gillespie algorithm gives exact realizations of stochastic processes underlying systems given by coupled ODEs. The question arises as to whether the FPE also provides an exact method to study fluctuations in such systems. In Tab. 3.1 a comparison is made between the standard deviation of fluctuations as modeled by the FPE (3.7)-(3.8) approach and by stochastic simulation. It seems that the FPE for the fluctuations in the system gives the same description for the variance of noise around the macroscopic values, as would be obtained from a large number of exact stochastic realizations. In Sect. 3.3 the



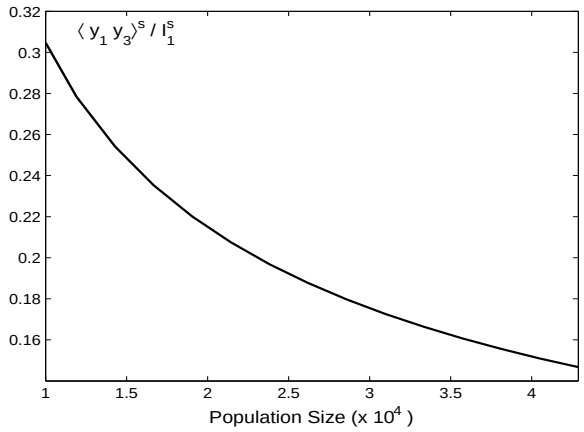
(a) stochastic and macroscopic dynamics



(b) up close



(c) average of five stochastic simulations



(d) relative influence of population size

Figure 3.1: (a) Macroscopic dynamics for categories S, E and I, i.e. the macroscopic state variables, are shown in blue. The standard deviation of the fluctuations around the macroscopic values is shown in red. Also shown is one stochastic realization in black. Note that the fluctuations are large during the exponential growth phase of the epidemic. During this time, the system undergoes a phase transition. It appears that fluctuations are also growing exponentially during this phase and the variance of fluctuations does not seem to be a useful quantity during this phase [105]. (b) An enlarged view of the fluctuations at the stationary state. (c) Fluctuations around the macroscopic values obtained by averaging five stochastic realizations. The average of five realizations lies more or less within one standard deviation of the mean. (d) The ratio between the size of the fluctuations in I_1 and macroscopic values for I_1 as a function of the total population size. Compartment I_1 is used because it is the smallest compartment in the model and the most relevant, being the compartment of active TB cases. This shows how the relative influence of the fluctuations decreases as a function of total population size, and gives an idea of what population size is required to give an accurate macroscopic description in which the influence of fluctuations will be minimal.

clustering of active TB events in the stationary state is studied using the Fokker-Planck description. A comparison is made with the clustering observed in stochastic realizations obtained with the Gillespie algorithm.

Autocorrelation matrix

The average fluctuations obey:

$$\frac{d}{dt}\langle y,t|y_0,t_0\rangle = \mathbb{A}\langle y,t|y_0,t_0\rangle\,,$$

for the time derivative of the average value of y at time t, given that the value was y_0 at time $t_0 = 0$. The solution is given by:

$$\langle y,t|y_0,t_0\rangle = e^{\mathbb{A}t}\cdot\langle y_0,t_0\rangle\,. \tag{3.14}$$

At equilibrium, the correlation matrix dissipates according to:

$$\begin{aligned}\langle y_i(0)y_j(t)\rangle_s &= e^{\mathbb{A}t}\cdot\langle y_i(0)y_j(0)\rangle_s \\ &= e^{\mathbb{A}t}\cdot\begin{pmatrix}\langle y_1y_1\rangle^s & \langle y_1y_2\rangle^s & \langle y_1y_3\rangle^s\\ \langle y_2y_1\rangle^s & \langle y_2y_2\rangle^s & \langle y_2y_3\rangle^s\\ \langle y_3y_1\rangle^s & \langle y_3y_2\rangle^s & \langle y_3y_3\rangle^s\end{pmatrix}\end{aligned}$$

using the stationary values for all fluctuations at time 0. We see that fluctuations from the equilibrium value take a long time to disappear. The slow reaction of the system results from the slow dynamics of exponentially distributed events.

The ideas presented here are captured by Onsager's principle (the so-called linear noise theory), which states that the influence of fluctuations about the equilibrium state decay, on average, according to the same macroscopic laws which govern the recovery of the macroscopic system to equilibrium [90, p. 532].

In Fig. 3.3 we show the dissipation of fluctuations in our TB-only model. The equilibrium size of the fluctuations are used as initial condition in Eq. (3.14). The simulation gives as estimate of the timescale over which fluctuations and correlations will influence the system. We proceed with a more thorough investigation into the statistical properties of fluctuations in this TB model.

3.3 Temporal clustering of active TB events

Data is available on the clustering of TB events. This data set comprises TB sputum samples for which the DNA types (i.e. strains) of TB have been determined. It shows that active TB episodes have a tendency to cluster, in the sense that a few cases of the same DNA type are registered in a short time interval. Analysis of these TB clusters make use of statistical

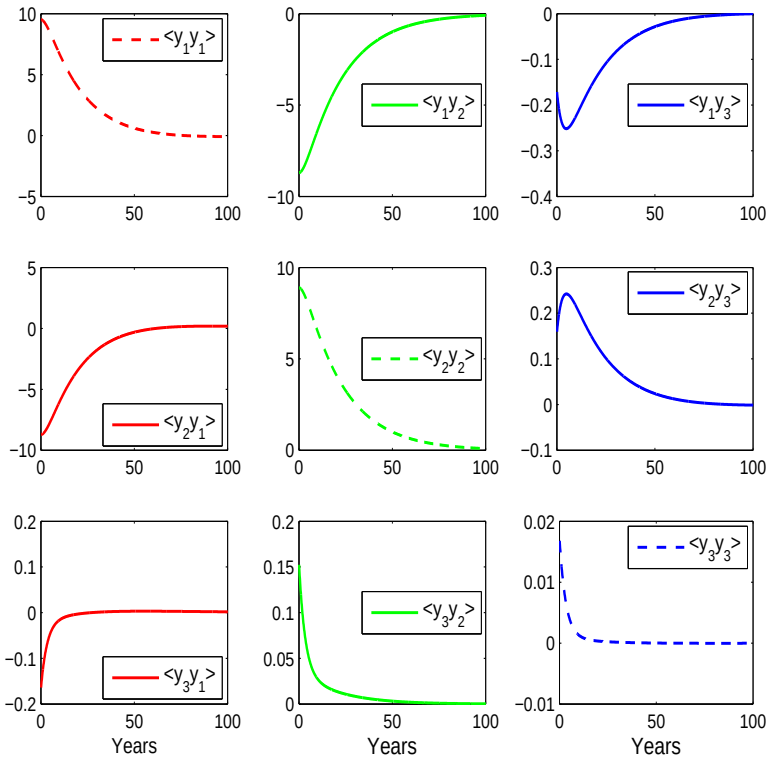


Figure 3.2: Visualization of dissipation of fluctuation near equilibrium. Note that the derivation relies on the global stability of the equilibrium state.

methods for clustered data, and ignores the fact that the TB episodes are generated by an underlying dynamical model. We study a model with one strain of TB to focus on the temporal aspects of TB event clustering.

Our technique for simulating realizations of the underlying stochastic process for TB infection (see Sect. 3.2) treats TB events as Poissonian. Fig. 3.3 shows simulated active TB events. A simulation result of all episode types, namely endogenous- or self-activation (Fig. 3.3(a)), primary infection events (Fig. 3.3(b)) and reinfection events (Fig. 3.3(c)), seems to show clustering. This raises the question as to whether a simulation algorithm, such as the Gillespie method, preserves correlation measures between different events. More specifically, whether it can be used to estimate timescales below which clustered events are likely, and beyond which events are uncorrelated.

Fig. 3.4 shows a histogram of the waiting times between active TB events. The waiting times to next endogenous-activation is approximately exponentially distributed with a mean of 0.98 years. This corresponds exactly to a average waiting time of $1/\left(a_1\langle E_1\rangle_s\right) = 0.98$ where $\langle S_1\rangle_s$ is the stationary value of E_1. The mean waiting time between simulated primary infection events is 0.83 years, which corresponds to $1/\left(p_1 k_1 \langle S_1\rangle_s \langle I_1\rangle_s/\Omega\right) = 0.82$. Similarly, the mean waiting time between simulated reinfection events is 0.64 years, which corresponds to $1/\left(q_1 k_1 \langle E_1\rangle_s \langle I_1\rangle_s/\Omega\right) = 0.65$. The Gillespie algorithm clearly preserves the average waiting time between events in the stationary state. However, since exponential waiting

times correspond to Poisson processes, which produce uncorrelated events, we do not expect the clusters seen in Fig. 3.3 to present a higher than average density of events.

Having derived a master equation description for individual TB events, we now use the theory of point processes to study te statistical properties of active TB events. We are able to derive a definition for the temporal elements of clustering, which includes a timescale over which points are expected to be correlated.

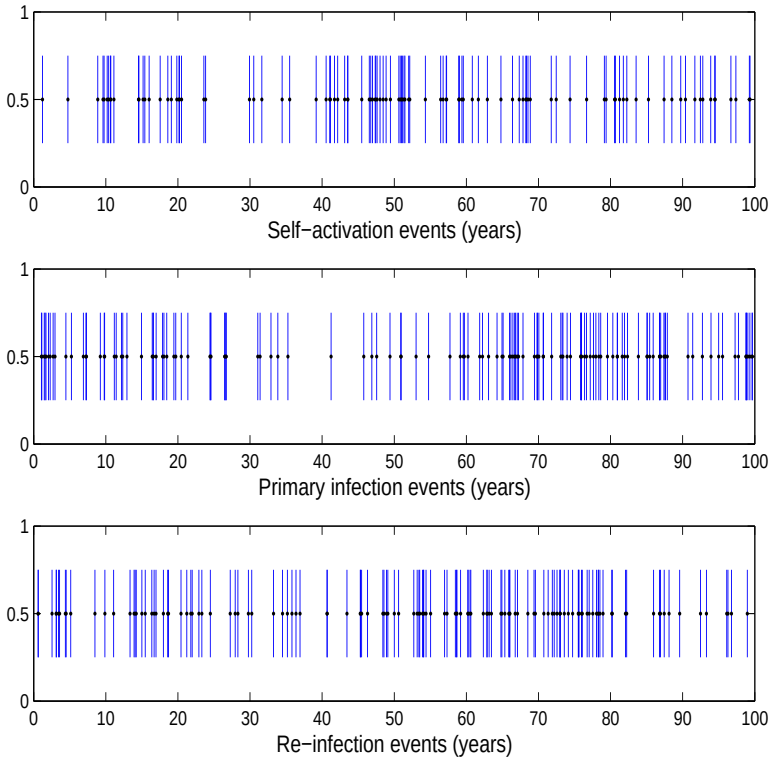


Figure 3.3: Visualization of TB endogenous-activation events as a point process.

A short overview of point process theory

The overview of point process theory presented here is practical and avoids most of the underlying mathematical machinery, for which the two volume work of Daley et al. [33] and the book of van Kampen [66] are excellent references. The language used here will be closer to that used for a physics approach to the problem, reviewing the work of [107].

The state space for a general point process consists of [107]:

- a nonnegative integer $s = 1, 2, 3, ...$
- for each s a set of real numbers obeying

$$-\infty < \tau_1 < \tau_2 < \tau_3 < ... < \tau_s < \infty$$

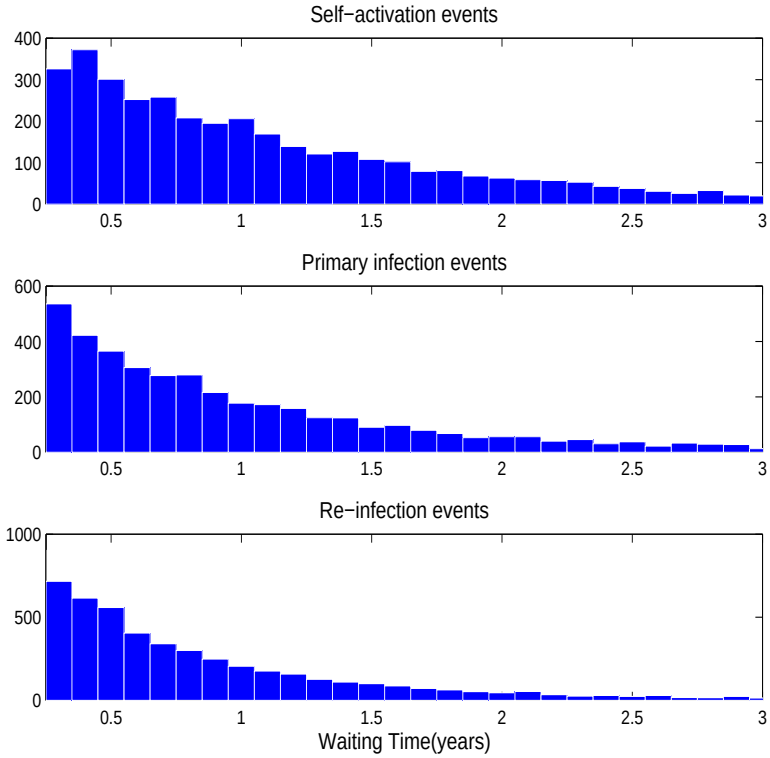


Figure 3.4: Distribution of simulated waiting times between active TB events.

- The state space can be thought of as the set of all sequences of points, where a state corresponds to a particular finite sequence of points.

A probability distribution over the state space is given by a sequence of non-negative functions $Q_s(\tau_1, \tau_2, ..., \tau_s)$. The Q_s are also known as Janossy densities, and they are normalized according to

$$Q_0 + \int_{-\infty}^{\infty} d\tau_1 \, Q_1(\tau_1) + \int_{-\infty}^{\infty} d\tau_1 \int_{\tau_1}^{\infty} d\tau_2 \, Q_1(\tau_1, \tau_2) + ... = 1 \, .$$

We now stipulate that the probability measures Q_s are symmetric and give equal weight to all $s!$ permutations of their arguments $(\tau_1, \tau_2, ..., \tau_s)$. The normalizing condition becomes

$$Q_0 + \sum_{s=1}^{\infty} \frac{1}{s!} \int_{-\infty}^{\infty} d\tau_1 d\tau_2 ... d\tau_s \, Q_s(\tau_1, \tau_2, ..., \tau_s) = 1 \, ,$$

An alternative specification of the point process is given by:

$$f_n(t_1, t_2, ..., t_n) = \sum_{s=n}^{\infty} \frac{1}{(s-n)!} \int_{-\infty}^{\infty} d\tau_{n+1} ... d\tau_s \; Q_s(t_1, ..., t_n, \tau_{n+1}, ..., \tau_s) \, , \tag{3.15}$$

with the intuitive meaning that $f_n(1, 2, ..., n) \, dt_1 dt_2 ... dt_n$ is the probability that there are

exactly n points, one in each of the infinitesimal intervals $(t_1, t_1+dt_1), (t_2, t_2+dt_2), ..., (t_n, t_n+dt_n)$ regardless of the number of points outside these intervals. This can be deduced directly from Eq. (3.15), where the first t_n arguments of Q_s are fixed, while averaging over or 'integrating out' the remaining $(n+1, ..., s)$ arguments. Note that f_1 represents the average density of dots (i.e. intensity of the process).

Another class of symmetric functions can be defined in terms of the functions g_m:

$$\begin{aligned} f_1(t_1) &= g_1(t_1) \\ f_2(t_1,t_2) &= g_1(t_1)g_1(t_2) + g_2(t_1,t_2) \\ f_3(t_1,t_2,t_3) &= g_1(t_1)g_1(t_2)g_1(t_3) \\ &\quad + g_1(t_1)g_2(t_1,t_2) + g_1(t_2)g_2(t_1,t_3) + g_1(t_3)g_2(t_1,t_2) \\ &\quad + g_3(t_1,t_2,t_3) \\ f_4(t_1,t_2,t_3,t_4) &= \cdots \\ &\vdots \end{aligned} \tag{3.16}$$

We shall make use of the g_m to study the clustering properties of active TB events. The reason for using the g_m functions is that they measure the degree to which events are uncorrelated. In particular, if the events are independent, then all the $g_m, m > 1$ vanish. A point process is said to 'have the cluster property' when [107, Ch. 2.5]:

$$\lim_{\tau\to\infty} g_{m+m'}(t_1, t_2, ..., t_m + \tau, t_{m+1} + \tau, ..., t_{m+m'} + \tau) = 0\,.$$

The function $g_2(t_1, t_2) = f_2(t_1, t_2) - f_1(t_1)f_1(t_2)$ represents a correlation function, and measures the degree to which the 'joint density' $f_2(t_1, t_1)$ is larger than the average density at t_1 and t_2. The timescale beyond which points are expected to be independent can be estimated from g_2. On shorter timescales it is likely that events occur in pairs, particularly on very short time intervals.

We derive this correlation function for events that lead to active TB episodes in a simple TB model (i.e. one without HIV) and use it to study the clustering properties of TB events around the stationary state.

Endogenous-activation TB events. The probability for a endogenous-activation TB event to take place between t_1 and $t_1 + dt_1$ is:

$$\begin{aligned} f_1(t_1) &= aE\,dt_1 \\ &= a\left\{\Omega\,\mathbb{E} + \Omega^{\frac{1}{2}}\langle y_2\rangle_s\right\} dt_1 \\ &= a\Omega\left\{\mathbb{E} + \Omega^{-\frac{1}{2}}\langle y_2\rangle_s\right\} dt_1\,, \end{aligned}$$

where y denotes fluctuations in the active TB events and $\langle y\rangle_s$ denotes the average fluctuation at steady state. The probability for having one event in (t_1, t_1+dt_1) and another independent event in (t_2, t_2+dt_2), with $t_2 \geq t_1$, is:

$$f_1(t_1)f_1(t_2) \;=\; a^2\Omega^2\mathbb{E}^2 + 2a^2\Omega^{\frac{3}{2}}\langle y_2\rangle_s + a^2\Omega\langle y_2\rangle_s^2\,.$$

Next we consider $f_2(t_1,t_2)dt_1\,dt_2$ which is the probability for having one event in $(t_1, t_1 + dt_1)$ and another possibly correlated event in the time interval $(t_2, t_2 + dt_2)$, with $t_2 \geq t_1$, regardless of whether events occur at other points in time:

$$\begin{aligned} f_2(t_1,t_2) \;=\; & a^2\Omega^2 \int\int \left\{\mathbb{E} + \Omega^{-\frac{1}{2}}\,P_2\,y_2\right\}\Psi\left(y_2,t_2|y_1+\overline{v},t_1\right)\times \\ & \left\{\mathbb{E} + \Omega^{-\frac{1}{2}}\,P_2\,y_1\right\}\Psi\left(y_1,t_1\right)\;d^3y_1 d^3y_2\,, \end{aligned} \tag{3.17}$$

where $\Psi(y)$ is the stationary distribution for fluctuations y. The vector $\overline{v} = \left(0, -\Omega^{-\frac{1}{2}}, \Omega^{-\frac{1}{2}}\right)^T$ is due to having one less latent and one more active TB case, after the endogenous-activation event at t_1. Note that y_1 and y_2 are 3-component vectors and the operator P_i projects to the i'th component of the vector to the right of the operator. When appearing in a subscript in an expectation, e.g. $\langle y_i\rangle$, i refers to the i'th component of the vector valued y.

To reduce Eq. (3.17) multiply the first $\{...\}$ factor with the second. Terms resulting from $\mathbb{E}$ in the first $\{...\}$ factor multiplied with the second $\{...\}$ factor are independent of y_2 and can easily be resolved. They are listed in the first line of the expression for f_2 below:

$$\begin{aligned} f_2(t_1,t_2) \;=\; & a^2\Omega^2\mathbb{E}^2 + a^2\Omega^{\frac{3}{2}}\mathbb{E}\langle y_2\rangle_s \\ & + a^2\Omega^{\frac{3}{2}}\mathbb{E}\int\int P_2\,y_2\Psi\left(y_2,t_2|y_1+\overline{v},t_1\right)\Psi\left(y_1,t_1\right)\;d^3y_1 d^3y_2 \\ & + a^2\Omega\int\int P_2\,y_2 y_1\Psi\left(y_2,t_2|y_1+\overline{v},t_1\right)\Psi\left(y_1,t_1\right)\;d^3y_1 d^3y_2\,. \end{aligned} \tag{3.18}$$

The second and third line of Eq. (3.18) can be reduced by means of Eq. (3.14). The integral in the second line states the conditional average of y_2, using $y_1 + \overline{v}$ as the initial condition at time t_1, averaged over the stationary value of y_1. Using Eq. (3.14) and averaging over the stationary distribution of y_1 we have:

$$a^2\Omega^{\frac{3}{2}}\mathbb{E}\int P_2\,e^{\mathbb{A}(t_2-t_1)}\left\{y_1+\overline{v}\right\}\;\Psi\left(y_1,t_1\right)d^3y_1\,.$$

We can now use the linearity of the integrand to rewrite the second line of Eq. (3.18):

$$a^2\Omega^{\frac{3}{2}}\mathbb{E}\,P_2\,e^{\mathbb{A}(t_2-t_1)}\langle y\rangle_s + a^2\Omega\mathbb{E}\,P_2\,e^{\mathbb{A}(t_2-t_1)}\,\overline{v}\,.$$

Similarly we can reduce the third line of Eq. (3.18):

$$a^2\Omega\int\int P_2\,y_2y_1\Psi\left(y_2,t_2|y_1+\overline{v},t_1\right)\Psi\left(y_1,t_1\right)\,d^3y_1d^3y_2\,,$$

to

$$a^2\Omega\,P_2\,e^{\mathbb{A}(t_2-t_1)}\langle y^2\rangle_s + a^2\Omega^{\frac{1}{2}}\,P_2\,e^{\mathbb{A}(t_2-t_1)}\langle y\rangle_s\,.$$

Collecting these results, and neglecting terms of order $\Omega^{\frac{1}{2}}$ and higher, we find:

$$\begin{aligned} g_2(t_1,t_2) &= f_2(t_1,t_2)-f_1(t_1)f_1(t_2) \\ &= a^2\Omega^{\frac{3}{2}}\mathbb{E}\,P_2\left(e^{\mathbb{A}(t_2-t_1)}-I\right)\langle y\rangle_s \\ &\quad +a^2\Omega\,P_2\left(e^{\mathbb{A}(t_2-t_1)}\langle y^2\rangle_s-\langle y\rangle_s^2\right) \\ &\quad +a^2\Omega\mathbb{E}\,P_2\,e^{\mathbb{A}(t_2-t_1)}\,\overline{v}\,. \end{aligned}$$

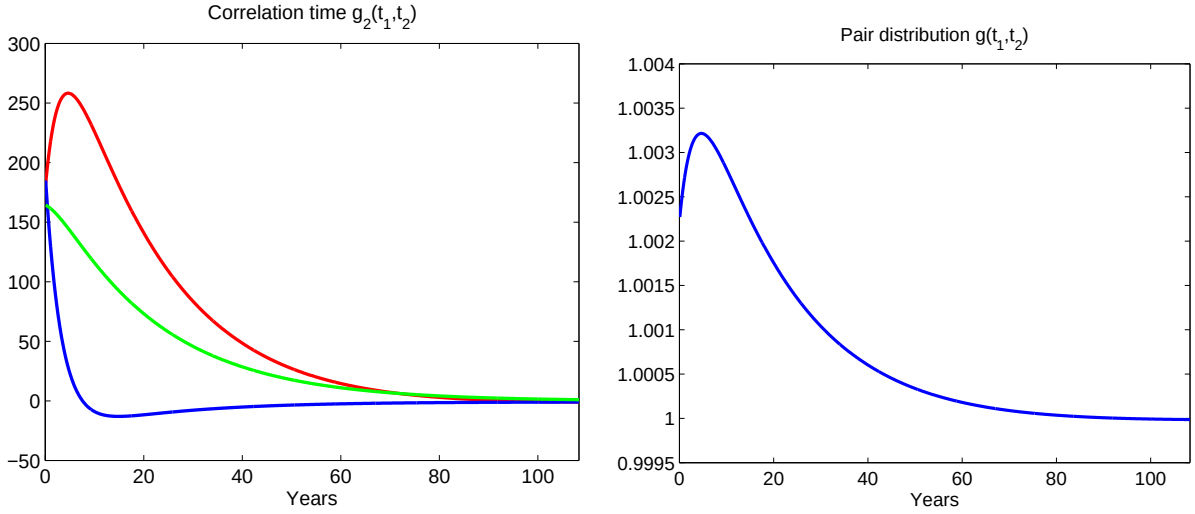
The correlation function $g_2(t_1,t_2)$ is normalized according to [25, Ch. 3.5]:

$$g_2(t_1,t_1+\tau)=\left[\lim_{\tau\to\infty}g_2(t_1,t_1+\tau)\right]^{-1}g_2(t_1,t_1+\tau)\,,$$

and displayed in Fig. 3.5(a,red), along with the correlation function for primary and reinfection events. The pair distribution function is given by:

$$g(t_a,t_b)=1+\frac{g_2(t_a,t_b)}{f_1(t_a)f_1(t_b)}\,.$$

From Fig. 3.5(b) we see that $g_2(t_1,t_2+\tau)<g_2(t_1,t_2)\;,\tau>0$. Thus, there is a tendency for endogenous-activation events to arrive in pairs (see [111, Chp 3.7]). On a long timescale, $\tau\to\infty$, it is certain that a second event will occur, but it is not correlated with the first event.



(a) correlation function (b) pair distribution function

Figure 3.5: (a) Normalized correlation function $g_2(t_1, t_2)$ of endogenous-activation events (red), primary infection event (blue) and reinfection events (green). (b) Pair distribution function for paired endogenous-activation events.

Primary TB infection events. The probability for a primary TB infection event to take place between t_1 and $t_1 + dt_1$ is:

$$\begin{aligned} f_1(t_1) &= \frac{k}{\Omega} SI\, dt_1 \\ &= k\Omega \left\{ \mathbb{S} + \Omega^{-\frac{1}{2}} \langle y_1 \rangle_s \right\} \left\{ \mathbb{I} + \Omega^{-\frac{1}{2}} \langle y_3 \rangle_s \right\} dt_1 \\ &= k\Omega \left\{ \mathbb{SI} + \Omega^{-\frac{1}{2}} \mathbb{S} \langle y_3 \rangle_s + \Omega^{-\frac{1}{2}} \mathbb{I} \langle y_1 \rangle_s + \Omega^{-1} \langle y_1 \rangle_s \langle y_3 \rangle_s \right\} dt_1 \,. \end{aligned}$$

The probability for having one primary infection event in $(t_1, t_1 + dt_1)$ and another independent event in $(t_2, t_2 + dt_2)$, where $t_2 \geq t_1$, is:

$$\begin{aligned} f_1(t_2) f_2(t_2) &= k^2 \Omega^2 \mathbb{S}^2 \mathbb{I}^2 + 2k^2 \Omega^{\frac{3}{2}} \mathbb{SI}^2 \langle y_1 \rangle_s + 2k^2 \Omega^{\frac{3}{2}} \mathbb{S}^2 \mathbb{I} \langle y_3 \rangle_s \\ &\quad + 4\Omega k^2 \mathbb{SI} \langle y_1 \rangle_s \langle y_3 \rangle_s + k^2 \Omega \mathbb{S}^2 \langle y_3 \rangle_s^2 + k^2 \Omega \mathbb{I}^2 \langle y_1 \rangle_s^2 \\ &\quad + 2k^2 \Omega^{\frac{1}{2}} \mathbb{I} \langle y_1 \rangle_s^2 \langle y_3 \rangle_s + 2k^2 \Omega^{\frac{1}{2}} \mathbb{S} \langle y_1 \rangle_s \langle y_3 \rangle_s^2 \\ &\quad + k^2 \langle y_1 \rangle_s^2 \langle y_3 \rangle_s^2 \,. \end{aligned}$$

The probability for having one primary infection event in $(t_1, t_1 + dt_1)$ and another possibly correlated event in the time interval $(t_2, t_2 + dt_2)$, with $t_2 \geq t_1$ is:

$$
\begin{aligned}
f_2(t_1,t_2) &= k^2\Omega^2 \int\int \left\{\mathbb{S}\mathbb{I} + \Omega^{-\frac{1}{2}}\mathbb{S}\, P_3\, y_2 + \Omega^{-\frac{1}{2}}\mathbb{I}\, P_1\, y_2 + \Omega^{-1}\, P_1\, y_2\, P_3\, y_2\right\} \\
&\quad \times \Psi\left(y_2, t_2 | y_1 + \overline{v}, t_1\right) \\
&\quad \times \left\{\mathbb{S}\mathbb{I} + \Omega^{-\frac{1}{2}}\mathbb{S}\, P_3\, y_1 + \Omega^{-\frac{1}{2}}\mathbb{I}\, P_1\, y_1 + \Omega^{-1}\, P_1\, y_1\, P_3\, y_1\right\} \\
&\quad \times \Psi\left(y_1, t_1\right)\; d^3y_1 d^3y_2\,,
\end{aligned}
\tag{3.19}
$$

where $\bar{v}$ is the vector $\bar{v} = \left(-\Omega^{-\frac{1}{2}}, 0, \Omega^{-\frac{1}{2}}\right)^T$. Upon calculation we have four terms corresponding to multiplying each term in the first {...} factor with the second {...} factor:

$$
\begin{aligned}
f_2(t_1,t_2) &= \sum_{i=1:4} T_i \text{ where,} \\
T_1 &= k^2\Omega^2\mathbb{S}^2\mathbb{I}^2 + k^2\Omega^{\frac{3}{2}}\mathbb{S}^2\mathbb{I}\langle y_3\rangle_s + k^2\Omega^{\frac{3}{2}}\mathbb{S}\mathbb{I}^2\langle y_1\rangle_s + k^2\Omega\mathbb{S}\mathbb{I}\langle y_1 y_3\rangle_s\,, \\
T_2 &= k^2\Omega^{\frac{3}{2}}\mathbb{S}^2\mathbb{I}\int\int y_2^{(3)}\Psi\left(y_2, t_2|y_1+\overline{v}, t_1\right)\Psi\left(y_1,t_1\right)\; d^3y_1\, d^3y_2 \\
&\quad + k^2\Omega\mathbb{S}^2\int\int y_2^{(3)}y_1^{(3)}\Psi\left(y_2, t_2|y_1+\overline{v}, t_1\right)\Psi\left(y_1,t_1\right)\; d^3y_1\, d^3y_2 \\
&\quad + k^2\Omega\mathbb{S}\mathbb{I}\int\int y_2^{(3)}y_1^{(1)}\Psi\left(y_2, t_2|y_1+\overline{v}, t_1\right)\Psi\left(y_1,t_1\right)\; d^3y_1\, d^3y_2 \\
&\quad + k^2\Omega^{\frac{1}{2}}\mathbb{S}\int\int y_2^{(3)}y_1^{(1)}y_1^{(3)}\Psi\left(y_2, t_2|y_1+\overline{v}, t_1\right)\Psi\left(y_1,t_1\right)\; d^3y_1\, d^3y_2\,, \\
T_3 &= k^2\Omega^{\frac{3}{2}}\mathbb{S}\mathbb{I}^2\int\int y_2^{(1)}\Psi\left(y_2, t_2|y_1+\overline{v}, t_1\right)\Psi\left(y_1,t_1\right)\; d^3y_1\, d^3y_2 \\
&\quad + k^2\Omega\mathbb{S}\mathbb{I}\int\int y_2^{(1)}y_1^{(3)}\Psi\left(y_2, t_2|y_1+\overline{v}, t_1\right)\Psi\left(y_1,t_1\right)\; d^3y_1\, d^3y_2 \\
&\quad + k^2\Omega\mathbb{I}^2\int\int y_2^{(1)}y_1^{(1)}\Psi\left(y_2, t_2|y_1+\overline{v}, t_1\right)\Psi\left(y_1,t_1\right)\; d^3y_1\, d^3y_2 \\
&\quad + k^2\Omega^{\frac{1}{2}}\mathbb{I}\int\int y_2^{(1)}y_1^{(1)}y_1^{(3)}\Psi\left(y_2, t_2|y_1+\overline{v}, t_1\right)\Psi\left(y_1,t_1\right)\; d^3y_1\, d^3y_2\,, \\
T_4 &= k^2\Omega\mathbb{S}\mathbb{I}\int\int y_2^{(1)}y_2^{(3)}\Psi\left(y_2, t_2|y_1+\overline{v}, t_1\right)\Psi\left(y_1,t_1\right)\; d^3y_1\, d^3y_2 \\
&\quad + k^2\Omega^{\frac{1}{2}}\mathbb{S}\int\int y_2^{(1)}y_2^{(3)}y_1^{(3)}\Psi\left(y_2, t_2|y_1+\overline{v}, t_1\right)\Psi\left(y_1,t_1\right)\; d^3y_1\, d^3y_2 \\
&\quad + k^2\Omega^{\frac{1}{2}}\mathbb{I}\int\int y_2^{(1)}y_2^{(3)}y_1^{(1)}\Psi\left(y_2, t_2|y_1+\overline{v}, t_1\right)\Psi\left(y_1,t_1\right)\; d^3y_1\, d^3y_2 \\
&\quad + k^2\int\int y_2^{(1)}y_2^{(3)}y_1^{(1)}y_1^{(3)}\Psi\left(y_2, t_2|y_1+\overline{v}, t_1\right)\Psi\left(y_1,t_1\right)\; d^3y_1\, d^3y_2\,.
\end{aligned}
$$

The vector $\bar{v} = \left(-\Omega^{-\frac{1}{2}}, 0, \Omega^{-\frac{1}{2}}\right)^T$ is due to having one less susceptible and one more active TB case, after the primary infection event at t_1. As before, these terms can reduced by replacing the conditional average of y_2 with $e^{\mathbb{A}(t_2-t_1)}\{y_1+\bar{v}\}$. However, it is worth pointing out that this wealth of terms shows how difficult it is to compute correlation functions in non-linear systems. The last equation shows that the highest order term will be proportional

to $\langle y_1 y_1 y_3 y_3 \rangle$. Computing higher order correlation functions, i.e. $g_3(t_1, t_2, t_3)$, will be very tedious. The 'Gaussian moment theorem' can be used to reduce or eliminate some of these terms. One of the corollaries of the Gaussian moment theorem states that correlations like $\langle y_1 y_1 y_3 \rangle$ are zero. It also says that terms like $\langle y_1 y_1 y_1 y_3 \rangle$ are functions of lower order correlation functions, viz $\langle y_1 y_3 \rangle$. Other than this tool, this large number of terms can only be reduced by phenomenologically motivated assumptions and approximations.

If terms of order $\Omega^{\frac{1}{2}}$ and higher are small enough to be neglected, $g_2(t_1, t_2)$ can be approximated by:

$$\begin{aligned} g_2(t_1,t_2) &= f_2(t_1,t_2) - f_1(t_1) f_1(t_2) \\ &= k^2 \Omega^{\frac{3}{2}} \mathbb{S}^2 \mathbb{I}\, P_3 \left(e^{\mathbb{A}(t_2-t_1)} - I \right) \langle y \rangle + k^2 \Omega \mathbb{S}^2\, P_{3,3} \left(e^{\mathbb{A}(t_2-t_1)} \langle y^2 \rangle - \langle y_3 \rangle \langle y_3 \rangle \right) \\ &\quad + k^2 \Omega^{\frac{3}{2}} \mathbb{S}\mathbb{I}^2\, P_1 \left(e^{\mathbb{A}(t_2-t_1)} - I \right) \langle y \rangle + k^2 \Omega \mathbb{I}^2\, P_{1,1} \left(e^{\mathbb{A}(t_2-t_1)} \langle y^2 \rangle - \langle y_1 \rangle \langle y_1 \rangle \right) \\ &\quad + k^2 \Omega \mathbb{S}\mathbb{I} P_{1,3} \left(3 e^{\mathbb{A}(t_2-t_1)} \langle y_1 y_3 \rangle_s - 4 \langle y_1 \rangle \langle y_3 \rangle \right) + k^2 \Omega \mathbb{S}\mathbb{I} \langle y_1 y_3 \rangle_s \\ &\quad + k^2 \Omega \mathbb{S}^2 \mathbb{I}\, P_3\, e^{\mathbb{A}(t_2-t_1)}\, \bar{v} + k^2 \Omega \mathbb{S}\mathbb{I}^2\, P_1\, e^{\mathbb{A}(t_2-t_1)}\, \bar{v}\,. \end{aligned} \tag{3.20}$$

The operator $P_{i,j}$ projects component i, j from a correlation matrix. I is the identity matrix. The two-point correlation function for primary infection events is shown in Fig. 3.5(blue). It seems to show that the correlation with a later second point is initially positive, with possible clustering of events. The correlation then becomes negative, still allowing clustering of events. Events then become anti-correlated (one event postpones the occurrence of the next), until there is no form of correlation.

Reinfection TB events. The two-time correlation function for reinfection events can be computed in exactly the same as above:

$$\begin{aligned} g_2(t_1,t_2) &= q^2 \Omega^{\frac{3}{2}} \mathbb{E}^2 \mathbb{I}\, P_3 \left(e^{\mathbb{A}(t_2-t_1)} - I \right) \langle y \rangle + q^2 \Omega \mathbb{E}^2\, P_{3,3} \left(e^{\mathbb{A}(t_2-t_1)} \langle y^2 \rangle - \langle y_3 \rangle \langle y_3 \rangle \right) \\ &\quad + q^2 \Omega^{\frac{3}{2}} \mathbb{E}\mathbb{I}^2\, P_1 \left(e^{\mathbb{A}(t_2-t_1)} - I \right) \langle y \rangle + q^2 \Omega \mathbb{I}^2\, P_{1,1} \left(e^{\mathbb{A}(t_2-t_1)} \langle y^2 \rangle - \langle y_1 \rangle \langle y_1 \rangle \right) \\ &\quad + 4 q^2 \Omega \mathbb{E}\mathbb{I} \left(e^{\mathbb{A}(t_2-t_1)} \langle y_1 y_3 \rangle_s - \langle y_1 \rangle \langle y_3 \rangle \right) \\ &\quad + q^2 \Omega \mathbb{E}^2 I\, P_3\, e^{\mathbb{A}(t_2-t_1)} \bar{v} + q^2 \Omega \mathbb{E}\mathbb{I}^2\, P_1\, e^{\mathbb{A}(t_2-t_1)} \bar{v}\,. \end{aligned} \tag{3.21}$$

The two-time correlation for reinfection events is shown in Fig. 3.5(green). It shows exponentially declining correlation, and hence a possibility of clustering, over a time period of 100 years.

3.4 Conclusions

Although deterministic approaches are widely used to model epidemics, the problem of uncertainty (variance) associated with deterministic models has received little attention. We used a 'population size' expansion technique to derive a Fokker-Planck equation for the fluctuations in a model of TB only. Using this FPE, we derived differential equations for the first two moments of an assumed Gaussian distribution of fluctuations around the macroscopic variables. An alternative approach to modeling fluctuations and variance in epidemic models is by means of stochastic simulation techniques. We used a Gillespie-type simulation technique to produce stochastic realizations. The statistical variance over many realizations were similar to those given by differential equations for the second moments of the Gaussian distributed fluctuations.

One interesting result of this investigation is that fluctuations are large during the phase transition of the epidemic. This is the stage when the epidemic loses sensitivity to its initial conditions (i.e. how it came about) and starts its exponential growth phase. This result suggests that when fitting the macroscopic model to data collected over the course of the epidemic, greatest uncertainty should be attached to the macroscopic description during the transition phase.

We showed how the expansion technique can be used to model the statistics of jump events in nonlinear epidemic models. We developed a simple point process theory for the occurrence of active TB episodes, comprising primary infection, re-infection and endogenous-activation events. We used the FPE to study the temporal aspects of TB clusters, and obtained an understanding of the timescale between active TB events: an age-independent model implies long correlation times. The method can be used as the basis for understanding the temporal component of TB clusters, which are currently only defined in terms of DNA type. If the modeled intensities of active TB events are validated, then the model can be used to warn against spurious conclusions from measured clustered data. For example, a correlation effect may be incorrectly attributed to an infection trend, or to a particular type of infection chain, while in reality it could be purely due to a fluctuation effect.

It is important to know if clustered active TB episodes are consistent with the dynamics of a closed community. If they are not, and therefore require exposure to external sources of infection to explain the observed clustering of TB events, it raises concerns for TB treatment programs. Treating TB cases only in a particular community will not reduce its TB burden: TB treatment programs must reach the wider community in order to be effective. One can also question whether a mass action model can produce the observed clustering of TB events. A model accounting for local contacts (household, schools) as well as global contacts (e.g. in the wider community) may be essential for modeling temporally clustered active TB events.

Chapter 4

An age-structured model of HIV and TB in a South African township

Guidelines for designing targeted TB interventions are desperately needed. The current strategy is to implement DOTS, despite the fact it has failed to reduce the burden of the TB epidemic in many of the communities where it has been tried [115]. Knowledge of the ARI can be used to evaluate the ability of DOTS to control TB epidemics in a setting with a high prevalence of HIV infection. An increases in the ARI in a dual HIV-TB setting is cause for alarm, and calls for modified DOTS intervention. On the other hand, a decreasing ARI could mean that DOTS is an effective control measure in this setting. Accurate estimates of ARI are already part of protocols used to evaluate TB control programs. The problem lies in improving the accuracy of these estimates.

Clinicians rely on indirect methods to estimate the ARI. In particular, they make use of cross-sectional prevalence surveys to estimate time trends in the ARI. The most commonly used method measures the prevalence of MTB among schoolchildren by means of TST surveys. We use a binomial-chain type model to compute the likelihood of the recent TST data set published in [80], which reports a very high prevalence of MTB among schoolchildren in Masiphumelele.

Clinical data indicate that the HIV epidemic has played a role in shifting the TB epidemic to younger age groups. In 1996-1997, the largest number of TB notifications was in the age group 40-49. By 2004 it had shifted to the 25-30-year-olds. Introducing age structure into the TB model of Sect. 2.1 adds a valuable layer of complexity, as it allows us to model age-dependent TB notification and HIV prevalence data. We investigate if age-dependent effects in the TB notification data can be reproduced by simply introducing age-dependent mixing into the HIV dynamics. This would indicate a correlation between the rise of the HIV and the TB epidemic.

We then build a next-generation matrix (NGM) and use it to explore the relative contri-

butions that HIV_- and HIV_+ active TB cases are making to the TB epidemic. Knowledge of the extent of the contribution that HIV_+ cases are making can help guide current and future TB control programs. If HIV_+ cases are making a significant contribution, then TB control programs should arguably be structured according to CD_4 count or other variables correlated to HIV progression. If HIV_+ cases are not making a significant contribution, then restructuring the DOTS program according to HIV disease staging criteria would not significantly increase its effectiveness.

4.1 Estimating the annual risk of infection

A recent article has reported a very high prevalence of infection with MTB among children aged 5 to 17 in a South African township [80]. It also showed that the annual risk of tuberculosis infection (ARI) has remained at roughly 4% for the last 15 years. The results are based on tuberculin skin tests performed in the year 2007 among 829 children aged between 5 and 17. From such age-specific prevalences, it was possible to estimate the mean annual risk of infection (ARI) experienced by each cohort (Tab. 4.1).

Age Category	Mean Age	Prevalence	ARI
5-8 yrs	7.8	$61/233 = 26.2\%$	3.8%
9-11 yrs	10.4	$70/222 = 31.5\%$	3.6%
12-13 yrs	13.0	$107/237 = 45.1\%$	4.5%
14-17 yrs	15.1	$73/139 = 52.5\%$	4.8%

Table 4.1: TB prevalence and ARI in 2007 by age quartile for the 10mm cut-off point. Reproduced from [80, Tab. 2].

Let q_x denote the proportion of children aged x with a negative TST result. If these children experienced a constant ARI a throughout their life, then q_x would be given by $q_x = (1-a)^x$, so $a = q_x^{1/x} - 1$. In [80], the value a is computed for four different age groups. All the results are close to 4%. From this it could be argued that the ARI may indeed have remained relatively constant in past 15 years.

The surprising thing is that the TB notification rate in the adult population has increased by a factor of about 5 over the same period because of the rapid rise of HIV ([69, Tab. 1]. If the sources of MTB have increased, why does the available data not show a corresponding increase in the ARI? A possible explanation is that the duration of active TB episodes for HIV_+ cases is short compared to that of HIV_- cases.

A comment on [80] pointed out that one should be careful when concluding that the ARI has remained constant [92]. The author assumed on the contrary that the ARI has been increasing each year by 10%, starting with an ARI of 1% in 1991. He claimed that this assumption gives a good fit to the data from [80, Tab. 2].

We use both deterministic and stochastic models to relate ARI and the resulting prevalence of MTB. We show that one cannot definitively conclude from the data in [80] that

the ARI has remained constant (as suggested in [80]), increased (as suggested in [92]) or decreased. However, we show that a decreasing ARI is the most likely possibility.

Stochastic model

Let $a(n)$ be the ARI in the year n. The probability for an individual aged x in 2007 to have escaped infection is:

$$q_x = (1 - a(2006)) \times (1 - a(2005)) \times ... \times (1 - a(2006 - x + 1)) \ .$$

If c_m is the size of the cohort born in year m, then the probability that i will be infected in 2007 is:

$$\left(\begin{array}{c} c_m \\ i \end{array} \right) q_x^{c_m - i} \left(1 - q_x\right)^i \ ,$$

where $x = 2007 - m + 1$. We use each cohort c_m (Tab 4.2) of the prevalence data instead of the specific cohort groupings used in [80, Tab. 2].

Fig. 4.1(a) shows the binomial distribution of the number of infected cases for a cohort aged 7 in 2007, assuming a constant ARI of 4%. Using $a(n) = a(2007) \times r^{(n-2007)}$, where r is the annual increase or decrease, we can construct a 'confidence region' for values of $a(2007)$ and r. For each choice of $a(2007)$ and r we impose the condition that for each age group the observed number of infected cases should not fall below the 2.5 percentile or above the 97.5 percentile, hence excluding 5% of each distribution.

Another approach is to give equal weight to all groups and to exclude the same amount y from each of the 13 distributions, in order to exclude 5% overall. Using $1-(1-y)^{13} = 0.05$ to find $y = 0.004$ we see that the observed number of infected cases should not fall below the 0.2 percentile or above the 99.8 percentile. The boundary of the first type of confidence region is shown by the narrow (inner) dashed contour in Fig. 4.1(b). The boundary of the second type of confidence region is shown by the wide (outer) dashed contour in Fig 4.1(b).

We can also make use of a chi-squared random variable and a standard hypothesis test to obtain a confidence region for the ARI. From the binomial distributions for 2007 we can compute the expectation value E_x for the number of children infected with MTB in each age group x. We also have observed values O_x for the number infected in each age group from the prevalence survey. The chi-squared variable is:

$$\chi^2 = \frac{(\mathrm{O}_5 - \mathrm{E}_5)^2}{\mathrm{E}_5} + \frac{(\mathrm{O}_6 - \mathrm{E}_6)^2}{\mathrm{E}_6} + ... + \frac{(\mathrm{O}_{16} - \mathrm{E}_{16})^2}{\mathrm{E}_{16}} + \frac{(\mathrm{O}_{17} - \mathrm{E}_{17})^2}{\mathrm{E}_{17}} \ .$$

We have included the children aged 5 with those aged 6 and children aged 17 with those aged 16 in the chi-squared calculation, due to small sample sizes in age groups 5 and 17. Accordingly, their observed and expected values are added to their respective groups and the

two degrees of freedom are deducted. A confidence region is constructed using the inverse of the chi-square cumulative density function (CDF) with 10 degrees of freedom at a value of 0.95. This region is depicted by a red contour in Fig. 4.1(b).

Also shown in Fig. 4.1(b) is the likelihood of observing the data set (Tab 4.2), calculated as follows:

$$L = q_5^{n_5} \times (1-q_5)^{p_5} \times ... \times q_{17}^{n_{17}} \times (1-q_{17})^{p_{17}} \ .$$

The likelihood is normalized so that the maximum value is one and the contour $L = 0.05$ is shown in Fig. 4.1(b).

The results show that although a decreasing ARI is more likely given the prevalence data, a constant or even increasing ARI cannot be ruled out. It does however rule out an ARI increasing by 10% per year starting from 1% in 1991 (and reaching 4.6% in 2007) as suggested in [92]. Based on this result, we adjusted a_1 from 0.0003/yr to 0.002/yr and k_1 from 11.4/yr to 9.0/yr and found these values to give an ARI of roughly 4%.

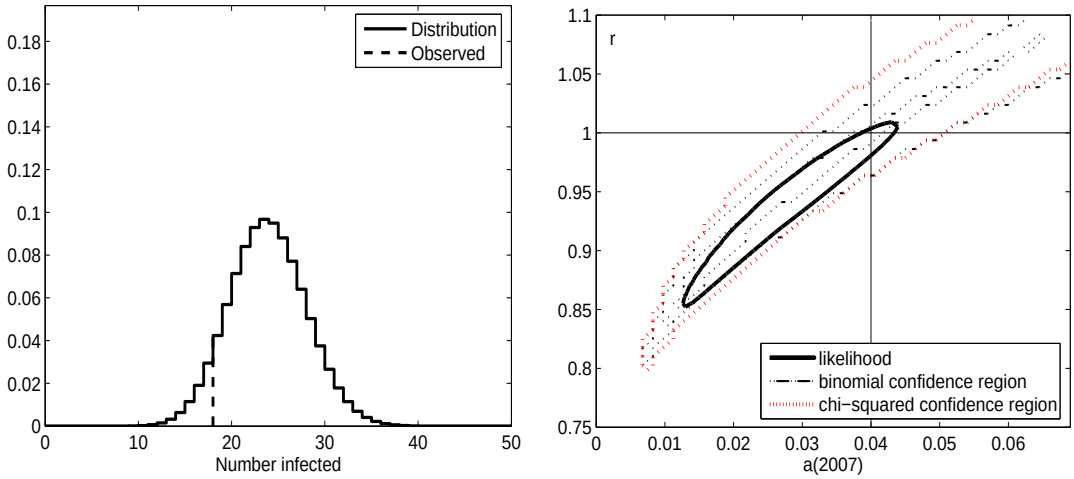
Additional prevalence surveys are needed to verify actual trends in the ARI. Estimating trends from this particular survey may be influenced by details of the histories of children participating in the survey. We have assumed that all children were equally susceptible to the sources of infection throughout their lives. However, children were eligible for the survey if they were merely resident in the town and registered at the local school at the time of the survey. Such uncertainties about the detailed cumulative exposure to the sources of infection adds to uncertainty in the analysis.

In addition, an age-dependent ARI may have been acting in the community, which would account for a relatively constant ARI in an environment of increasing exposure to MTB. Notice that the ARI is lower for young children according to Tab. 4.2. It is unfortunately not possible to disentangle time-dependent and age-dependent effects when estimating the ARI from one prevalence survey.

The analysis improves our understanding of the ARI, and serves to calibrate parameters in Ch. 2 (Tab. 2.5). Values of $k_1 = 9.0$/yr, $a_1 = 0.002$/yr and $\gamma_2 = 6.0$/yr, for example, can be justified in the context of HIV-TB epidemics in Masiphumelele and would give a ARI of roughly 4%.

age	x	5	6	7	8	9	10	11	12	13	14	15	16	17
Negative	n_x	1	27	66	78	62	47	43	65	63	45	15	5	1
Positive	p_x	0	9	18	34	30	24	16	42	65	40	19	9	5

Table 4.2: Tuberculin skin test performed in 2007



(a) binomial probability distribution (b) ARI likelihood of TST data set

Figure 4.1: (a) Probability distribution of the number of infected children in a cohort born in 2000 assuming a constant ARI of 4%. (b) A constant ARI of 4% is indicated by lines $a = 0.04\%$ and $r = 1$. We see that a decreasing ARI is more compatible with the prevalence data, but that a constant or increasing ARI also lie within the 'confidence region', and cannot be ruled out.

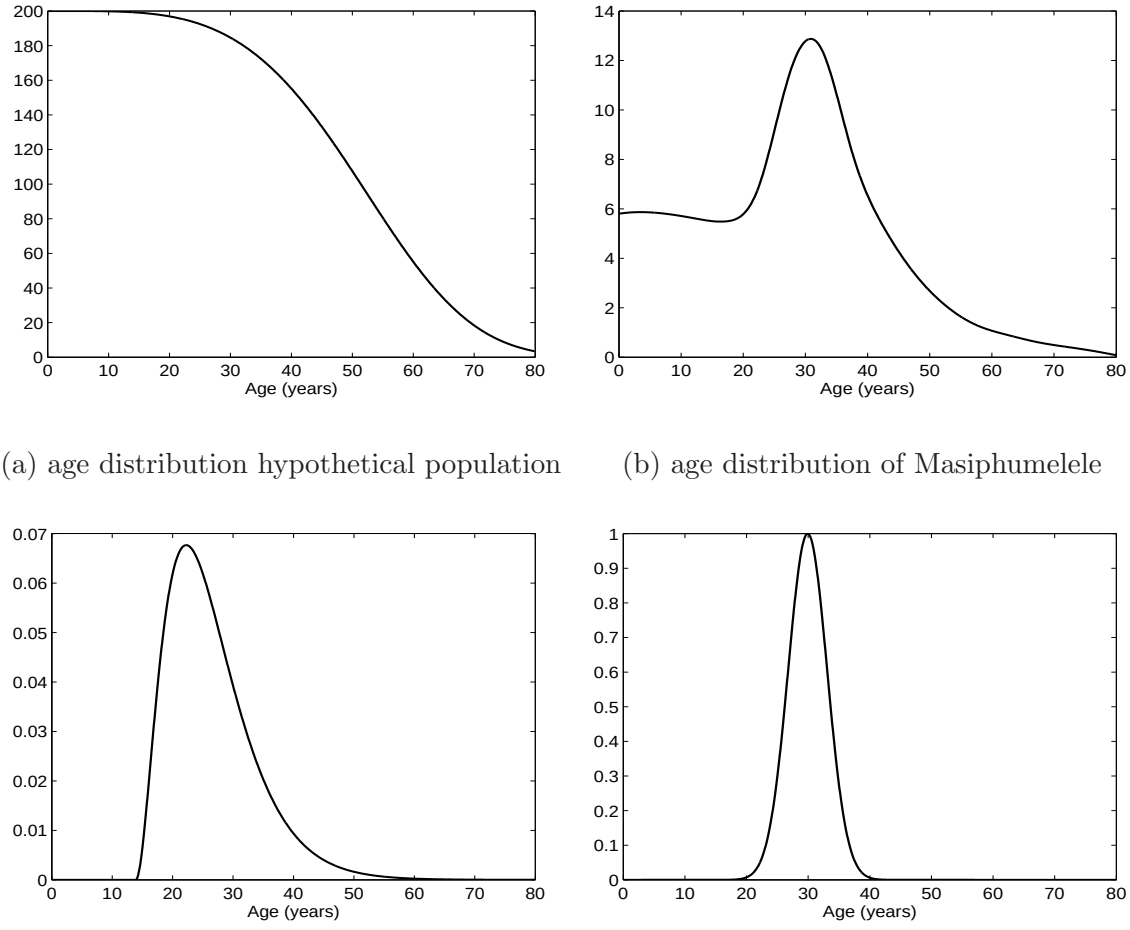
4.2 An age-structured TB and HIV model

Demography is integral to the epidemics and the tools of demography can help us understand them better (for a review see appendix B.3). The linearized epidemic (i.e. the early epidemic, when depletion of susceptibles can be ignored and when interactions between individuals are linear) can usually be studied as a linear integral equation, which is analogous to the characteristic equation of demography [15]. As a result some of the successful tools used in demography, such as the basic reproduction number and the reproduction value, can be adapted to epidemiology.

We present an age-structured model for HIV and TB interaction, adding age structure to each compartment, while keeping the model as simple as possible. Age-dependence is allowed only in the HIV submodel, while the TB dynamics (transmission and progression) are kept independent of age. The model shows that age-dependent trends in TB notifications can be explained by age-dependent effects in HIV transmission. The model is further used to derive an age-structured NGM, and to show how this matrix can help us understand the driving factors of the dual TB and HIV epidemic.

Modeling the unusual age structure of Masiphumelele

The age structure of Masiphumelele is unusual, as shown in Fig. 4.2(b). The community has a surplus of young adults compared to children and elderly people. This is thought to be the result of immigration, mostly from the Eastern Cape. Instead of modeling Masiphumele, we model the age structure of a hypothetical township of which Masiphumele is a small sample.



(a) age distribution hypothetical population

(b) age distribution of Masiphumelele

(c) risk as a function of age

(d) partner choice as a function of age

Figure 4.2: Parameter values: (a) Survivor function in a hypothetical population without HIV or TB. (b) Age pyramid of the population of Masiphumelele according to 2006 census. (c) Rate at which risky contacts are made as a function of age. (d) Distribution of partner preference for a 30-year-old.

This township is closed with respect to immigration. The model we present in this section aims to capture the population dynamics of this closed population.

The 'skewed' age distribution of Masiphumele was taken into account when computing aggregated HIV prevalence and TB notification rates. Our 'sampling' procedure can be viewed as the process by which clinics register HIV and TB cases in the community [2]. This sampling process is also required to 'maintain' the age structure of Masiphumelele, which has remained relatively unchanged between 1996 and 2004. The age distributions of the hypothetical township and the actual age distribution of Masiphumele are depicted in Fig. 4.2(a,b).

4.2.1 Model equations

Building on the model presented in Sect. 2.1, age-structure is added in a simple way to capture an age-dependent feature of HIV infection. As in Sect. 2.1, the HIV submodel is a one-sex model, and does not explicitly model how time since infection influences the spread of HIV. As in Sect. 2.1, the incidence of HIV is moderated by a response to HIV prevalence. An age-dependent risk of contracting HIV is added, and it decreases exponentially in response to HIV prevalence.

The HIV submodel is of course an over-simplification of HIV dynamics, and does not include many of the known mechanisms of HIV models, such as concurrency [39, 77, 52], commercial sex workers (CSW) [85], level of male circumcision [24], however it fits the HIV prevalence data (structured by time and age) reasonably well. Masiphumelele does not have many commercial sex workers, but various forms of commercial sex transactions make young women especially vulnerable to contracting HIV. A high percentage of men are circumcised in the community, being predominantly Xhosa–a culture where male circumcision is a custom. We avoided all of these complexities, and included only one gender in a minimally structured model for HIV dynamics.

The TB submodel is kept the same as in Sect. 2.1 and TB infection and progression rates are not modeled as a function of age, although individuals age in the model. These rates are in reality age-dependent, and the model presented shortly somewhat naïve, but the HIV-TB model presented in this section offers an explanation of how HIV and TB spreads to younger ages. The model is stated in three parts.

For those not at risk of contracting HIV:

[2] The idea of modeling an 'ordinary' population and sampling from it using the 'unusual' population of Masiphumelele was proposed by N. Bacaër, IRD, France

$$\begin{aligned}
\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right) S_0(t,a) &= -S_0(t,a)\ \left(k_1\, I_1(t)+k_2\, I_2(t)\right)/P(t) - \mu_1(t,a)\, S_0(t,a)\,,\\
\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right) E_0(t,a) &= \left(p_1'\, S_0(t,a) - q_1\, E_0(t,a)\right)\left(k_1\, I_1(t)+k_2\, I_2(t)\right)/P(t)\\
&\quad -(a_1+\mu_1(t,a))\, E_0(t,a) + b_1\, I_0(t,a)\,,\\
\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right) I_0(t,a) &= \left(p_1\, S_0(t,a) + q_1\, E_0(t,a)\right)\left(\left(k_1\, I_1(t)+k_2\, I_2(t)\right)/P(t)\right.\\
&\quad -(b_1+m_1(t,a))\, I_0(t,a) + a_1\, E_0(t,a)\,.
\end{aligned}$$

For those at risk of contracting HIV:

$$\begin{aligned}
\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right) S_1(t,a) &= -S_1(t,a)\ \left(k_1\, I_1(t)+k_2\, I_2(t)\right)/P(t) - \mu_1(t,a)\, S_1(t,a)\\
&\quad -\phi(t,a) S_1(t,a)\,,\\
\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right) E_1(t,a) &= \left(p_1'\, S_1(t,a) - q_1\, E_1(t,a)\right)\left(k_1\, I_1(t)+k_2\, I_2(t)\right)/P(t)\,,\\
&\quad -(a_1+\mu_1(t,a))\, E_1(t,a) + b_1\, I_1(t,a) - \phi(t,a) E_1(t,a)\\
\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right) I_1(t,a) &= \left(p_1\, S_1(t,a) + q_1\, E_1(t,a)\right)\left(k_1\, I_1(t)+k_2\, I_2(t)\right)/P(t)\\
&\quad -(b_1+m_1(t,a))\, I_1(t) + a_1\, E_1(t,a) - \phi(t,a) I_1(t,a)\ .
\end{aligned}$$

For those who are HIV+:

$$\begin{aligned}
\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right) S_2(t,a) &= S_2(t,a)\left(k_1\, I_1(t)+k_2\, I_2(t)\right)/P(t) - \mu_2(t,a)\, S_2(t,a)\\
&\quad +\phi(t,a) S_1(t,a)\,,\\
\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right) E_2(t,a) &= \left(p_2'\, S_2(t,a) - q_2\, E_2(t,a)\right)\left(k_1\, I_1(t)+k_2\, I_2(t)\right)/P(t)\\
&\quad -(a_2+\mu_2(t,a))\, E_2(t,a) + b_2\, I_2(t,a) + \phi(t,a) E_1(t,a)\,,\\
\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right) I_2(t,a) &= \left(p_2\, S_2(t,a) + q_2\, E_2(t,a)\right)\left(k_1\, I_1(t)+k_2\, I_2(t)\right)/P(t)\\
&\quad -(b_2+m_2(t,a))\, I_2(t,a) + a_2\, E_2(t,a) + \phi(t,a) I_1(t,a)\ .
\end{aligned}$$

The boundary conditions are:

$$\begin{aligned}
S_0(t,0) = pB,\ E_0(t,0) = 0,\ I_0(t,0) = 0,\\
S_1(t,0) = p'B,\ E_1(t,0) = 0,\ I_1(t,0) = 0,\\
S_2(t,0) = 0,\ E_2(t,0) = 0,\ I_2(t,0) = 0\,,
\end{aligned}$$

where $p = 1 - p'$ is the probability, assigned at birth and kept for life, that an individual

is not at risk of getting HIV. In this simple model, TB infection and TB dynamics are not age-dependent. In particular, the effective TB rate is calculated using total numbers of TB cases divided by the total population (i.e. standard incidence). To this end, we define:

$$\begin{aligned} I_1(t) &= \int_0^\infty \big(I_0(t,a) + I_1(t,a)\big)\, da\,, \\ I_2(t) &= \int_0^\infty I_2(t,a)\, da\,, \\ P(t) &= \sum_{i=0}^{2} \int_0^\infty \big(S_i(t,a) + E_i(t,a) + I_i(t,a)\big)\, da\,. \end{aligned}$$

The risk of HIV infection $\phi(t,a)$ is assumed to be age-dependent. The age-dependent force of infection for HIV is based on a proportional mixing argument. It is stated in Eq. (4.1) and is elaborated upon in Sect. 6.1.1, in a discussion about HIV dynamics.

4.2.2 Parameter values

Mortality. We assume that the survival curve (without HIV and TB) is given by $S(a) = \exp(-\phi a^v)$ with $\phi = 10.7$ and $v = 4$ (a in years). The natural mortality is then defined by $m(a) = -S'(a)/S(a)$. Mortality with HIV is given by $\mu_2(a) = \mu_1(a) + \frac{1}{10}$, capturing a life expectancy of 10 years with HIV.

TB-related mortality is assumed to be independent of age, as in Tab. 2.5. The age-distribution in our hypothetical population is shown in Fig. 4.2(a), while the age-distribution of Masiphumelele, according to a 2006 census in the community, is shown in Fig. 4.2(b)

Risk of contracting HIV. The force of infection for contracting HIV is formulated in terms of an age-dependent risk of contracting HIV and an age-dependent preference for partners:

$$\phi(t,a) = F(H)\beta(a)\frac{\int_0^\varpi \beta(b) f(a,b)\big[S_2(t,b)\big]\, db}{\int_0^\varpi \beta(b,\tau) f(a,b)\big[S_1(t,b) + S_2(t,b)\big]\, db}\,, \tag{4.1}$$

where $\beta(a)$ is a measure of the risk of contracting HIV at age a and $b \mapsto f(a,b)$ is the age-distribution of partners of a person aged a. The function $F(H) = e^{-\lambda H}$, with H representing the prevalence of HIV and λ behavior change, reduces the transmission rate in response to HIV. Fig. 4.2(c) depicts the following risk function:

$$\beta(a) = \begin{cases} 0 & \text{if } a < 15 \\ 5 \times 10^{-5} \times 12^2 \times (a-15)^2 e^{-0.24(a-15)} & \text{if } a \geq 15 \end{cases}\,. \tag{4.2}$$

For $f(a, b)$ we use a normal distribution (see Fig. 4.2(d)). Although in reality men and woman have different preferences, this risk function is sufficient for our purposes.

4.3 Simulation results

It is easy to be overconfident about epidemic models which fit available data. This confidence can be unwarranted in models with many parameters, because good fit to data could well be the result of the choice of structure and parameters in the model, and not of the correctness of the model as a whole [20, Ch 6.2]. The following is a short outline of the procedure we followed to fit the model to available data, as displayed in Fig. 4.3, in an attempt to apply method to what is, by its very nature, an imprecise science.

The simulation (in the form of a finite difference scheme solved by Euler's method) is run until a demographical steady state is reached. At this point an HIV_+ individual is introduced, and the proportion of people at risk of contracting HIV adjusted (using $p = 0.4$ and the age-dependent parameters specified in Sect. 4.2.2, and $\lambda = 3$) until a reasonable fit to the aggregated HIV prevalence curve displayed Fig. 4.3(b) is found. The parameter p influence the overall scale of the HIV epidemic, while λ influences the rate at which the prevalence curve levels off. It is worth pointing out that the steepness of the HIV prevalence curve to the age of the index HIV case. It is also sensitive to the birth rate which have been set to constant $B = 200$ newborns per year. Our motivation for this choice is that we do a steady-state analysis. Adjusting the age-dependent risk function $\beta(a)$ influences the steepness of the initial HIV epidemic. It directly affects (by design) the spread of HIV from older to younger age groups, and indirectly has the same effect on TB notification as a function of age.

Having found the HIV-related parameters, a latent TB case is introduced into the compartment not at risk of HIV infection (E_0) and the model is run from demographical steady state until an endemic TB steady state is reached. A 16-year-old HIV_+ TB susceptible is then introduced. The age-independent TB parameters specified in Tab. 2.5 give the simulated epidemic curves shown in Fig. 4.3(a).

Fig. 4.3(c and d) is an attempt to fit the TB notification and HIV prevalence data when they are disaggregated by age group and by year, as published in [69]. Fig. 4.3(c) shows that in 1996-1997 the maximum number of TB notifications was in the age group 40-49. Fig. 4.3(d) shows the estimated age structure of people living with HIV [123]. Since the HIV epidemic affects young people and since HIV_+ people progress more rapidly from latent to active TB, the maximum number of TB notifications has shifted to younger age groups.

In Fig. 4.3(c) there is a notable misfit between the modeled and the measured TB notification data, particularly within younger age groups. The fit can be improved, although not to complete satisfaction, by adjusting the age-dependent risk of contracting HIV $\beta(a)$. However, children may well experience different levels of MTB infection risk than adults. Visual inspection of Fig. 4.3(c) also suggests that the risk of contracting TB could be age-

dependent for HIV_+ cases, but at this point we do not have to investigate this possibility.

The mass-action assumption used in the TB submodel is not *a priori* true. An obvious next step would be to evaluate if different households are at different risk of infection. One can also attempt to model two levels of mixing, one for local contacts (household, school, workplace, and so on) and a lower level of risk due to mixing in the general community (a theoretical framework for this type of mixing is developed by Ball, et al. [19]). We note that only limited data on different TB strains, which in principle could provide an empirical basis for mixing assumptions, are available for Masiphumelele.

With regards to TB mixing patterns, it is worth pointing out that a clustering analysis of different TB strains, circulating in a community near Cape Town with a similar TB burden, shows no association between age and 'risk of being in a particular cluster' [108]. However, our focus in the remainder of the chapter is to study the influence of HIV_+ individuals in spreading TB in the community rather than how they are contracting MTB as individuals.

4.4 The next-generation matrix for an age-structured HIV-TB model

In this section we use the conceptual framework for building a NGM for an age-structured TB model, adapting a general framework discussed in appendix B.1. The formulation will be kept simple at first, and elaborated upon later, once the main ideas are illustrated. The NGM formalism is thoroughly treated in Sect. 5.1, and applied to more sophisticated age-structured models. Consider the following definition for the next-generation operator:

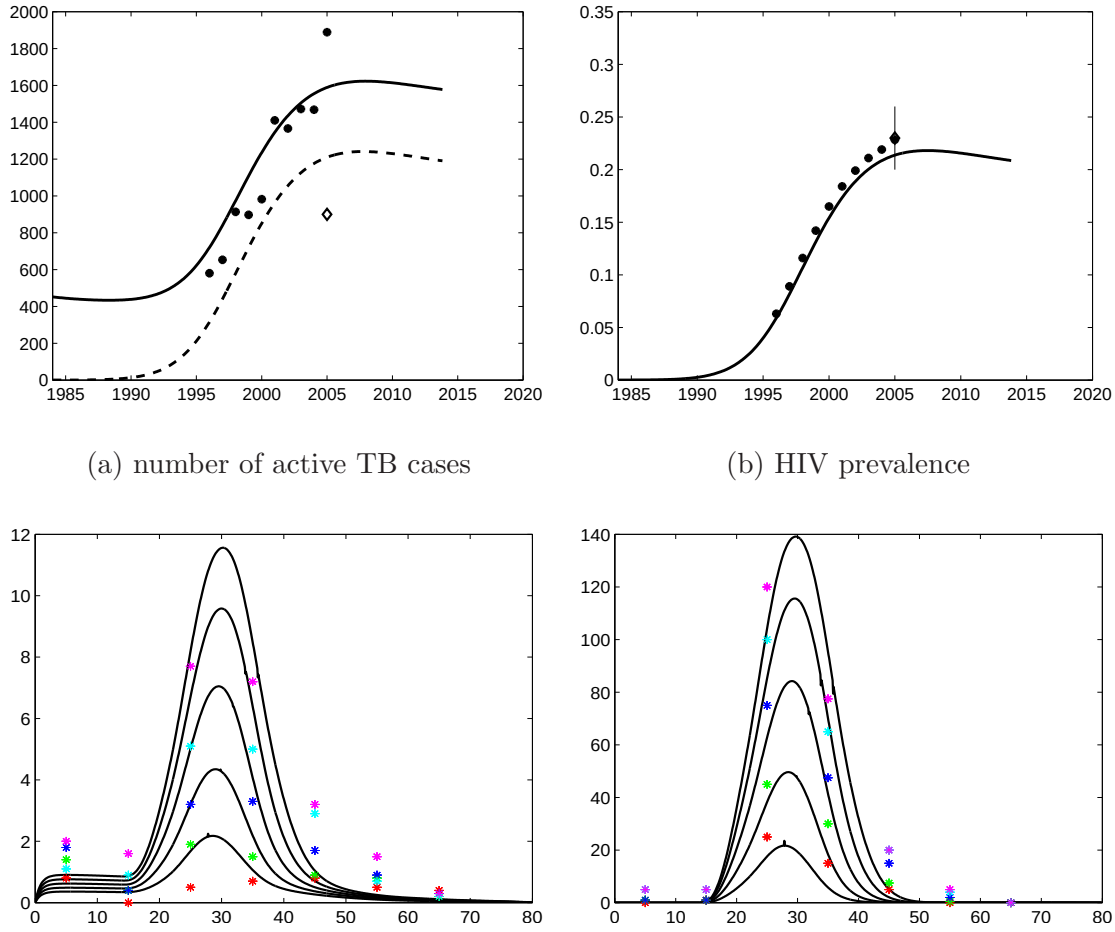
$$(\mathbb{K}\varphi)(a) = \int_{D(\eta)} K(a,\alpha)\varphi(\alpha)\, d\alpha\,,$$

where $\varphi(a)$ represents an infective distributed over some state. In this model, infectives are distributed according to age, duration of infection, being latently or actively infected and being HIV_- or HIV_+. We use the simplest formulation of the kernel K, given by the following general form:

$$K(a,\alpha) = \hat{S}(a) \int_0^\infty h(\tau,\alpha)\, c(a,\alpha+\tau)\, \frac{F_d(\alpha+\tau)}{F_d(\alpha)}\, d\tau\,, \tag{4.3}$$

with the following interpretations for the terms in the integrand:

- $\frac{F_d(\alpha+\tau)}{F_d(\alpha)}$ is the probability of escaping death τ time units after the start of an active TB episode at age α. This survival probability includes a factor $e^{-\int_\alpha^{\alpha+\tau}\mu(a)\,da}$ for natural mortality and a factor $e^{-m_i\tau}$ $(i=1,2)$ for surviving active TB while HIV_- or HIV_+. We make the simplifying assumption that disease-related mortality is independent of time since infection.



(a) number of active TB cases

(b) HIV prevalence

(c) disaggregated TB notification rate

(d) disaggregated HIV prevalence

Figure 4.3: TB and HIV burden becoming more peaked at younger age groups. The data in (a) and (b) are stratified according to age and year: 1996–1997 in red, 1998–1999 in green, 2002-2003 in blue, 2002–2003 in cyan and 2004 in magenta. (a) Aggregated number of notified TB cases per 100000. (b) Aggregated number of HIV_+ individuals. (c) Disaggregated number of notified TB cases per 100000. (d) Disaggregated number of HIV_+ individuals. Data adapted from [69].

- $c(a,\alpha)$ is an age-dependent contact rate for MTB infection, while $h(\tau,\alpha)$ is the probability of transmission, which in principle depends on the duration of (i.e. time since) infection. In our model age does not directly influence the effective contact rate, but influences the probability of becoming HIV_+. The two contact rates are k_1, k_2 for HIV_- and HIV_+ individuals respectively, as in Sect 2.1.

- $\hat{S}(a)$ is the number of individuals, aged a, who are susceptible to TB, when the population is at steady state with endemic HIV.

Description of infectious episodes. To study this sequence of TB episodes, let us first consider the category of HIV_- and smear positive $I_1(t,a)$ or smear negative $E_1(t,a)$. The recovery and activation events are modeled by:

$$\begin{aligned}\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right) I_1(t,a) &= -b_1\, I_1(t,a) + a_1\, E_1(t,a) - m_1\, I_1(t,a)\,,\\ \left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right) E_1(t,a) &= b_1\, I_1(t,a) - a_1\, E_1(t,a) - \mu_1(t,a)\, E_1(t,a)\,,\end{aligned}$$

for HIV_- individuals, and a similar equation holds for HIV_+ individuals. In this model, TB progression is characterized by sequential episodes. A newly introduced infective undergoes a series of recovery and reactivation events. The initial reinfection rate is assumed to be low enough to be ignored. The probabilities are (conditional on survival) $e^{b_i\,\tau}(i=1,2)$ for HIV_- and HIV_+ active cases and $e^{a_i\,\tau}(i=1,2)$ for HIV negative and positive latently infected cases respectively.

Each individual undergoes, given that he survives, transitions according to the matrix of transition rates:

$$G=\begin{pmatrix}-b_1 & a_1\\ b_1 & -a_1\end{pmatrix},$$

and survives according to the survival matrix:

$$F(\tau)=\begin{pmatrix}m_1 & 0\\ 0 & \mu_1(\tau)\end{pmatrix}.$$

Collecting these ideas we see that the number of infections from the state (I_1,E_1) toward susceptibles aged a, of which there are $S(a)$ near the (TB) disease-free steady state, is:

$$k(a,\alpha)=\frac{k_1\,S(a)}{\bar{P}}\begin{pmatrix}p_1 & p_1'\\ p_1 & p_1'\end{pmatrix}\int_0^\infty e^{G\,\tau}\, e^{-\int_\alpha^{\alpha+\tau} F(s)\,ds} e^{-\int_\alpha^{\alpha+\tau}\phi_{hiv}(t_0+\tau,s)\,ds}\,d\tau\,, \tag{4.4}$$

where $\bar{P}$ is the size of the total population in HIV steady state and $\phi_{hiv}(t,a)$ is the force of HIV infection acting on an individual aged a at time t. The probability for an individual aged α at t_0 (the year HIV is introduced into the community) to escape infection until at least time $(t_0+\tau)$ is $e^{-\int_\alpha^{\alpha+\tau}\phi_{hiv}(t_0+\tau,s)\,ds}$.

A general model for TB infection. Generalizing the NGM of Sect. 2.2 to model infection of between all types, we see that:

$$K(a,\alpha)=\int_0^\infty \mathcal{A}(a,\theta,\tau)\,\mathcal{P}(\theta,\alpha,\tau)\,d\tau\,. \tag{4.5}$$

$\mathcal{A}(a,\theta,\tau)$ models infectiousness towards susceptibles aged a of an infective in state θ, τ years into the infectious period. $\mathcal{P}(\theta,\alpha,\tau)$ is the conditional probability that the infective is in state θ at time τ, given that he was in state α when his infectious period started. The infective is aged α when his infectious period started (as an $\text{HIV}_-(I_1)$ or $\text{HIV}_+(I_2)$) and he was introduced into a population in a disease-free steady state. Other than his age, his state is specified by whether he is latently or actively infective, and HIV_- or HIV_+. $\mathcal{P}(\theta,\alpha,\tau)$ is the solution of the following matrix equation:

$$\frac{d}{d\tau}\mathcal{P}(\theta,\alpha,\tau)=\mathcal{B}(\theta,\alpha,\tau)\,\mathcal{P}(\theta,\alpha,\tau),\ \mathcal{P}(\theta,\alpha,0)=I\,. \tag{4.6}$$

$\mathcal{B}(\theta,\alpha,\tau)$ is a 4×4 time-dependent transition matrix acting on a state probability with an initial value $\mathcal{P}(0)=I$. For this model, $\mathcal{B}$ is given by:

$$\mathcal{B}(\tau)=\begin{pmatrix} -(a_1+\mu_1(b)+p'\phi_{hiv}(b)) & b_1 & 0 & 0\\ a_1 & -(b_1+m_1+p'\phi_{hiv}(b)) & 0 & 0\\ p'\,\phi_{hiv}(b) & 0 & -(a_2+\mu_2(b)) & b_2\\ 0 & p'\,\phi_{hiv}(b) & a_2 & -(b_2+m_2)\end{pmatrix},$$

where $b=\alpha+\tau$, $p'\phi_{hiv}(\alpha+\tau)$ is a time-dependent risk of getting HIV and $\mu_1(\alpha+\tau)$ is an age-dependent mortality, τ years after being infected with MTB at age α. Note that transitions from one state to another are captured by off-diagonal elements of the matrix $\mathcal{B}(i,j,\tau)$. $\mathcal{A}(\tau)$ is a matrix of contact rates and susceptible numbers in the relevant categories:

$$\mathcal{A}(\tau)=\frac{1}{\bar{P}}\begin{pmatrix} 0 & p_1'k_1S_1(a) & 0 & p_1'k_2\,S_1(a)\\ 0 & p_1k_1S_1(a) & 0 & p_1k_2\,S_1(a)\\ 0 & p_2'\,k_1S_2(a) & 0 & p_2'k_2\,S_2(a)\\ 0 & p_2\,k_1S_2(a) & 0 & p_2\,k_2\,S_2(a)\end{pmatrix}.$$

This matrix was previously assumed to be independent of time since infection (Sect. 2.2), and was computed at a disease-free equilibrium value, in order to study the invasion of a situation with endemic HIV by TB.

Numerical investigation into the properties of the NGM

Due to the time-dependence of $\mathcal{B}(t)$ we cannot write:

$$\mathcal{P}(t) = e^{\int_0^t \mathcal{B}(s)ds} \cdot P(0)\,,$$

in analogy with the formula for a time-independent $\mathcal{B}$:

$$\mathcal{P}(t) = e^{\mathcal{B}\,t} \cdot \mathcal{P}(0)\,.$$

An analytical solution for $\mathcal{P}(t)$ can be obtained by means of a time ordered series expansion in the time-dependent operator $\mathcal{B}(t)$ (e.g. a Dyson series expansion [93, ch. 4.2]). We can approximate a numerical solution for $\mathcal{P}(\theta, \alpha, \tau)$ in Eq. (4.6) by means of the discretized transition matrix $N(\tau_n)$, which remains constant over each interval $\Delta\tau$ of a discretized time since infection:

$$\mathcal{P}_k(\tau_n) = e^{\mathcal{B}_k(\tau_n)\Delta\tau}\, e^{\mathcal{B}_k(\tau_{n-1})\Delta\tau}\, \cdots\, e^{\mathcal{B}_k(\tau_1)\Delta\tau}\, \mathcal{P}_k(0)\,,$$

where, dropping reference to θ and α, $\mathcal{P}(0)$ is the 4×4 identity matrix. The NGM is found by multiplying this by $\mathcal{A}(a, \theta)$, in the case where $\mathcal{A}(a, \theta)$ is independent of time since infection. In the case where $\mathcal{A}(a, \theta, \tau)$ does depend on time since infection τ, the kernel K is given by:

$$\begin{aligned} K(a, \alpha) \;=\; & M(\tau_n)\, e^{\mathcal{B}(\tau_n)\Delta\tau}\, \mathcal{P}(\tau_{n-1}) + \mathcal{A}(\tau_{n-1})\, e^{\mathcal{B}(\tau_{n-1})\Delta\tau}\, \mathcal{P}(\tau_{n-2}) + \cdots \\ & + \mathcal{A}(\tau_1)\, e^{\mathcal{B}(\tau_1)\Delta\tau}\, \mathcal{P}(0)\,. \end{aligned}$$

A more direct way to find $\mathcal{P}(\theta, \alpha, \tau)$ is to solve the ODE (4.6) directly. However, the approach of a product of year-on-year exponential matrices fits conveniently within our finite difference solution, which is updated at yearly intervals.

Distribution of infection period over infectious states. Taking a closer look at Eq. (4.6) one sees that it can model the time an individual spends from when he is introduced into a state. For example, the time spent in state i at time t by an individual who was introduced into state j at time $t = 0$, is given by the *i'th* component of the vector $T = \int_0^t \mathbb{P}(s)ds \cdot \mathbb{P}(0)$, where the *j'th* component of $\mathbb{P}(0)$ is 1 and the other components are zero.

The distribution of states over time since infection for a 30- and 50-year-old infective, is shown in Fig. 4.4(a) and (b) respectively. From these figures we see that 30-year-olds spend

more time in state E_2 than in E_1. At 50 years the role is reversed, because of the reduced risk of becoming HIV$_+$ at age 50. It can be seen that the average time spent in an active TB episode (the dotted blue line) is much less than one year.

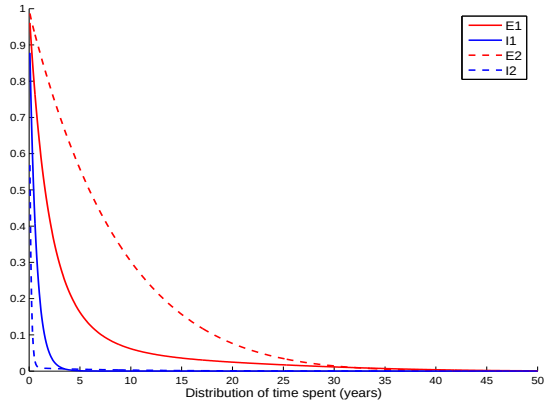
The structure and spectral properties of the NGM. The structure of the NGM is shown in Fig. 4.4(c). The reproductive value, i.e. the left eigenvector corresponding to the largest eigenvalue, is shown in Fig. 4.4(d). Note that HIV$_-$ individuals with active TB make the greatest contribution to the TB epidemic. Their relative contribution decrease between the ages 25 and 50, when they are at risk of contracting HIV. The second largest contribution comes from HIV$_+$ individuals on ART with latent and active TB.

4.5 Conclusions

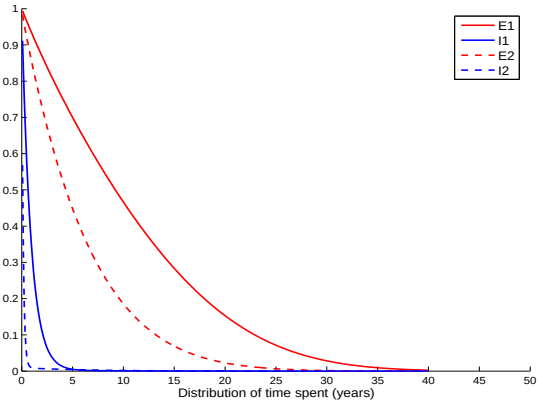
We developed a stochastic binomial model for MTB infection among schoolchildren in order to estimate the annual risk of MTB infection in Masiphumelele. Using MTB prevalence data from a recent TST survey among 5- to 17-year-old children in the community, we computed binomial distributions for the number of infected children in each age group. We constructed a confidence region for the value of the ARI in 2007, and its annual increase or decrease over the preceding years. The likelihood of the observed data set was computed as a function of these two parameters and its shape was compared with that of the confidence region. Our analysis of the TST data set suggests that the ARI has most likely been decreasing over the last 15 years. This implies that the DOTS program is in fact limiting transmission of TB to HIV$_-$ cases, despite a sharp rise in notified TB cases.

To investigate the role that the escalating number of TB cases who are HIV$_+$ play in spreading TB in this community, we built an age-structured model for the dual HIV and TB epidemics. We were able to reproduce the age-dependent trends of TB notifications and HIV prevalence in Masiphumelele, without making any of the TB parameters age-dependent. Our model suggests that trends in TB notification follow trends in the spread of HIV. In particular, both epidemics peak at younger ages.

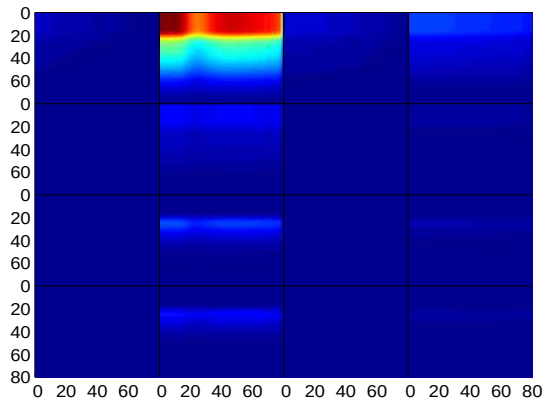
We developed a next-generation operator framework in order to understand the relative influence of different TB types on the spread of the TB epidemic. The left eigenvalue of this matrix, associated with the largest eigenvalue R_0, is a measure of the relative influence different infective types are making to the TB epidemic. We showed that the TB epidemic is driven by HIV$_-$ active TB cases. HIV$_+$ cases spend too little time in the community to make a significant contribution to the spread of TB. The next-generation operator formalism is developed to study HIV and TB epidemics in South Africa as a whole. We focus on the impact of a proposed universal test and treat strategy for HIV, which could potentially reconstitute immune system health in HIV-TB co-infected individuals. This effect has the potential of enabling DOTS to reduce the incidence of new HIV$_+$ TB cases.



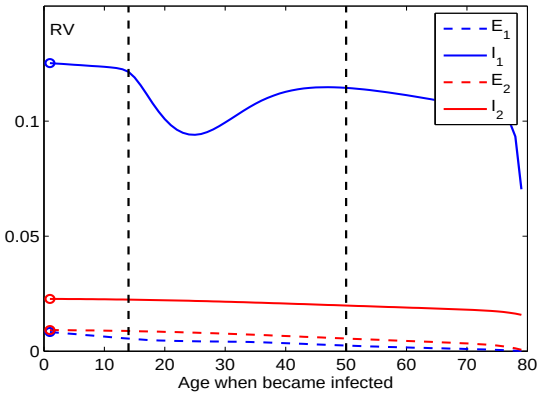
(a) infectious history of 30-year-old



infectious history of 50-year-old



(c) NGM of TB infection



(d) reproduction value of TB infection

Figure 4.4: (a) Infectious history of an infective who became active at age 30. (b) Infectious history of an infective who became active at age 50. (c) The NGM. (d) The reproduction value as the left eigenvector of the NGM corresponding to the largest eigenvalue. For these parameter values the largest eigenvalue is $R_0 = 1.5$. It seems to show that active HIV$_-$ individuals are making the greatest contribution to a TB epidemic in a community with high HIV prevalence. The contribution is reduced during the ages of 15 to 50 years when they have a high risk of becoming HIV$_+$, reducing their contribution to the spread of TB.

Chapter 5

An age-structured model of HIV and TB in South Africa

In South Africa many HIV_+ people are never tested for HIV before they die of AIDS-related symptoms. Many others are tested only at a late stage of HIV infection, when symptoms are already present. Those that do have access to ART after an HIV test often already have a very low CD4 cell count. Although ART saves lives and decreases viral load to very low levels that are much less infectious, it is typically implemented too late in the infection process to prevent the spread of HIV.

Recent modeling work [47] suggested that a very active program of HIV testing, with all detected HIV_+ people immediately receiving ART, could be an efficient way not just to control the HIV epidemic but even to eradicate it within two decades. The model used in [47] was relatively simple in structure and the assumptions somewhat optimistic. Nevertheless, the modeling results were striking enough to make experts consider ART not only as a treatment for HIV but also as a possible tool for preventing the spread of the disease, like condoms or circumcision.

There are now more than 6 million HIV_+ people in South Africa alone, and more than 300,000 AIDS-related deaths occur each year in the country despite the current ART program. ART costs at least 300 US dollars a year for each patient. Even if a universal test and treat strategy (UTTS) could save several millions of lives, it would cost more than one billion US dollars per year during its first decade after deployment. Funders need to be convinced of the feasibility of such a costly project. Modeling can help produce a clearer picture of the situation and how it could evolve.

In this chapter, building on our earlier study of a very simple HIV-TB model [17], we investigate the advantages of the proposed strategy for the control of tuberculosis (TB), which is the leading cause of mortality among HIV_+ individuals in South Africa. Whereas until now in this project we have focused on data from a single township with exceptionally high TB notification rates, we will now consider an imaginary 'average' community in which TB and HIV statistics would be typical of South Africa at the national level. Modeling a

hypothetical community is both deliberate and necessary: it is not feasible to model all 48 million South Africans. However, we have based this community on real data obtained from antenatal clinics throughout South Africa.

The HIV submodel offers several improvements to the model of Sect. 2.1 by including age, a variable for time since HIV infection, separate compartments for people receiving ART, and various parameters taking into account the history of interventions against HIV. (For a review of HIV-TB models, we refer to [17], to which one should add the more recent article [21].) Of course, the question of which ingredients should go into the model is controversial. There is no general agreement, for example, as to why the HIV epidemic spread so fast in South Africa when compared to Central or Western Africa. Some think that the large migrant male population working in mines and the associated female sex workers were a key factor [51, 61]. Some put forward the low level of male circumcision [122]. Others emphasize the fact that age at first marriage is relatively high in South Africa [22]. We have avoided these difficulties in our HIV submodel, since we can already obtain, with relatively few ingredients, a reasonable fit to age-specific HIV data from South Africa.

5.1 The next-generation matrix for structured epidemic models

The following is a generalization of the next-generation formalism, discussed in appendix B.2, for epidemic models which account for demographical detail. This formulation will be used to study age-structured HIV and TB epidemics. X denotes infectives, which can be structured according to age and gender. $\mathcal{A}$ is an infection matrix, stating how many cases of each susceptibility type are infected when it is assumed that the epidemic is linear, i.e. it models the 'early' epidemic when susceptibles are not yet appreciably depleted. This assumption is particularly important when modeling time since infection. It achieves an analogue of the homogenous mixing assumption as it implies that the infectious period does not influence the probability of making a potentially infectious contact with a susceptible [20, Ch. 8.1]. $\mathcal{B}$ is a matrix of transitions an infective may undergo during an infectious period.

Consider a linearized age- and infection-age structured epidemic model of the form

$$\frac{\partial X}{\partial t} + \frac{\partial X}{\partial x} + \frac{\partial X}{\partial \tau} = -\mathcal{B}(x,\tau)\, X(t,x,\tau) \tag{5.1}$$

$$X(t,0,\tau) = 0\ , \tag{5.2}$$

$$X(t,x,0) = \int_0^\infty \int_0^y \mathcal{A}(x,y,\tau)\, X(t,y,\tau)\, d\tau\, dy\ . \tag{5.3}$$

We assume that there is no mother-to-child transmission. Here, $X(t,x,\tau)$ is a vector whose components are the "infected" populations at time t with age x and infection-age τ. The square matrix $\mathcal{A}(x,y,\tau)$ is nonnegative. The square matrix $\mathcal{B}(x,\tau)$ has positive diagonal elements and non-positive off-diagonal elements. Let $\mathcal{C}(x,\tau)$ be the transition matrix from

age x to age $x+\tau$, defined by

$$\frac{\partial \mathcal{C}}{\partial \tau}(x,\tau) = -\mathcal{B}(x+\tau,\tau)\,\mathcal{C}(x,\tau), \quad \mathcal{C}(x,0) = I,$$

where I is the identity matrix. Then one can show that the "incidence" $Y(t,x) = X(t,x,0)$ satisfies the renewal equation

$$Y(t,x) = \int_0^\infty \int_0^\infty \mathcal{A}(x,y+\tau,\tau)\,\mathcal{C}(y,\tau)\,Y(t-\tau,y)\,d\tau\,dy\,. \tag{5.4}$$

Indeed, set $F(t,x,\tau) = X(t+\tau,x+\tau,\tau)$. Then

$$\frac{\partial F}{\partial u}(t,x,\tau) = \Big(\frac{\partial X}{\partial t} + \frac{\partial X}{\partial x} + \frac{\partial X}{\partial \tau}\Big)(t+\tau,x+\tau,\tau) = -\mathcal{B}(x+\tau,\tau)\,F(t,x,\tau).$$

So $F(t,x,\tau) = \mathcal{C}(x,\tau)F(t,x,0) = \mathcal{C}(x,\tau)Y(t,x)$. Therefore, $X(t,y,\tau) = \mathcal{C}(y-\tau,\tau)Y(t-\tau,y-\tau)$. Eq. 5.3 then gives

$$\begin{aligned} Y(t,x) &= \int_0^\infty \int_0^y \mathcal{A}(x,y,\tau)\,\mathcal{C}(y-\tau,\tau)\,Y(t-\tau,y-\tau)\,d\tau\,dy \\ &= \int_0^\infty \int_\tau^\infty \mathcal{A}(x,y,\tau)\,\mathcal{C}(y-\tau,\tau)\,Y(t-\tau,y-\tau)\,dy\,d\tau\,, \end{aligned}$$

which is the same as (5.4). Looking for a solution of the form $Y(t,x) = e^{rt}Z(x)$ of (5.4), we see that r is the unique real number such that equation

$$Z(x) = \int_0^\infty \Big[\int_0^\infty \mathcal{A}(x,y+\tau,\tau)\,\mathcal{C}(y,\tau)\,e^{-r\tau}\,d\tau\Big] Z(y)\,dy \tag{5.5}$$

has a positive solution $Z(x)$. Finally, the basic reproduction number R_0 is the spectral radius associated with the eigenvalue problem

$$R_0\,\Omega(x) = \int_0^\infty K(x,z)\,\Omega(z)\,dz\,, \tag{5.6}$$

where

$$K(x,y) = \int_0^\infty \mathcal{A}(x,y+\tau,\tau)\,\mathcal{C}(y,\tau)\,d\tau \tag{5.7}$$

is the expected number of "secondary cases" aged x produced by one "primary case" that got infected at age y.

5.2 Model equations

5.2.1 HIV submodel

We separate women ($k=1$) from men ($k=2$). Let $S_k(t,x)$ be the density of HIV_- individuals at time t, aged x, and of gender k, in the sense that $\int_{y_1}^{y_2} S_k(t,y)\,dy$ is the number of S_k individuals in the age interval (y_1,y_2). Let $S_k^*(t,x,\tau)$ be the density of HIV_+

individuals at time t, aged x, of gender k, that have been infected for τ units of time, and that do not receive ART. Finally let $S_k^\diamond(t,x)$ be the density of HIV_+ individuals at time t, aged x, of gender k, that receive ART. This structure is sketched in Fig. 5.1.

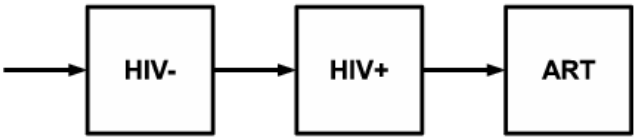


Figure 5.1: HIV structure.

Let $\mu_k(x)$ (resp. $\mu_k^*(x,\tau)$, $\mu_k^\diamond(x)$) be the gender- and age-specific mortality for HIV_- (resp. HIV_+, ART) individuals. Notice that mortality for HIV_+ individuals also depends on the time τ since infection. Let $\phi_k(t,x)$ be the force of infection, i.e., the rate at which HIV_- individuals become HIV_+. Let $g^\diamond(t,x,\tau)$ be the rate for starting ART, which depends on the time τ since infection and can change with time t as access to ART improves. Let $\varphi(t)$ be the mother-to-child transmission probability of HIV if the mother is HIV_+ but did not receive ART before pregnancy. Let $M(t)$ be the proportion of such women among all pregnant women. We assume that women who received ART before pregnancy do not transmit HIV to their children. We neglect HIV transmission as a result of breastfeeding. Let B be the birth rate, which is assumed to be constant. With these notations, our model is:

$$\frac{\partial S_k}{\partial t} + \frac{\partial S_k}{\partial x} = -\left[\phi_k(t,x) + \mu_k(x)\right] S_k, \tag{5.8}$$

$$\frac{\partial S_k^*}{\partial t} + \frac{\partial S_k^*}{\partial x} + \frac{\partial S_k^*}{\partial \tau} = -\left[g^\diamond(t,x,\tau) + \mu_k^*(x,\tau)\right] S_k^*, \tag{5.9}$$

$$\frac{\partial S_k^\diamond}{\partial t} + \frac{\partial S_k^\diamond}{\partial x} = \int_0^\infty g^\diamond(t,x,\tau)\, S_k^*(t,x,\tau)\, d\tau - \mu_k^\diamond(x)\, S_k^\diamond, \tag{5.10}$$

with $k \in \{1,2\}$, $x > 0$, $\tau > 0$, with the boundary conditions:

$$S_k(t,0) = \frac{B}{2}\left[1 - M(t) + M(t)(1-\varphi(t))\right], \tag{5.11}$$

$$S_k^*(t,x,0) = \phi_k(t,x)\, S_k(t,x), \tag{5.12}$$

$$S_k^*(t,0,\tau) = \frac{B}{2}\, M(t)\, \varphi(t)\, \delta_{\tau=0}\,, \tag{5.13}$$

$$S_k^\diamond(t,0) = 0\,, \tag{5.14}$$

and some initial conditions. Here $\delta_{\tau=0}$ stands for the "Dirac delta function". We assumed that half of all births are male and half are female. The system is completely specified by defining the force of infection $\phi_k(t,x)$ and the function $M(t)$. Let σ_1 (resp. σ_2) be the transmission probability of HIV per partnership from man to woman (resp. from woman to man). Let $c(x)$ be the turnover rate of sexual partners of women aged x. Women

aged x choose a male sexual partner aged x' with a probability density $\omega(x, x')$ such that $\int_0^\infty \omega(x, x')\, dx' = 1$ for all x. Let $\vartheta(t, x)$ represent the probability that a partner of a woman aged x uses a condom. For a given age x, it is a monotone increasing function of time t with $0 \le \vartheta(t, x) \le 1$. Set $c^*(t, x) = (1 - \vartheta(t, x))\, c(x)$. We then assume that the forces of infection for women and men are:

$$\phi_1(t, x) = \int_0^\infty \sigma_1\, c^*(t, x)\omega(x, y)\, \frac{S_2^*(t, y)}{P_2(t, y)}\, dy\,, \tag{5.15}$$

$$\phi_2(t, x) = \int_0^\infty \sigma_2\, c^*(t, y)\, \omega(y, x)\, \frac{S_1^*(t, y)}{P_2(t, x)} dy\,, \tag{5.16}$$

where we set $S_k^*(t, x) = \int_0^\infty S_k^*(t, x, \tau)\, d\tau$ and $P_k(t, x) = S_k(t, x) + S_k^*(t, x) + S_k^\diamond(t, x)$. Notice in Eqs. (5.15)-(5.16) that individuals receiving ART are assumed to have a "normal" sexual life and to be non-infectious. Let $f(x)$ be the fertility of women aged x: the probability of giving birth between age x and $x+dx$ is $f(x)\, dx$ for an infinitesimal dx. Then the prevalence of HIV_+ women who did not receive ART before pregnancy among all pregnant women is

$$M(t) = \frac{\int_0^\infty f(x)\, S_1^*(t, x)\, dx}{\int_0^\infty f(x)\, P_1(t, x)\, dx}\,. \tag{5.17}$$

Notice again that HIV status or ART is assumed not to change the fertility $f(x)$. The use of a constant input of births B that does not take $f(x)$ into account may seem strange but we wanted to avoid the problem of exponential population growth, which in fact requires geographic expansion whereas we will focus on a relatively small homogeneously mixing population as explained below.

In summary, system Eqs. (5.8)-(5.10) is an SIR-type model structured by age, time since infection and gender. The boundary conditions (5.11)-(5.14) reflect vertical transmission of HIV. The forces of infection (5.15)-(5.16) follow a "mass action" principle.

The basic reproduction number R_0 for HIV. For simplicity, we shall neglect mother-to-child transmission. The number of susceptibles in the disease-free steady state $\bar{S}_k(x)$ is given by

$$\bar{S}_k(x) = \exp\Big(-\int_0^x \mu_k(y)\, dy\Big)\, B/2. \tag{5.18}$$

We assume that the rate of starting ART is a time-independent function $g^\diamond(x, \tau)$. We assume that condom use (from the point of view of women) is a time-independent function $\vartheta(x)$ and set $c^*(x) = (1 - \vartheta(x))\, c(x)$. The linearized system near the disease-free steady state is

of the form (5.1)-(5.3), as given in Sect. 5.1, with

$$X = \begin{pmatrix} S_1^* \\ S_2^* \end{pmatrix}, \quad \mathcal{B} = \begin{pmatrix} \mathcal{B}_1 & 0 \\ 0 & \mathcal{B}_2 \end{pmatrix}, \quad \mathcal{A} = \begin{pmatrix} 0 & \mathcal{A}_{1,2} \\ \mathcal{A}_{2,1} & 0 \end{pmatrix},$$
$$\mathcal{B}_k(x,\tau) = g^\diamond(x,\tau) + \mu_k^*(x,\tau),$$
$$\mathcal{A}_{1,2}(x,y,\tau) = \sigma_1\, c^*(x)\, \omega(x,y)\, \bar{S}_1(x)/\bar{S}_2(y),$$
$$\mathcal{A}_{2,1}(x,y,\tau) = \sigma_2\, c^*(y)\, \omega(y,x).$$

R_0 is the spectral radius of the integral operator (5.6) with a kernel given by (5.7). With the notations of the appendix, matrix $\mathcal{C}$ is a diagonal matrix whose diagonal elements are $\mathcal{C}_k(x,\tau) = \exp(\int_0^\tau \mathcal{B}_k(x+y,y)\,dy)$.

5.2.2 TB submodel

In the TB submodel we draw a distinction between individuals who are susceptible to *Mycobacterium tuberculosis* (MTB)(S), who are latently infected with MTB (E), who are actively infected with TB (I) and who are on treatment for TB infection (T). Let $S_k(t,x)$ be the density of uninfected individuals aged x at time t and of gender k. Let $E_k(t,x)$ be the density of individuals with latent TB, $I_k(t,x)$ be the density of individuals with active TB, $T_k(t,x)$ be the density of individuals receiving treatment for active TB. Fig. 5.2 is a diagrammatic presentation of the model.

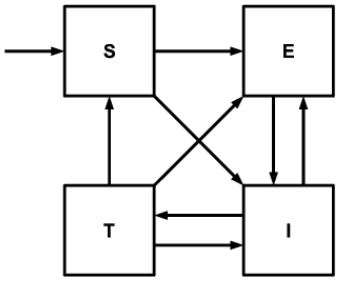


Figure 5.2: TB structure.

The active TB compartment I_k includes all active cases: extrapulmonary cases, smear-negative pulmonary cases and smear-positive pulmonary cases. It would not make much difference to use three separate compartments, with a fixed percentage entering in each. We assume a constant infectiousness at the average of the population.

Let $\psi(t)$ be the force of infection, $p(x)$ (resp. $q(x)$) the probability that primary infection (resp. reinfection) of a person aged x leads shortly afterwards to active TB, $a(x)$ the reactivation rate, γ the rate at which active cases start treatment, s the rate at which individuals stop treatment, ε the probability of being cured after stopping treatment, χ the probability of still having active TB among those that are not cured after treatment. Let m (resp. n) be the TB-specific mortality during active TB (resp. during treatment) and let

β be the natural recovery rate without treatment. With these notations, our model is

$$\begin{aligned}
\frac{\partial S_k}{\partial t} + \frac{\partial S_k}{\partial x} &= \varepsilon\, s\, T_k - \big[\psi(t) + \mu_k(x)\big]\, S_k, \\
\frac{\partial E_k}{\partial t} + \frac{\partial E_k}{\partial x} &= (1-\chi)\,(1-\varepsilon)\, s\, T_k + (1-p(x))\,\psi(t)\, S_k + \beta\, I_k \\
&\quad - \big[q(x)\,\psi(t) + a(x) + \mu_k(x)\big]\, E_k, \\
\frac{\partial I_k}{\partial t} + \frac{\partial I_k}{\partial x} &= \chi\,(1-\varepsilon)\, s\, T_k + \big[p(x)\, S_k + q(x)\, E_k\big]\psi(t) + a(x)\, E_k \\
&\quad - \big[\gamma + \mu_k(x) + m + \beta\big]\, I_k, \\
\frac{\partial T_k}{\partial t} + \frac{\partial T_k}{\partial x} &= \gamma\, I_k - \big[s + \mu_k(x) + n\big]\, T_k,
\end{aligned}$$

with the boundary conditions $S_k(t,0) = B/2$ and $E_k(t,0) = I_k(t,0) = T_k(t,0) = 0$. Let κ be an effective contact rate including transmission probability. We assume that the force of infection $\psi(t)$, i.e., the rate at which individuals move out of the susceptible compartment, is simply the product of κ and of the prevalence of active TB, the latter being the probability that the contact is with an active TB case. Therefore

$$\psi(t) = \kappa\, \frac{\int_0^\infty (I_1 + I_2)\, dx}{\int_0^\infty (P_1 + P_2)\, dx}\,,$$

where $P_k = S_k + E_k + I_k + T_k$ is the total population of gender k.

The basic reproduction number R_0 for TB. The disease-free steady state $\bar{S}_k(x)$ is again given by (5.18). Adding an additional variable τ for the time since infection, the linearized system near the disease-free steady state can be put in the form Eqs. (5.1)-(5.3), as given in Sect. 5.1, with

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix},\ \mathcal{B} = \begin{pmatrix} \mathcal{B}_1 & 0 \\ 0 & \mathcal{B}_2 \end{pmatrix},\ A = \begin{pmatrix} \mathcal{A}_1 & \mathcal{A}_1 \\ \mathcal{A}_2 & A_2 \end{pmatrix},\ X_k = \begin{pmatrix} E_k \\ I_k \\ T_k \end{pmatrix},$$

$$\mathcal{B}_k(x,\tau) = \begin{pmatrix} -(a(x) + \mu_k(x)) & \beta & (1-\chi)(1-\varepsilon)s \\ a(x) & -(\gamma + \mu_k(x) + m + \beta) & \chi(1-\varepsilon)s \\ 0 & \gamma & -(s + \mu_k(x) + n) \end{pmatrix},$$

$$\mathcal{A}_k(x,y,\tau) = \begin{pmatrix} 0 & 1-p(x) & 0 \\ 0 & p(x) & 0 \\ 0 & 0 & 0 \end{pmatrix} \frac{\kappa\, \bar{S}_k(x)}{\int_0^\infty (\bar{S}_1(z) + \bar{S}_2(z))\, dz}\,.$$

R_0 is the spectral radius of the integral operator (5.6) with a kernel given by (5.7).

5.2.3 HIV-TB model

Combining the HIV submodel and the TB submodel, we obtain $4\times3\times2 = 24$ compartments, as shown in Tab. 5.1. Newborns can be HIV_+ due to the possible vertical transmission of HIV, but are born susceptible to MTB.

Table 5.1: Structure of the HIV-TB model

		TB			
		susceptible	latent	active	treated
Women(1)	HIV_-	S_1	E_1	I_1	T_1
	HIV_+	S_1^*	E_1^*	I_1^*	T_1^*
	ART	$S_1^\diamond$	$E_1^\diamond$	$I_1^\diamond$	$T_1^\diamond$
Men(2)	HIV_-	S_2	E_2	I_2	T_2
	HIV_+	S_2^*	E_2^*	I_2^*	T_2^*
	ART	$S_2^\diamond$	$E_2^\diamond$	$I_2^\diamond$	$T_2^\diamond$

The model equations for HIV_- individuals are

$$\begin{aligned}
\frac{\partial S_k}{\partial t} + \frac{\partial S_k}{\partial x} &= \varepsilon\, s(t)\, T_k - \big[\phi_k(t,x) + \psi(t) + \mu_k(x)\big]\, S_k, \\
\frac{\partial E_k}{\partial t} + \frac{\partial E_k}{\partial x} &= (1-\chi)(1-\varepsilon)\, s(t)\, T_k + (1-p(x))\, \psi(t)\, S_k + \beta\, I_k \\
&\quad - \big[\phi_k(t,x) + q(x)\, \psi(t) + a(x) + \mu_k(x)\big]\, E_k, \\
\frac{\partial I_k}{\partial t} + \frac{\partial I_k}{\partial x} &= \chi\,(1-\varepsilon)\, s(t)\, T_k + \psi(t)\, \big[p(x)\, S_k + q(x)\, E_k\big] + a(x)\, E_k \\
&\quad - \big[\phi_k(t,x) + \gamma + \mu_k(x) + m + \beta\big]\, I_k, \\
\frac{\partial T_k}{\partial t} + \frac{\partial T_k}{\partial x} &= \gamma\, I_k - \big[\phi_k(t,x) + s(t) + \mu_k(x) + n\big]\, T_k\,,
\end{aligned}$$

with the "boundary" conditions

$$S_k(t,0) = [1 - M(t) + M(t)(1-\varphi(t))]B/2\,, \quad E_k(t,0) = I_k(t,0) = T_k(t,0) = 0.$$

We note that, compared to the TB submodel, we have now assumed that the rate $s(t)$ at which individuals stop TB treatment may depend on time. This is because we want to take into account increased detection rates and coverage under the World Health Organization's

DOTS strategy during the 1990s. The model equations for HIV_+ individuals are

$$\begin{aligned}
\frac{\partial S_k^*}{\partial t} + \frac{\partial S_k^*}{\partial x} + \frac{\partial S_k^*}{\partial \tau} &= \varepsilon\, s(t)\, T_k^* - \left[\psi(t) + g^\diamond(t,x,\tau) + \mu_k^*(x,\tau)\right] S_k^*, \\
\frac{\partial E_k^*}{\partial t} + \frac{\partial E_k^*}{\partial x} + \frac{\partial E_k^*}{\partial \tau} &= (1-\chi)\,(1-\varepsilon)\, s(t)\, T_k^* + (1-p^*)\,\psi(t)\, S_k^* + \beta^*\, I_k^* \\
&\quad - \left[q^*\,\psi(t) + a^*(\tau) + g^\diamond(t,x,\tau) + \mu_k^*(x,\tau)\right] E_k^*, \\
\frac{\partial I_k^*}{\partial t} + \frac{\partial I_k^*}{\partial x} + \frac{\partial I_k^*}{\partial \tau} &= \chi\,(1-\varepsilon)\, s(t)\, T_k^* + \psi(t)\left[p^*\, S_k^* + q^*\, E_k^*\right] + a^*(\tau)\, E_k^* \\
&\quad - \left[\gamma^* + g^\diamond(t,x,\tau) + \mu_k^*(x,\tau) + m^* + \beta^*\right] I_k^*, \\
\frac{\partial T_k^*}{\partial t} + \frac{\partial T_k^*}{\partial x} + \frac{\partial T_k^*}{\partial \tau} &= \gamma^*\, I_k^* - \left[s(t) + g^\diamond(t,x,\tau) + \mu_k^*(x,\tau) + n^*\right] T_k^*\,,
\end{aligned}$$

with the "boundary" conditions

$$\begin{aligned}
&S_k^*(t,x,0) = \phi_k(t,x)\, S_k(t,x), \quad S_k^*(t,0,\tau) = M(t)\,\varphi(t)\,(B/2)\,\delta_{\tau=0} \\
&E_k^*(t,x,0) = \phi_k(t,x)\, E_k(t,x), \quad E_k^*(t,0,\tau) = 0, \\
&I_k^*(t,x,0) = \phi_k(t,x)\, I_k(t,x), \quad I_k^*(t,0,\tau) = 0, \\
&T_k^*(t,x,0) = \phi_k(t,x)\, T_k(t,x), \quad T_k^*(t,0,\tau) = 0.
\end{aligned}$$

Notice that the TB reactivation rate $a^*(\tau)$ is assumed to depend essentially on the time τ since infection and not on the age x. Parameters γ^*, β^*,m^*,n^* are analogous to γ, β, m, n but for HIV_+ individuals. Parameters p^* and q^* are analogous to $p(x)$ and $q(x)$ except that the age dependence probably does not matter much (most HIV_+ individuals are adults). The model equations for individuals receiving ART are

$$\begin{aligned}
\frac{\partial S_k^\diamond}{\partial t} + \frac{\partial S_k^\diamond}{\partial x} &= \int_0^\infty g^\diamond(t,x,\tau) S_k^*\, d\tau + \varepsilon\, s(t)\, T_k^\diamond - \left[\psi(t) + \mu_k(x)\right] S_k^\diamond, \\
\frac{\partial E_k^\diamond}{\partial t} + \frac{\partial E_k^\diamond}{\partial x} &= \int_0^\infty g^\diamond(t,x,\tau)\, E_k^*\, d\tau + (1-\chi)\,(1-\varepsilon)\, s(t)\, T_k^\diamond + (1-p^\diamond)\,\psi(t)\, S_k^\diamond \\
&\quad + \beta^\diamond\, I_k^\diamond - \left[q^\diamond\,\psi(t) + a^\diamond + \mu_k(x)\right] E_k^\diamond, \\
\frac{\partial I_k^\diamond}{\partial t} + \frac{\partial I_k^\diamond}{\partial x} &= \chi\,(1-\varepsilon)\, s(t)\, T_k^\diamond + \psi(t)\left[p^\diamond\, \widehat{S}_k^\diamond + q^\diamond\, E_k^\diamond\right] + a^\diamond\, E_k^\diamond \\
&\quad - \left[\gamma^\diamond + \mu_k(x) + m^\diamond + \beta^\diamond\right] I_k^\diamond, \\
\frac{\partial T_k^\diamond}{\partial t} + \frac{\partial T_k^\diamond}{\partial x} &= \int_0^\infty g^\diamond(t,x,\tau)\Big[I_k^* + T_k^*\Big]\, d\tau + \gamma^\diamond\, I_k^\diamond - \left[s(t) + \mu_k(x) + n^\diamond\right] T_k^\diamond\,,
\end{aligned}$$

with the "boundary" conditions $S_k^\diamond(t,0) = E_k^\diamond(t,0) = I_k^\diamond(t,0) = T_k^\diamond(t,0) = 0$. Notice that individuals with active TB are assumed to get treatment against TB if they have been chosen to start ART (the term $\int_0^\infty g^\diamond(t,x,\tau)\, I_k^*\, d\tau$ enters the equation for $T_k^\diamond$ and not for $I_k^\diamond$). For $k \in \{1,2\}$, let $R_k = S_k + E_k + I_k + T_k$ be the total number of HIV_- individuals, $R_k^*(t,x) = \int_0^\infty [S_k^* + E_k^* + I_k^* + T_k^*]\, d\tau)$ the total number of HIV_+ individuals without ART, $R_k^\diamond = S_k^\diamond + E_k^\diamond + I_k^\diamond + T_k^\diamond$ the total number of individuals receiving ART, and $P_k = R_k + R_k^* + R_k^\diamond$.

We assume that the forces of HIV infection are given by

$$\phi_1(t,x) = \int_0^\infty \sigma_1\, c^*(t,x)\omega(x,y)\, \frac{S_2^*(t,y)}{P_2(t,y)}\, dy$$
$$\phi_2(t,x) = \int_0^\infty \sigma_2\, c^*(t,x')\, \omega(y,x)\, \frac{S_1^*(t,y)}{P_2(t,x)} dy$$

Let $P(t)$ be the total population at time t. We assume that

$$\psi(t) = \frac{\kappa}{P(t)} \int_0^\infty \Big[I_1 + I_2 + I_1^\diamond + I_2^\diamond + \rho \int_0^\infty [I_1^* + I_2^*]\, d\tau \Big]\, dx$$

is the TB infection rate, where $\rho < 1$ takes into account the fact that HIV_+ TB cases are less likely to have smear-positive pulmonary TB.

The basic reproduction number R_0 for TB invading HIV. It is widely thought that HIV 'drives the TB epidemic in South Africa. The influence of HIV on TB can be studied by formulating a next generation matrix for TB 'invading' HIV in endemic steady state. The linearized system near the disease-free steady state can be put in the form (5.1)-(5.3),

as given in Sect. 5.1, with

$$X = \begin{pmatrix} \mathbb{X}_1 \\ \mathbb{X}_2 \end{pmatrix},\ \mathcal{B} = \begin{pmatrix} \mathcal{B}_1 & 0 \\ 0 & \mathcal{B}_2 \end{pmatrix},\ A = \begin{pmatrix} \mathcal{A}_1 & \mathcal{A}_1 \\ \mathcal{A}_2 & A_2 \end{pmatrix},$$

$$\mathbb{X}_k = \begin{pmatrix} X_k \\ X_k^* \\ X_k^\diamond \end{pmatrix},\ X_k = \begin{pmatrix} E_k \\ I_k \\ T_k \end{pmatrix}, X_k^* = \begin{pmatrix} E_k^* \\ I_k^* \\ T_k^* \end{pmatrix}, X_k^\diamond = \begin{pmatrix} E_k^\diamond \\ I_k^\diamond \\ T_k^\diamond \end{pmatrix}$$

$$\mathcal{B}_k(x,y,\tau) = \left(\begin{array}{ccc|ccc|c} & B_k & & & 0 & & 0 \\ \hline \phi_k(z) & & & & & & \\ & \phi_k(z) & & & B_k^* & & 0 \\ & & \phi_k(z) & & & & \\ \hline & & & g^\diamond(z) & & & \\ & 0 & & & g^\diamond(z) & & B_k^\diamond \\ & & & & & g^\diamond(z) & \end{array}\right)$$

$$B_k = \begin{pmatrix} -(\mu_k(z)+a(z)) & \beta & (1-\chi)(1-\varepsilon)s \\ a(z) & -(\mu_k(z)+m+\gamma+\beta) & \chi(1-\varepsilon)s \\ 0 & \gamma & -(\mu_k(z)+n+s) \end{pmatrix}$$

$$B_k^* = \begin{pmatrix} -(\mu_k^*(z)+a^*(z,\tau)) & \beta & (1-\chi)(1-s)s \\ a^*(z,\tau) & -(\mu_k^*(z)+m^*+\gamma^*+\beta) & \chi(1-\varepsilon)s \\ 0 & \gamma^* & -(\mu_k^*(z)+n^*+s) \end{pmatrix}$$

$$B_k^\diamond = \begin{pmatrix} -(\mu_k(z)+a^\diamond(z,\tau)) & \beta & (1-\chi)(1-\varepsilon)s \\ a^\diamond(z,\tau) & -(\mu_k(z)+m^\diamond+\gamma^\diamond+\beta) & \chi(1-\varepsilon)s \\ 0 & \gamma^\diamond & -(\mu_k(z)+n^\diamond+s) \end{pmatrix}$$

$$\mathcal{A}_k(x,y,\tau) = \frac{1}{\bar{P}} \left(\begin{array}{ccc|ccc|ccc} 0 & \bar\kappa\,(1-p)\,\bar S(x) & 0 & 0 & \rho\bar\kappa\,(1-p)\,\bar S(x) & 0 & 0 & \bar\kappa\,(1-p)\,\bar S(x) & 0 \\ 0 & \bar\kappa\,p\,\bar S(x) & 0 & 0 & \rho\bar\kappa\,p\,\bar S(x) & 0 & 0 & \bar\kappa\,p\,\bar S(x) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & \bar\kappa\,(1-p)\,\bar S^*(x) & 0 & 0 & \rho\bar\kappa\,(1-p)\,\bar S^*(x) & 0 & 0 & \bar\kappa\,(1-p)\,\bar S^*(x)0 & \\ 0 & \bar\kappa\,p\,\bar S^*(x) & 0 & 0 & \rho\bar\kappa\,p\,\bar S^*(x) & 0 & 0 & \bar\kappa\,p\,\bar S^*(x) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & \bar\kappa\,(1-p)\,\bar S^\diamond(x) & 0 & 0 & \rho\bar\kappa\,(1-p)\,\bar S^\diamond(x) & 0 & 0 & \bar\kappa\,(1-p)\,\bar S^\diamond(x) & 0 \\ 0 & \bar\kappa\,p\,\bar S^\diamond(x) & 0 & 0 & \rho\bar\kappa\,p\,\bar S^\diamond(x) & 0 & 0 & \bar\kappa\,p\,\bar S^\diamond(x) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \end{array}\right)$$

where $z = y + \tau$ and $\bar S(x), \bar S^*(x)$ and $\bar S^*(x)$ refers to the equilibrium density of individuals who are HIV_-, HIV_+ and on ART treatment respectively. $\bar P$ is the size of the total population in HIV steady state.

5.3 Parameter values

5.3.1 Demography and HIV

For the gender- and age-specific natural mortality $\mu_k(x)$ above 15 years of age, we used the 1997 South African data from [4], which we smoothed somewhat to cancel out excess mortality among adults in their 30s due to the rising HIV epidemic (Fig. 5.3(a)). We know from [2] that infant mortality under age 1 was 43 (resp. 47) deaths per 1,000 female (resp. male) births in 1990. We also know from [3] that under-5 mortality was 57 (resp. 63) per 1,000 live female (resp. male) births in 1990. Between age 5 and 15, we tried to calculate a reasonable estimate. For the age-specific fertility of women $f(x)$, we used the data from [102], which is given in 5-year age groups (Fig. 5.3(b)).

The assumed turnover rate of male sexual partners $c(x)$ for women is shown in Fig.5.3(c), with $c(x) = 0$ for ages under 14 years. We assume that function $\omega(x, \cdot)$, describing the choice of the partner's age, is a "triangular" probability distribution, i.e., equal to 0 for $x' < x_{\min}(x)$ and $x' > x_{\max}(x)$, with a maximum at $x' = x_{\text{opt}}(x)$ and piecewise linear (Fig. 5.3(d)).

For the level of condom use $\vartheta(t, x)$, we assume that it is maximum at sexual debut (14 years) and decreases linearly with age until a certain age above which it stays constant. The maximum of $\vartheta(t, x)$ is assumed to increase from 0 in 1993 to 70% in 2007 and to stay constant afterwards. The minimum is assumed to increase from 0 before 2000 to 50% in 2020 (the horizon of our simulations). The age above which condom use stays constant is assumed to increase from 15 in 1990 to 60 in 2000. Fig. 5.4(e) shows, e.g., condom use at ages 20, 30, 40 and 50 as a function of time. Reported condom use at last sex in the age group 15–24 for the years 1998, 2003, 2005 and 2006 and in the age group 25–49 for the year 2005 was taken from [1], which summarizes different surveys (see data points in Fig. 5.4(e)).

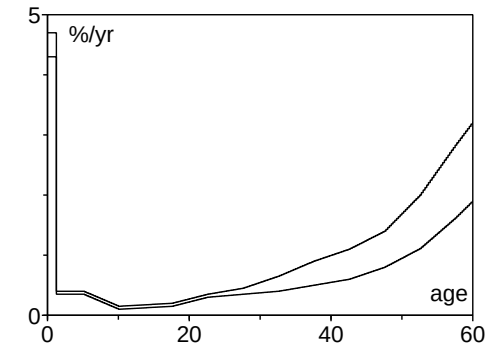
We assume that the first person infected with HIV is introduced in 1985. One large HIV survey done in 1986 showed that there was almost no HIV in South Africa at that time. We assume that the first infected person is a woman aged 17. We assumed that the transmission probability per partnership was $\sigma_2 = 45\%$ from woman to man and $\sigma_1 = 90\%$ from man to woman. We assume that there are only $B = 20$ births per year in the model community. With a median survival prior to HIV era of about 60 years, the homogeneously mixing population consists of approximately 1,200 individuals. We note that this assumption on B is critical to be able to reach a prevalence of HIV close to 1% in 1990 starting from just one initial case in 1985 and with an initial doubling time of the epidemic of about 1.5 years.

We assume for simplicity that the mortality of HIV_+ individuals is

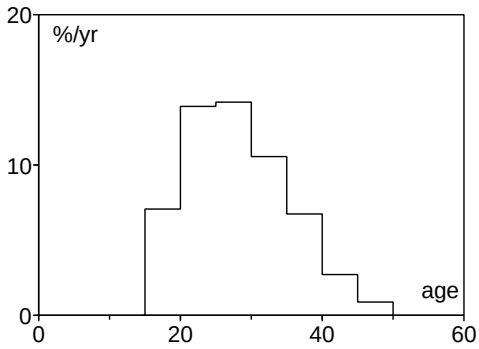
$$\mu_k^*(x, \tau) = \mu_k(x) + \nu(\tau)\,,$$

where $\nu(\tau)$ is an age-independent AIDS-specific mortality. Fig. 5.4(a) shows $\nu(\tau)$ as a function of the time since infection τ. It gives a median survival time of 7 years. We assume that individuals receiving ART have the same mortality as HIV_- individuals: $\mu_k^\diamond(x) = \mu_k(x)$.

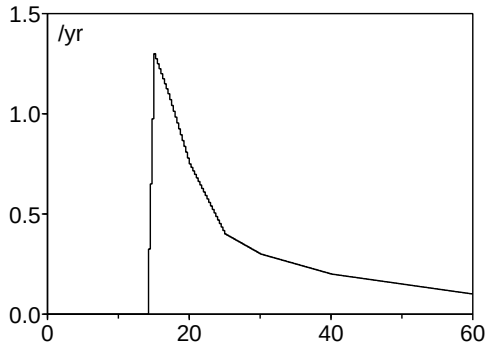
In South Africa, nevirapine became available in public hospitals in 2001. We assume



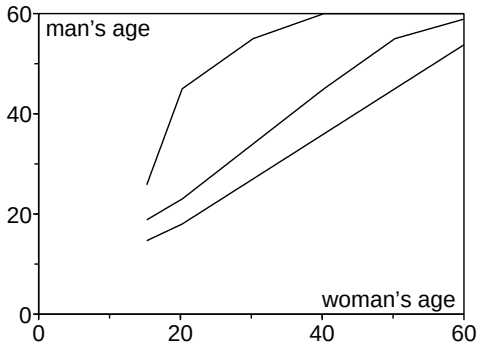
(a) mortality



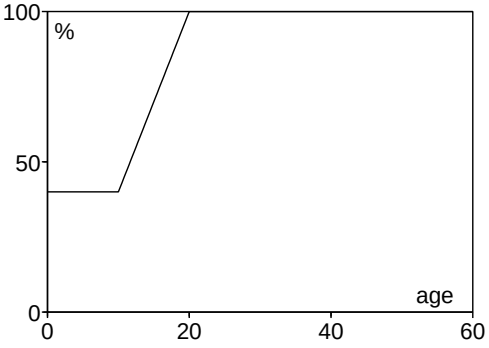
(b) fertility



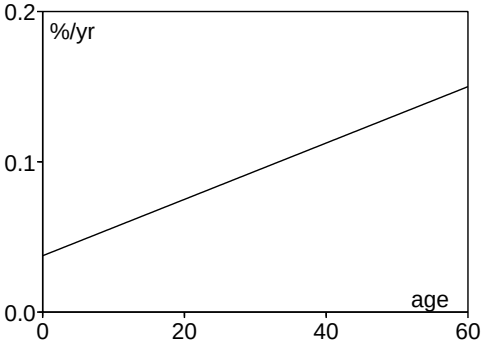
(c) turnover of sexual partners



(d) partner choice



(e) reduced TB risk



(f) TB reactivation rate

Figure 5.3: Age-dependent parameters for the HIV-TB model for South Africa – Sect.5.2.3.

that it reduces the mother-to-child transmission probability of HIV from 20% to 5% and that access to nevirapine increased from 0 to 100% between 2001 and 2006 (Fig. 5.4(d)). Fig. 5.4(c) shows the rate $g_0^\diamond(\tau)$ at which HIV_+ individuals start ART as a function of time since infection τ in the basic scenario. ART became available in public hospitals only in 2004. We assume a progressive access $A(t)$ to ART between 2003 and 2008 (Fig. 5.4(f)). For UTTS, we call F the frequency at which individuals are tested for HIV and assume that $F = 1/\text{yr}$. For the implementation of the UTTS we assume a schedule $C(t)$ as in Fig. 5.4(f), i.e., full implementation within the first quarter of 2010. More precisely, we assume that the rate of starting ART is

$$g^\diamond(t, x, \tau) = A(t)\, g_0^\diamond(\tau) + F\, C(t)\, U,$$

where $U = 1$ or 0 depending on whether the "universal test and treat strategy" is used or not.

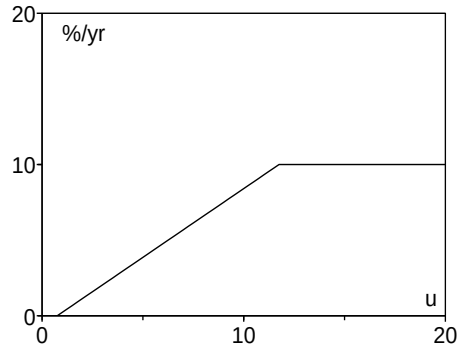
5.3.2 TB parameters before the HIV era

We choose an effective contact rate $\kappa = 10/\text{yr}$, an age-dependent probability of fast progression after primary infection $p(x) = 15\%\, \theta(x)$ with $\theta(x)$ as in Fig. 5.3(e), a probability of fast progression after reinfection $q(x) = 9\%\, \theta(x)$, a reactivation rate $a(x)$ as in Fig. 5.3(f). The coefficient $\theta(x)$ in Fig. 5.3(e) is similar to [109, Fig. 2]. Active TB cases are assumed to suffer of a TB-specific mortality $m = 0.25/\text{yr}$. We also include a natural recovery rate $\beta = 0.25/\text{yr}$. In this way, the probability of surviving from active TB without any treatment would be $\beta/(m + \beta) = 50\%$, and the average length of disease without treatment would be $1/(m + \beta) = 2$ years.

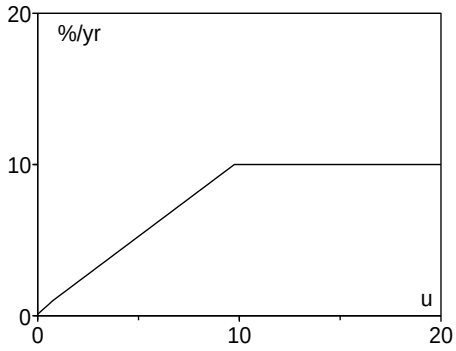
We assume that active cases are "detected", i.e., start treatment at a rate $\gamma = 0.75/\text{yr}$, which corresponds to a case detection "rate" equal to $\gamma/(\gamma + m + \beta) = 60\%$. The average time spent with active TB (excluding the period under treatment) is $1/(\gamma + m + \beta) \simeq 0.8$ years or approximately 10 months. The average length of treatment is assumed to be 18 months before DOTS and 6 months with DOTS. In South Africa, DOTS coverage increased approximately linearly from 0 to 100% between 1996 and 2002 (Fig. 5.4(f)). Therefore we assume that $s(t) = (1 - D(t))\, s_0 + D(t)\, s_1$, where $D(t)$ is as in Fig. 5.4(f), $1/s_0 = 1.5$ years, and $1/s_1 = 0.5$ years. Mortality during treatment is assumed to be $\mu = 0.05/\text{yr}$. We choose a fraction of successful treatments equal to $\varepsilon = 70\%$. We assume that half of unsuccessful treatments return to the compartment with active TB: $\chi = 0.5$.

5.3.3 Parameters related to HIV-TB interaction

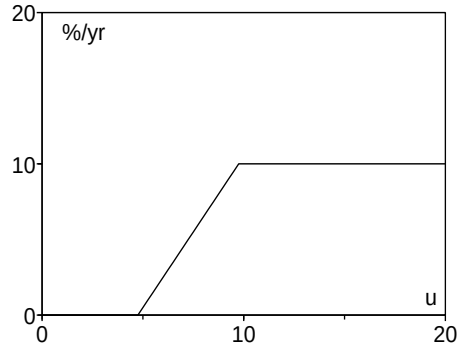
We assume that the infectiousness of HIV_+ TB cases is $\rho = 2/3$ of that of HIV_- TB cases. We choose $p^* = 30\%$ and $q^* = 20\%$. The estimate for p^* is substantiated by [101], who estimated that TB incidence doubled within the first year of HIV infection, with a further slight increase in HIV_+ miners for longer periods. Another study suggests that the



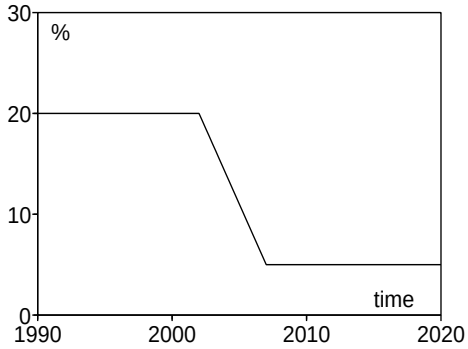
(a) AIDS mortality



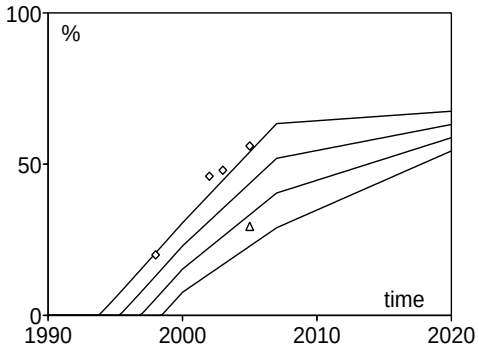
(b) TB reactivation rate if HIV_+



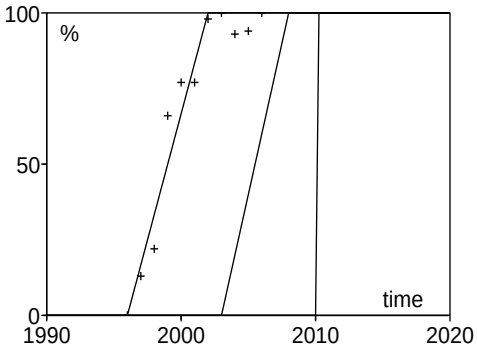
(c) access to ART



(d) mother-to-child transmission



(e) condom use (age 20,30,40,50)



(f) DOTS, ART and UTTS

Figure 5.4: Model parameters that depend on time since infection or that change over time.

cumulative lifetime risk of MTB infection may be 5 times higher for HIV_+ than for HIV- individuals [120].

Keeping a natural recovery rate $\beta^* = \beta = 0.25/\text{yr}$, we choose a TB-specific death rate $m^* = 0.4/\text{yr}$ so that the probability of surviving without treatment is $\beta^*/(m^* + \beta^*) \simeq 38\%$, while the average length of disease without treatment is $1/(m^* + \beta^*) \simeq 1.5$ years. We assumed that active cases are detected at a rate $\gamma^* = 1.5/\text{yr}$. The corresponding case detection rate is $\gamma^*/(\gamma^* + m^* + \beta^*) \simeq 70\%$. The average length of active TB (excluding treatment) is $1/(\gamma^* + m^* + \beta^*) \simeq 0.47$ years (5.5 months). We assume that mortality during treatment for TB in HIV_+ individuals is $\mu^* = 0.1/\text{yr}$.

We assumed that the reactivation rate with ART is $a^\diamond = 0.01/\text{yr}$, independently of the stage of HIV-infection at which ART starts (this is obviously a crude assumption). Notice that the value of $a^\diamond$ is the same as $a^*(\tau)$ after $\tau = 1$ year of HIV-infection, but still 6 times higher than the maximum value of the reactivation rate $a(x)$ for HIV_- individuals. We assumed that $p^\diamond = 15\%$, $q^\diamond = 10\%$, $m^\diamond = m$, $\beta^\diamond = \beta$, $\mu^\diamond = \mu$ (same mortality as HIV_- individuals), $\gamma^\diamond = \gamma^*$ (same detection rate as HIV_+ individuals).

All parameter values are summarized in Tab. 5.2.

5.4 Simulation results

The model is solved numerically using a finite difference scheme and Euler's method. As in Sections 2.3 and 4.3, the HIV-related parameters are fitted first in a model with HIV only, justified on the grounds that TB does not directly influence the spread of HIV.

The overall scale of the HIV epidemic is set by adjusting the transmission probability per partnership from women to men (between 0 and 50%), assuming that the transmission probability from men to women is twice this value. Together with a reasonable partner turnover rate, for which our model gives about 10 lifetime partners at age 30, the model gives a realistic logistically-shaped HIV prevalence over time. The steepness of the initial rise is sensitive to the birth rate B, as pointed out in Sect. 5.3.1. The steepness of the increase among 15- to 25-year-olds is sensitive to the partner turnover rate of women $c(x)$. The difference between the ages at which HIV prevalence peaks between men and women is adjusted by means of the partner-choice function $\omega(x, y)$. This difference can also be attributed to the difference in ages of sexual debut for women and men [50] (see also Sect. 6.2), a dynamical effect we did not include. Adjusting these parameters to the values listed in Tab. 5.2 gives a reasonable fit to the age-specific antenatal clinic data.

Fig. 5.5 shows the results of simulating the model with the parameter values with or without UTTS. Fig. 5.5(a) shows the prevalence of HIV in antenatal clinics. Using data from [10, p. 9]. Fig. 5.5(b) shows the crude death rate. The time series is the one estimated by the US Census Bureau. Fig. 5.5(c) shows the prevalence of ART in the total population (data obtained from [37]). For example, an estimated 488,739 individuals were receiving ART in South Africa in November 2007, about 1% of a population of 48 million. Fig. 5.5(d)

Table 5.2: Parameter values.

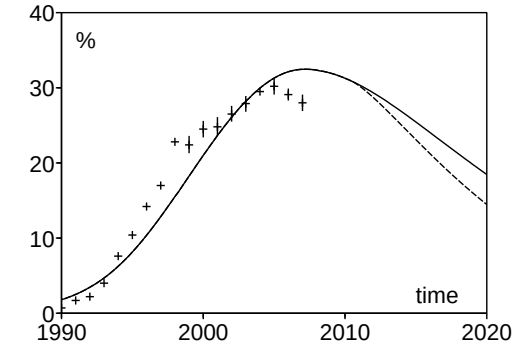
year of introduction of HIV	t_0	1985
age of first infected case (a woman)		17 yrs
births in the community	B	20/yr
mortality	$\mu_k(x)$	Fig. 5.3(a)
fertility	$f(x)$	Fig. 5.3(b)
turnover of male sexual partners	$c(x)$	Fig. 5.3(c)
choice of male sexual partner	$\omega(x, x')$	Fig. 5.3(d)
reduced risk of fast progression	$\theta(x)$	Fig. 5.3(e)
AIDS mortality	$\nu(\tau)$	Fig. 5.4(a)
starting rate of ART in the basic program	$g_0^\diamond(\tau)$	Fig. 5.4(c)
mother-to-child transmission probability	$\varphi(t)$	Fig. 5.4(d)
condom use	$\vartheta(t, x)$	Fig. 5.4(e)
DOTS implementation	$D(t)$	Fig. 5.4(f)
implementation of the new strategy	$C(t)$	Fig. 5.4(f)
access to the basic ART program	$A(t)$	Fig. 5.4(f)
frequency of HIV testing in the new strategy	F	1/yr
transmission probability from man to woman	σ_1	90%
transmission probability from woman to man	σ_2	45%
effective TB contact rate	κ	10/yr
self-recovery rate	β	0.25/yr
length of treatment without DOTS	$1/s_0$	1.5 yrs
length of treatment with DOTS	$1/s_1$	0.5 yrs
treatment success	ε	70%
unsuccessful treatments with active TB	χ	0.5
reduced infectiousness of HIV+ TB cases	ρ	2/3

	HIV_-	HIV_+	ART
primary infection	$p(x) = 15\%\,\theta(x)$	$p^* = 30\%$	$p^\diamond = 15\%$
reinfection	$q(x) = 9\%\,\theta(x)$	$q^* = 15\%$	$q^\diamond = 10\%$
reactivation	$a(x)$, Fig. 5.3(f)	$a^*(\tau)$, Fig. 5.4(b)	$a^\diamond = 0.01$/yr
detection rate	$\gamma = 0.75$/yr	$\gamma^* = 1.5$/yr	$\gamma^\diamond = \gamma^*$
tb mortality	$m = 0.25$/yr	$m^* = 0.5$/yr	$m^\diamond = m$
tb treatment mortality	$n = 0.05$/yr	$n^* = 0.1$/yr	$n^\diamond = n$

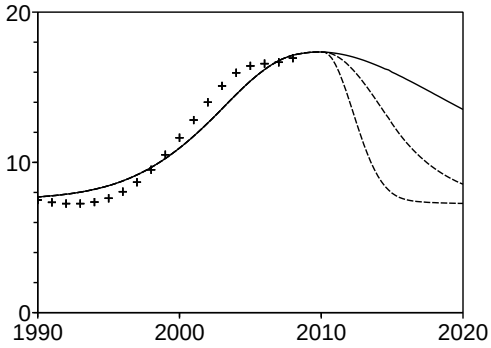
shows the age-specific prevalence of HIV among men in 2005.

The choice of TB parameters is guided by similar considerations as in [17]. The effective contact rate κ and the rate at which TB cases are detected for treatment γ, is adjusted to fit the TB notification data for South Africa. Fig. 5.5(e) shows the TB notification rate and Fig. 5.5(f) the TB-specific mortality. The data come from [11]. We do not show TB prevalence or incidence as these statistics are not observed directly but estimated using uncertain assumptions.

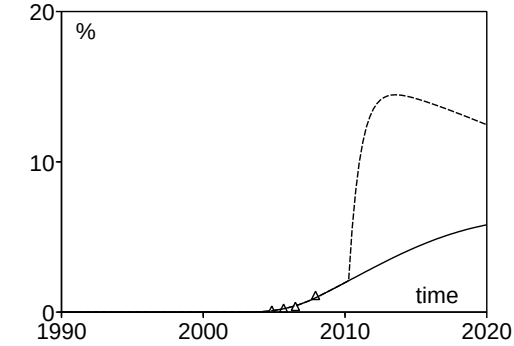
Fig. 5.6 shows the prevalence of HIV disaggregated into in 5-year age groups in women. Notice that this prevalence coincides with the prevalence in antenatal clinics because we assumed that $f(x)$ is piecewise constant with 5-year age steps (Fig. 5.3(b)). The antenatal clinic data by age group and the confidence intervals come from [10, p. 19] for the years 2005–2007, from [6, p. 8] for 2002-2003, from [5, p. 11] for 1999-2001. A national survey of HIV prevalence [7] in 2005 gives extra data points for female age groups and the only data points available for male age groups.



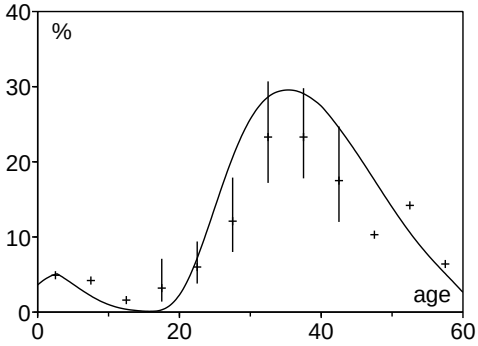
(a) HIV in antenatal clinics



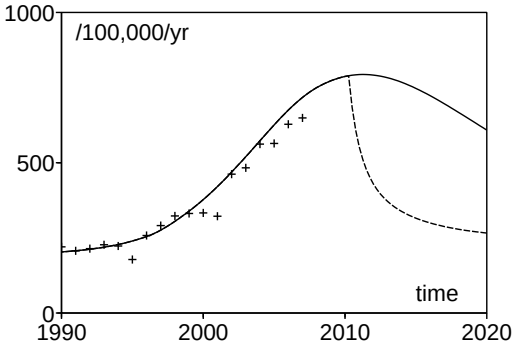
(b) crude death rate



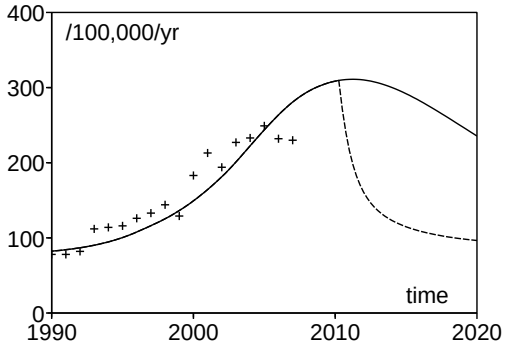
(c) prevalence of ART



(d) HIV in men in 2005



(e) TB notification rate



(f) TB mortality

Figure 5.5: Simulations and data. Plain line: no extra intervention. Dashed line: with UTTS.

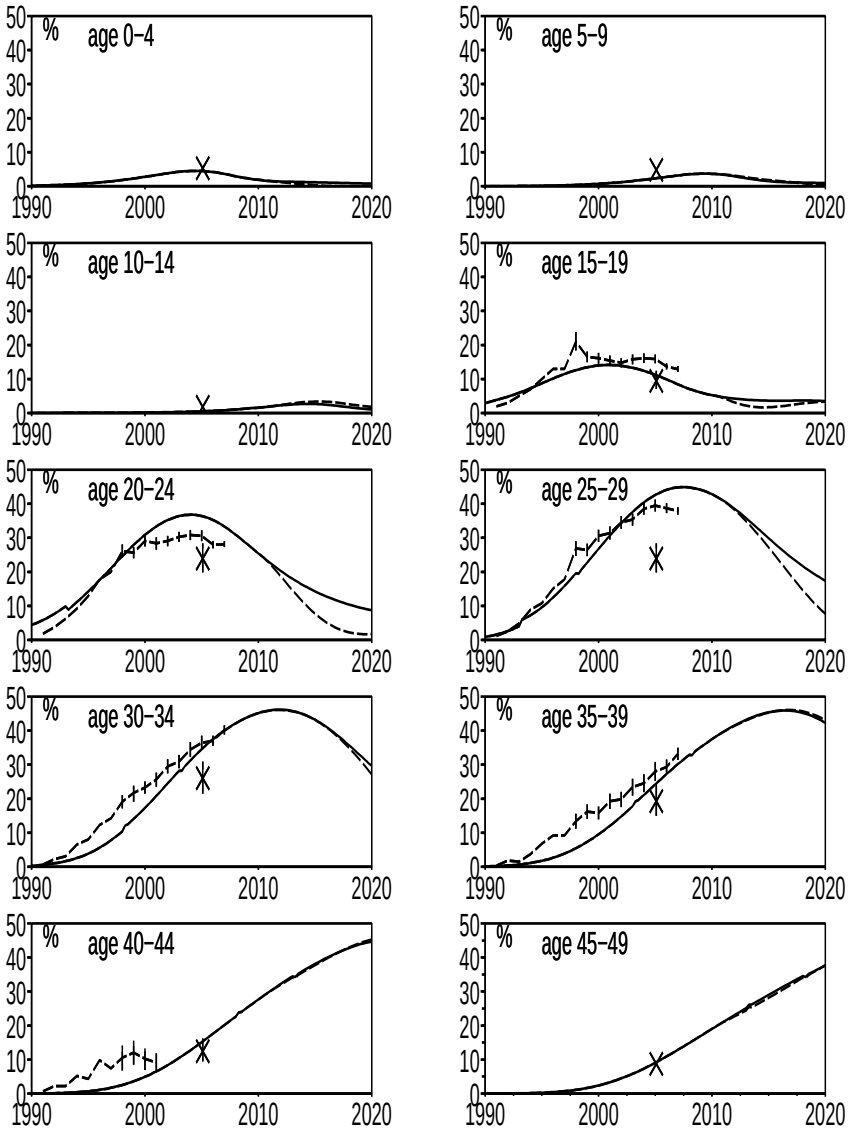


Figure 5.6: Prevalence of HIV among women in 5-year age groups. Simulation and data. $*$ is a data point from a 2005 national survey of HIV prevalence [7].

5.5 Modeling the impacts of intervention

5.5.1 Universal testing and treatment for HIV

The basic reproduction number for HIV can be used to study the possible impact of interventions targeting the HIV epidemic. The kernel K of the next-generation operator, defined by Eq. 5.7 in Sect. 5.2.1, is calculated with susceptibles in demographical steady state. The age distributions of susceptible women (solid line) and men (dashed line) $S_k(x)$ are shown in Fig. 5.7(a). If there were no condom use, no testing for HIV, and no subsequent ART, for the whole duration of the epidemic, then computing R_0, the spectral radius of this matrix, gives $R_0 = 2.00$. The corresponding left-eigenfunction, which is the reproduction value (RV) of women (solid line) and men (dashed line) in the HIV submodel, is shown in Fig. 5.7(b). It appears that men are making the greatest contribution to the modelled HIV epidemic. The reproduction value of women decreases from sexual debut onwards, due to a decrease in partner turnover rate $c(x)$, as shown in Fig. 5.3(c). The partner turnover rate of men is not explicitly modelled, and the RV for men is a function the RV of women, and the age-dependent preference women have for their partners.

This value for $R_0 = 2.0$ is lower than our estimate of $R_0 = 7.0$ in Masiphumelele in Sect. 2.4 and is also lower than the value reported by Williams, et al. [118, p.562], who used a model with no age structure. It is similar, however, to the estimate in [54], who studied the impact of HIV disease progression on R_0 of HIV. $R_0 = 2.0$ as calculated in Sect. 5.2.1 cannot be directly compared to the value of $R_0 \approx 8$ [118, p.562], estimated in a model without age structure. If one adds the rows in the matrix K, one can indeed find age groups of individuals who infect more than 8 partners of any age. Fig. 5.7(c) shows the number of secondary infections (of al ages) caused by women (solid line) and men (dashed line) as a function of the age at which they were infected. The averaging procedure across all age categories, by means of the spectral radius of K, brings down considerably the 'average' number of secondary infections caused over an 'average' lifetime.

We see that $\mathcal{R}_0$ is a kind of average of the number of secondary cases caused by men and women of different ages, as depicted in Fig. 5.7(c). However, the basic reproduction number $\mathcal{R}_0$ does not have a simple interpretation for epidemic models with many infective types. Formally $\mathcal{R}_0$ is the eventual growth rate of one generation infectives to the next. It can be interpreted as the average number of secondary cases produced by a 'typical' index case, once a stable distribution of infectives are reached with respect to age and time since infection.

The average number of lifetime partners for a 14-year-old HIV$_-$ woman, is given by:

$$\int_{14}^{\infty} c(x)\, e^{-\int_{14}^{x} v_1(y)\, dy}\, dx \approx 15. \tag{5.19}$$

On these grounds, the partner turnover rate in the HIV submodel and a value of $R_0 = 2.0$ is plausible.

Solving Eq. (5.5) for the growth rate numerically gives $r = 0.28$ per year, which gives

doubling time of $\log(2)/r = 2.5$ years. It should be noted that this doubling time corresponds to the eventual per infective generation growth rate, once a stable distribution of infectives are reached with respect to age and time since infection. This value for the doubling time is understandably higher than a value that would be obtained by fitting an exponential curve to the first few prevalence data points.

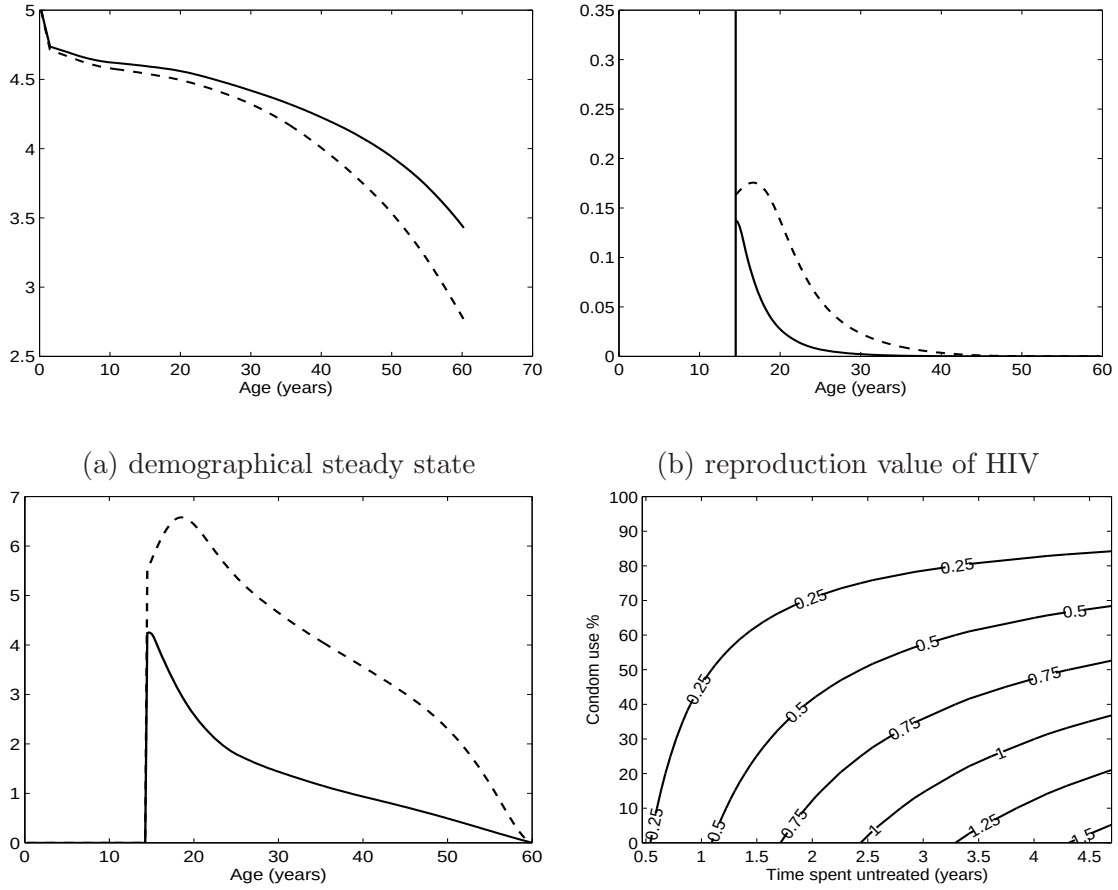
R_0 for HIV can be used to get an idea of how generalized testing followed by immediate ART enrolment could reduce HIV incidence at the population level. We compute and plot R_0 as a function of the rate at which HIV_+ individuals enrol for ART and the percentage of sexual contacts protected by condom use. Condom use is currently the HIV intervention with widest coverage in South Africa. On the vertical axis of Fig. 5.7(d) is the degree of condom use and on the horizontal axis is the time spent before an HIV_+ individual tests for HIV and enrols for ART. The time spent untreated is given by $1/\left(g^{\diamond}+\nu\right)$ years, where $g^{\diamond}$ is the rate at which HIV_+ individuals enroll for ART, and is varied in this calculation, while $\nu = \frac{1}{7}$ /year is the rate at which HIV_+ individuals die of HIV without treatment. This value for μ gives an average survival time of 7 years without ART, and is fixed in the simulation.

The contour line where $R_0 = 1$ shows the relationship between these interventions and the stability of the disease-free steady state. Note that without any ART program, i.e. where the horizontal axis tends to 7 years, 40% condom use would have kept $R_0 < 1$, averting a major HIV epidemic. Further, the R_0 contouring result of Fig. 5.7(d) suggests that a testing rate of once every 2.5 years together with low levels of condom use, would have averted a major HIV epidemic, had it been in place from the outset.

5.5.2 Increased TB detection rates

The basic reproduction number for TB only can be used to evaluate whether South Africa's implementation of the WHO DOTS strategy would have controlled TB, if there had been HIV. Using the formulae for the kernel of the next-generation operator for TB at the end of Sect. 5.2.2, and the parameter values summarized in Tab. 5.2, gives $R_0 = 1.5$. The left eigenfunction corresponding to this eigenvalue is shown in Fig. 5.8(a). The result suggests that active TB cases are making the largest contribution to the TB epidemic, and intervention should target these cases first.

R_0 for TB can be used to get an idea of how a DOTS program should function in terms of treatment success ratios and case detection rates, in order to control TB in a (now hypothetical) situation without HIV. We compute and plot in Fig. 5.8(b) R_0 as function of the percentage of cases successfully treated (vertical axis) and the average time spent with active TB (horizontal axis), excluding the period under treatment, given by $1/(\gamma + m + \beta)$ years. γ (detection rate) is varied in the simulation while m (TB mortality rate) and β (self-recovery rate) remain fixed.



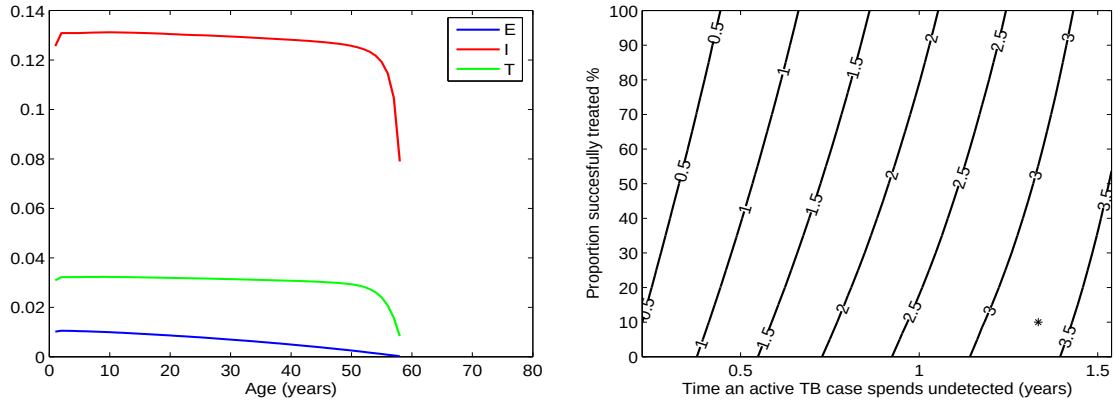
(a) demographical steady state

(b) reproduction value of HIV

(c) infections caused by age and gender

(d) impact of ART and condom use

Figure 5.7: (a) Distribution of women (solid) and men (dashed) at demographical steady state. The vertical axis denotes number of individuals. (b) The left eigenfunction (the reproduction value) corresponding to the largest eigenvalue (the basic reproduction number) for woman (solid) and men (dashed). (c) The would-be impact of UTTS acting from the start of the epidemic. (d) Number of secondary cases caused, disaggregated by age and gender – women (solid) and men (dashed).



(a) reproduction value of TB (b) impact of DOTS intervention

Figure 5.8: (a) The reproduction value for a model of TB infection. E denotes latent, I active and T on-treatment TB cases respectively. Active TB cases make the greatest contribution to the TB epidemic. (b) R_0 as a function of the effective contact rate κ and the time an active TB case spends undetected $1/(\gamma + m + \beta)$. (b)

5.5.3 The impact on TB of UTTS for HIV

We create a hypothetical situation in which TB invades a population where HIV has reached endemic steady state. However, the HIV-control parameter values (condom use and ART enrolment rate) as given by Tab. 5.2 leads to eventual eradication of HIV. Therefore, we construct a scenario where HIV reaches an endemic steady state of 24% (see Fig. 5.9(b)). To achieve this situation, we set the rate at which HIV_+ individuals test for HIV to 1/10 per year and the condom-use level constant at 10% for all age categories. The resulting age distributions of women susceptible to HIV (blue), infected with HIV (red) and on ART (green) are displayed in Fig. 5.9(a) in solid lines. The corresponding age distributions for susceptible men are also shown in Fig. 5.9(a), in dashed lines.

The reproduction value of a model where individuals are tested and immediately treated, once every ten years, on average, is shown in Fig. 5.9(c). Note that HIV_- individuals with active TB make the greatest contribution to the TB epidemic. The second largest contribution comes from HIV_+ individuals on ART with latent and active TB. We note that the reproduction value for HIV_- active TB cases decreases between the ages of 15 and 50 years. These are the ages when they are at risk of becoming HIV_+, which will reduce the contribution they make to the TB epidemic. Fig. 5.9(d) shows the impact that testing every year for HIV could have. The role of HIV_+ cases on ART with latent and active TB is reduced. A DOTS intervention in a situation of UTTS for HIV, will have the properties of the pre-HIV DOTS program. Isoniazid preventive therapy will have a greater impact if it is applied to HIV_+ individuals on ART, rather than applying it to HIV_- latent TB cases.

Fig. 5.9(e) shows R_0 as a function of the time an active TB case remain undetected in the community, with HIV_- cases on the horizontal axis and HIV_+ cases on the vertical axis. In this scenario HIV_+ individuals test every 10 years. Even though R_0 is more sensitive to

changes in the detection rate for HIV_+ cases, it is likely that detecting HIV_- cases will have a greater impact in terms of reducing the spread of MTB. The short duration of TB disease for HIV_+ cases offers little time for detection. In Fig. 5.9(f) HIV_+ individuals test every year under UTT. The impact on the TB disease is that the detection time of HIV_- cases can be extended while still having an impact in terms of reducing the spread of MTB.

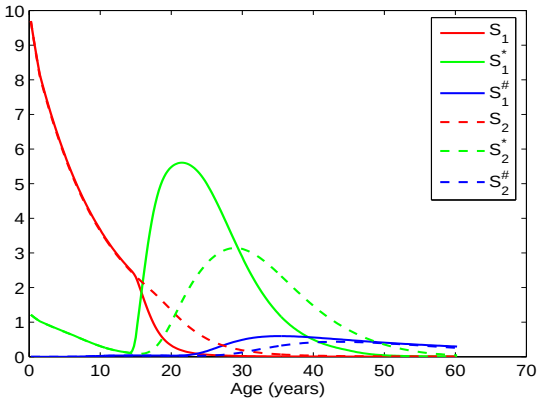
5.6 Conclusions

We developed an age-structured model to evaluate the potential impact of a proposed universal test and treat (UTTS) program on the course of the HIV epidemic in South Africa. Based on the knowledge that ART reduces viral load and thereby infectiousness of HIV_+ cases, the strategy aims to use ART not only as treatment but also as prevention tool to reduce the incidence of HIV. Our model indicates that even if all HIV_+ cases on ART are not at all infectious (an optimistic scenario to say the least) it will take some time to see the impact of the proposed intervention, as measured by prevalence, should it be deployed at a national level.

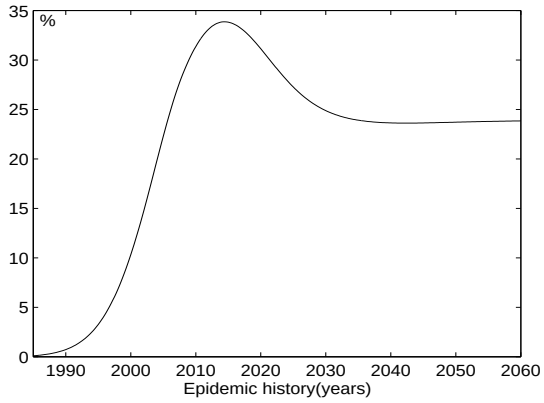
At this point we must give a word of warning about the limitations of our data and the effect on our modeling work. Our most complete data on HIV prevalence and sexual behavior come from women only and are collected at antenatal clinics. Our model is based on this data - it offers a picture of HIV as seen from the 'point of view' of women. Further, it is likely that if UTTS discussed above is approved, antenatal clinics will become key players in implementing the strategy. Our model was formulated to represent HIV statistics in South Africa at the national level. However, it is a simple matter to relate prevalence at a national level to prevalence data collected at antenatal clinics. For this purpose we use Eq. (5.17) which 'weighs' prevalence as a function of age and the age-dependent likelihood of attending antenatal clinics.

To evaluate the potential impact of UTTS, we studied its impact on HIV prevalence. In the absence of accurate direct measurement of incidence at a national level, we assume that HIV prevalence will be used to monitor the immediate post-UTTS epidemic. Our modeling result shows a drastic reduction in the crude death rate (Fig. 5.5b), returning within one decade to pre-HIV rates, should UTTS be implemented immediately. However, our model shows that there would not be a corresponding drastic reduction in HIV prevalence measured in antenatal clinics (Fig. 5.5a). It is likely that health authorities will redefine 'prevalence' should millions of HIV_+ individuals receive ART, in order to use prevalence as an effective measure to monitor a post-UTTS epidemic.

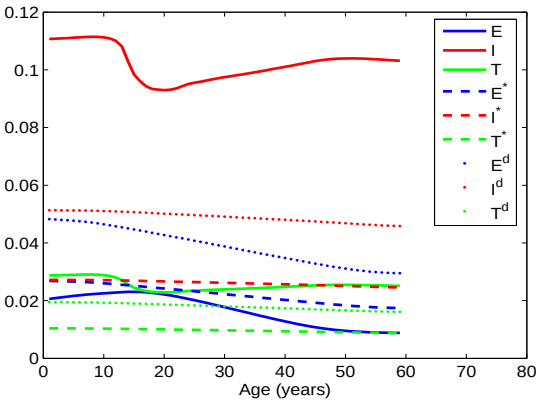
The HIV submodel can be used to evaluate the impact of condom use and circumcision among men, and ART for all, on the spread of HIV. A useful feature of the model is that it can compare an intervention which is applied at different rates to men and women respectively. The basic reproduction number for the population is equal to the square root of the product of the basic reproduction numbers of women and men. An intervention that



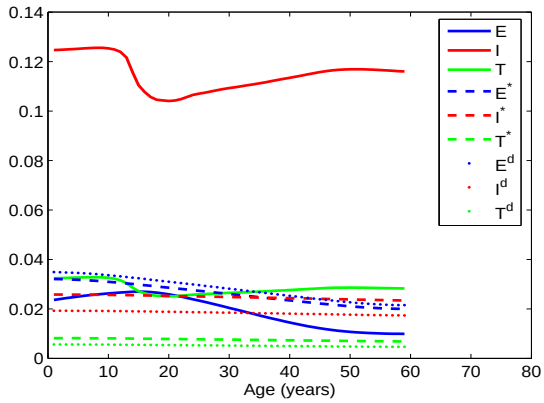
(a) age distribution in HIV steady state



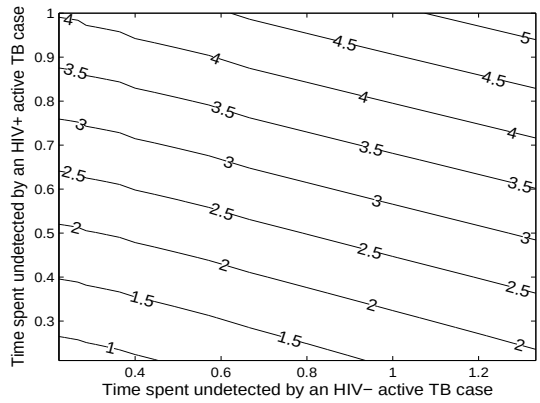
(b) HIV prevalence up to steady state



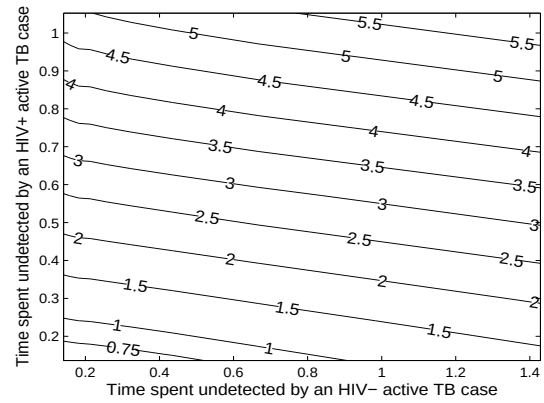
(c) RV of TB, $g^{\diamond} = 1/10$ per year



(d) RV of TB, $g^{\diamond} = 1$ per year



(e)



(f)

Figure 5.9: (a) Age distribution of HIV categories (HIV_- in blue, HIV_+ in red and those receiving treatment in blue, women subscript 1 and men subscript 2) at steady state. (b) A hypothetical situation where HIV reached a steady state of 24% before TB is introduced. (c) Reproduction value of TB classes, where $g^{\diamond} = 1/10$ per year. (d) RV of TB classes, where $g^{\diamond} = 1$ per year. (e) R_0 as a function of the detection rate of HIV$_-(I_k)$ and HIV$_+(I_k^*)$ active TB cases, where $g^{\diamond} = 1/10$ per year (f) R_0 as a function of the detection rate of HIV$_-(I_k)$ and HIV$_+(I_k^*)$ active TB cases, where $g^{\diamond} = 1$ per year.

would impact on only one of the two factors, such as male circumcision, requires a factor of 0.75 reduction to achieve a reduction of 0.5 in R_0 overall.

We also used the model to study HIV-TB dynamics at a national level, focusing on the impact of UTTS on HIV-TB epidemics. The extent of the modeled impact depends strongly on parameter choices for HIV-TB interaction. In particular, whether the TB epidemic will recover to what it was in the pre-HIV era will depend on how comparable TB disease is among HIV_- cases and HIV_+ cases on treatment. We assumed that the primary infection rate for HIV_+ cases on treatment is the same as that of HIV_- individuals. Under this assumption, UTTS can have a significant impact on the TB burden and mortality. Both would fall significantly, as shown in Fig. 5.5(e) and Fig. 5.5(f) respectively, back to levels of the pre-HIV era.

Chapter 6

The role of age separation in the spread of HIV

In order to explore how age separation between partners can influence the spread of HIV, we must consider interactions between demographical, partnering and HIV transmission dynamics. These interactions are subtle but crucially important. The fact that HIV has an incubation period of the same order of magnitude as the generation time of the population means that demographical processes can influence the transmission dynamics of HIV [49, 60, 44]. Conversely, HIV-mediated mortality will influence the demographical structure of a population. It is therefore necessary to consider HIV/AIDS within the framework of demographical models [49].

Understanding the nature of these relationships is key to investigating HIV transmission. The spread of HIV through heterosexual contact is the dominant route of HIV transmission in Sub-Saharan Africa [42, 99]. It has been recognized that pair formation and dissociation play a central role in the spread of HIV and different approaches have been used to model this dynamic [49, 59]. In models of serial monogamous relationships, the time spent either between relationships or within relationships minimizes the potential of infected cases to spread the disease. This effect can be significant when transmission probabilities are high, i.e. during the acute phase of HIV infection. Susceptible cases, on the other hand, are protected from disease if they are in a mutual monogamous relationship with another susceptible. Models which do not account for the fact that a fraction of the population is engaged in stable relationships will overestimate the transmission rate in a heterosexual population [49]. Therefore, our model explicitly includes relationship dynamics.

Data suggest that infectivity varies with time since infection. Measuring viral RNA levels after seroconversion shows that viral load peaks within the weeks following infection. If transmission probabilities are proportional to viral load, it implies high transmission probabilities during the acute phase of infection, followed by reduced transmission during the so-called asymptomatic phase. Thus far, time since infection has not been systematically studied in pair formation models, although its importance has been suggested [59] (and

references therein). To understand the impact of variable infectivity on R_0, it is necessary to structure population models by time since infection. We do not focus on its influence on transmission probabilities, but rather on its influence on a more fundamental aspect: survival.

Age separation between partners in developing countries seems to be higher than in developed countries [26], a fact which may help explain why HIV has spread much more quickly in the former. Recent behavioral studies show that age separation between partners may be associated with the persistence of the HIV epidemic in South Africa [63]. It could be argued that behavioral intervention could contribute more to reducing the spread of HIV if it discourages intergenerational sex. In this chapter we use a mathematical model to study the extent to which variance in age difference in partner choice is necessary to establish and maintain an epidemic.

It is believed that concurrent relationships played a crucial role in establishing a major HIV epidemic, and that they continue to play a significant role in the persistence of the epidemic [39]. The model we propose does not model concurrent relationships explicitly, but approximates the influence of concurrency on the sexual network structure, by allowing individuals to interact with 'pairs' of individuals in random contacts. We derive the NGM of this model and use it to investigate if concurrency influences the criticality of variance in the age difference between partners.

6.1 HIV spread in partnering models

Serial monogamy models

Standard SI models (as opposed to SIR models which track susceptible, infected and recovered individuals) make the implicit assumption that contacts are made randomly, but in a model with pair formation there are repeated contacts between the same individuals within a relationship. The time spent in a monogamous loyal relationship has a quarantining effect for susceptible couples. Furthermore, contacts restricted to relationships with one infective (i.e. discordant relationship) or two (i.e. both partners are infected) offer a degree of protection to the community. It is useful to investigate this effect within the NGM formalism. However, including relationship dynamics is rather difficult as one must consider all possible relationship histories. For example, the protective effect (for the community), is greatest when an infective is in a steady relationship during the initial (and very infectious) stage of the infectious period.

We start from a straightforward extension of the model in appendix C, where an age-structured partnering model is studied, introducing an extra level of indices for disease status (0-susceptible, 1-infectious). The subscript k is used to indicate gender: $k = 1$ for women and $k = 2$ for men. Let $p_{yk}(t, a)$ be the number of young women and men aged a, not yet eligible for relationships. Let $p_{0k}(t, a)$ be the number of eligibles aged a who are susceptible to infection and $p_{1k}(t, a)$ the number of eligibles who are already infected. Susceptible indi-

viduals in relationships are denoted by $p_{c,00}(t,a,b)$ representing the number of relationships between a susceptible woman aged a and a susceptible man aged b, $p_{c,10}(t,a,b)$ representing infected women in a relationship with susceptible men, $p_{c,01}(t,a,b)$ representing susceptible women in a relationship with susceptible men and $p_{c,11}(t,a,b)$ representing couples where both partners are infected. The mortality rate for individuals aged a is modeled by $\mu_k(a)$. Disease related mortality rate is given by $\mu_{dk}(a)$. The rate at which young individuals aged a mature into becoming eligible for relationships is given by $\eta_k(a)$. The rate at which eligible woman aged a meet eligible men aged b is given by $\kappa(a,b)$. Individuals are born at a constant rate of B.

The model is an SI-type model structured by age and gender and tracks relationships explicitly. Other forms of heterogeneity, e.g. in sexual activity patterns and other types of behaviorial variation, are not explicitly modelled as these are sufficiently captured through age dependencies (see Fig. 6.1). With these notations, the equations for the model are:

- For young men and women:

$$\begin{aligned}\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right)p_{yk}(t,a) &= -\mu_k(a)\,p_{yk}(t,a)-\eta_k(a)\,p_{ky}(t,a)\,,\\ p_{yk}(t,0) &= \frac{B}{2}\,.\end{aligned}$$

- For susceptible men and women:

$$\begin{aligned}\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right)p_{0k}(t,a) &= -\mu_k(a)\,p_{0k}(t,a)+\eta_k(a)\,p_{yk}(t,a)\\ &\quad+\int_0^\infty p_{c,00}(t,a,b)\big[\sigma(a,b)+\mu_{k^*}(b)\big]\,db\\ &\quad+\int_0^\infty p_{c,k^\diamond}(t,a,b')\big[\sigma(a,b)+\mu_{dk^*}(b)\big]\,db\\ &\quad-\int_0^\infty \Psi(t,a,b,p_{01},p_{02})\,db-\int_0^\infty \Psi\left(t,a,b,p_{k'},p_{k''}\right)\,db\\ &\quad-\phi_k(a)\,p_{0,k}(t,a)\,,\end{aligned}$$

  where $k^*=2, k^\diamond=01, k'=01, k''=12$ *if* $k=1$ and $k^*=1, k^\diamond=10, k'=11, k''=02$ *if* $k=2$.

- For infected men and women:

$$\begin{aligned}\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right)p_{1k}(t,a) &= -\mu_{dk}(a)\,p_{1k}(t,a)\\ &\quad+\int_0^\infty p_{c,11}(t,a,b')\big[\sigma(a,b)+\mu_{dk^*}(b)\big]\,db\\ &\quad+\int_0^\infty p_{c,k^\diamond}(t,a,b')\big[\sigma(a,b)+\mu_{dk^*}(b)\big]\,db\\ &\quad-\int_0^\infty \Psi(t,a,b,p_{11},p_{12})\,db-\int_0^\infty \Psi\left(t,a,b,p_{k'},p_{k''}\right)\,db\\ &\quad+\phi_k(a)\,p_{0,k}(t,a)\,,\end{aligned}$$

where $k^*=2, k^\diamond=10, k'=11, k''=02$ if $k=1$ and $k^*=1, k^\diamond=01, k'=01, k''=12$ if $k=2$.

- For couples:

$$\begin{aligned}(D_{t,a,b})\,p_{c,00}(t,a,b) &= -\big[\sigma(a,b)+\mu_1(a)+\mu_2(b)\big]\,p_{c,00}(t,a,b)\\ &\quad+\Psi(t,a,b,p_{01},p_{02})\\ (D_{t,a,b})\,p_{c,01}(t,a,b) &= -\big[\sigma(a,b)+\mu_1(a)+\mu_{d2}(b)\big]\,p_{c,01}(t,a,b)\\ &\quad-\vartheta_1(a,b)p_{c,01}(t,a,b)+\Psi(t,a,b,p_{01},p_{12})\\ (D_{t,a,b,})\,p_{c,10}(t,a,b) &= -\big[\sigma(a,b)+\mu_{d1}(a)+\mu_2(b)\big]\,p_{c,10}(t,a,b)\\ &\quad-\vartheta_2(a,b)p_{10}(t,a,b)+\Psi(t,a,b,p_{11},p_{02})\\ (D_{t,a,b})\,p_{c,11}(t,a,b) &= -\big[\sigma(a,b)+\mu_{d1}(a)+\mu_{d2}(b)\big]\,p_{c,11}(t,a,b)\\ &\quad+\vartheta_1(a,b)p_{c,01}(t,a,b)+\vartheta_2(a,b)p_{c,10}(t,a,b)\\ &\quad+\Psi(t,a,b,p_{11},p_{12})\,,\end{aligned}$$

where $D_{t,a,b}=\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}+\frac{\partial}{\partial b}\right)$.

- The same function Ψ is used for pairing eligible men and women over all disease-status combinations:

$$
\begin{aligned}
\Psi(p_{02}, p_{01})(a,b) &= \frac{\kappa(a,b)}{\Omega_e(t)}\big[p_{02}(t,a)p_{01}(t,b)\big] \\
\Psi(p_{02}, p_{11})(a,b) &= \frac{\kappa(a,b)}{\Omega_e(t)}\big[p_{02}(t,a)p_{11}(t,b)\big] \\
\Psi(p_{12}, p_{01})(a,b) &= \frac{\kappa(a,b)}{\Omega_e(t)}\big[p_{12}(t,a)p_{01}(t,b)\big] \\
\Psi(p_{12}, p_{11}(a,b) &= \frac{\kappa(a,b)}{\Omega_e(t)}\big[p_{12}(t,a)p_{11}(t,b)\big] \\
\Omega_e(t) &= \sum_{k=1,2}\int_0^\infty \big[p_{0k}(t,a) + p_{1k}(t,a)\big]\,da\,. \qquad (6.1)
\end{aligned}
$$

There are a number of caveats concerning the model and remarks regarding notation:

- Disease duration (i.e. the incubation period) is a crucial detail in realistic models for studying certain aspects of HIV epidemiology, such as disease-related mortality and the influence of progression on disease spread. Adding this level of realism necessitates the use of numerical methods to study disease dynamics.

  The focus of our model is to track couples where one or both partners can be infected. Adding a variable to take account of a distribution over disease duration adds an extra dimension to objects (usually a matrix) used in the simulation. The object becomes too large when the many age categories (e.g. 1–80 age categories) and many years of disease duration (e.g. 30 years) are modeled.

  With age structure being our main focus, we model disease-related mortality as independent of disease duration, by adding a constant rate to the disease free mortality rate, see Sect. 6.1.1. This amounts to using an exponential survival function for disease-related related mortality.

- We make the simplifying assumption of a constant birth rate, which together with mortality can yield any desired age structure for the population. The population is studied at demographical steady state, which makes this assumption less restrictive [59].

  In an STD model what really matters is the rate at which new susceptibles enter the population. Knowledge of immigrants can be combined with fertility data to suggest a correct 'birth rate' although data on their disease status are rare, especially in the demographical situations we consider. If some of the immigrants are infectious, they will introduce a time-dependent infection intensity into the community which will be very difficult to disentangle from age-dependent effects, given the scarcity of data [20, Ch. 8]. For this reason we do not model immigration.

- The rate at which relationships form is determined by a harmonic mean function Ψ, allowing relationships to form between all ages and disease (S,I=0,1)

categories. The harmonic mean function to model individuals forming pairs in steady relationships has its origin in marriage models, from demography (see appendix C). To keep things as simple as possible, we do not let the intensity of relationship formation, $\kappa(a, b)$, depend on disease duration.

– A force of infection due to random casual contacts, is modeled by ϕ_1 and ϕ_2 for eligible women and men respectively. The functional form used for this force of infection is explained in Sect. 6.1.1.

Table 6.1 is a diagrammatic presentation of the overall dynamics. Individuals enter as 'young', after which they 'mature' into eligibility. From here they may undergo a sequence of eligible-relationship-separate-eligible cycles. The red arrow shows which categories eligibles can infect, while the blue line shows which categories are involved in steady-relationship infection.

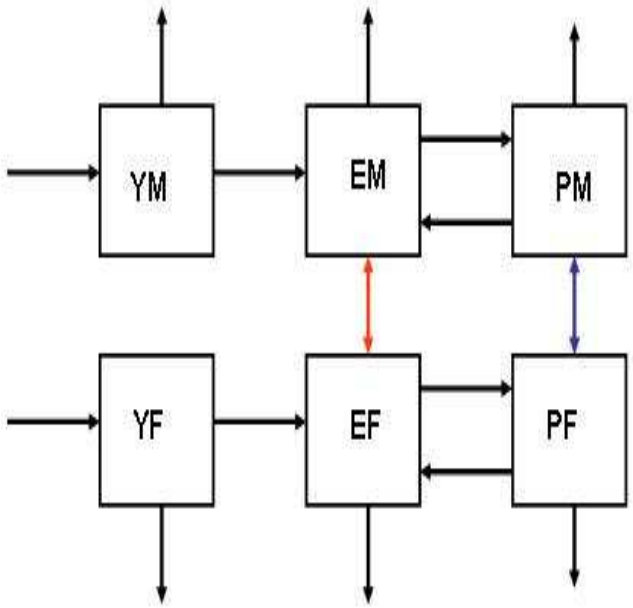


Figure 6.1: A diagrammatic presentation of the partnering dynamics. YM-young men, YF-young women, EM-eligible men, EF- eligible women, PM-partnered men, PF-partnered women

6.1.1 Parameter values

The model of HIV dynamics is based on that of a hypothetical community. This community is set up so as to give a realistic picture of HIV dynamics in a typical South African township, as discussed in Sect. 4.2. Results specific to a particular township community, like Masiphumelele, can be obtained by sampling from the hypothetical distributions in the

model. The model does not capture the role of commercial sex workers (CSW) or concurrent partners. Both of these processes will make age separation between partners less of a critical factor in HIV becoming a major epidemic.

Demographical parameters

Natural mortality. We assume that the survival curve (without HIV and TB) is given by the two-parameter function $S(a) = \exp(-v_1 a^{v_2})$ with $v_1 = 10.7$ and $v_2 = 4$, with a (age) given in years. The natural mortality is then defined by $m(a) = -S'(a)/S(a)$. This gives a life expectation of approximately 50 years in the absence of HIV, and the corresponding age distribution is shown in Fig. 6.2(a). We use the same survival function for men and women. The age distribution of the particular township community (i.e. Masiphumelele), according to a 2006 demographical survey, is shown in Fig. 6.2(b).

Birth rate. To avoid clouding the issue with the complexities of the relationship between relationship dynamics and a birth rate, we make the simplifying assumption that the birth rate is constant at 200 newborns per year. As in [59], we study how HIV 'invades' a population in demographical steady state, the state in which R_0 is calculated, and a realistic steady-state situation can be achieved with a constant birth rate. We use a birth fraction of 0.5 for men and women, ignoring the fact that slightly more men are born than women. The model does not include mother-to-child transmission, which is an important route of intergenerational transmission of HIV in South Africa.

Sexual debut. We use a standard curve from applied demography to model the process of maturing into eligibility for relationships. The curve depicted in Fig. C.1(b) is the cumulative distribution of the extreme value distribution $\eta(a) = 0.17e^{-4.411e^{0.309a}}$, where a is the age of sexual debut [57, p. 48]. We use the values $a = 13$ and $a = 16$ for the first age of eligibility for women and men respectively.

In demographic studies this curve was originally used to model the age of eligibility for a first marriage, and where men are generally modeled as being eligible for first marriage at a younger age than women. We will use this function to model the process of maturing into participation in sexual relations. For Sub-Saharan Africa, it is known that the age of first participation for women is generally younger than for men.

Parameters related to relationship dynamics

Relationship formation. To model the 'waiting time' between being eligible for a relationship and participating in one, we model $\kappa(a, b)$ as $f(a, b) * 1/w(a, b)$ where $w(a, b)$ is the average waiting time between relationships and $f(a, b)$ is the distribution of partners. We use a constant $w(a, b) = w$ to adjust the intensity of relationship formation.

An ongoing behavioral study in Guguletu [62], another township near Cape Town, recorded the age difference between 600 couples. Fig. 6.2(e) is a scatter plot which shows

large variance in the age difference between men and women. In studies of heterosexual couples, men are mostly older than women. Fig. 6.2(e) shows that a normal distribution with a mean of -3 years and a standard deviation of roughly 5.5 years^2, fits the data reasonably well. In this chapter we use a normal distribution with a mean of 0, as in Sect. 4.2 (see Fig. 4.2(d)) and vary the standard deviation of this distrinution - it is the variance in age difference between partners which is critical, not the mean of their age difference.

To model partner preference we use a normal distribution with a mean of 0 and we adjust the standard deviation in different scenarios. We make the assumption that partner preference is normally and symmetrically distributed, while noting that in reality men generally choose younger women as partners. Our aim is to study the impact of the age difference in relationships on establishing and maintaining HIV, and a symmetric partner choice function is adequate for this purpose.

Relationship duration. The average duration of relationships is approximately $\frac{1}{\sigma}$. The deviation from an exact exponential is due to the mortality of one of the partners within the relationship. For the simulation depicted in Fig. 6.3 we used a value of $\sigma = 1/5$ per year.

Parameters related to HIV

Fraction of individuals not at risk of contracting HIV. A feature of most epidemic models is that simulated prevalence is too high. To remedy this we introduce a category of individuals who are never at risk of contracting HIV. Another way of reducing the simulated prevalence of HIV would be to adjust the risk of contracting HIV according to changes in prevalence (see Sect. 2.1) or to increasing HIV-related mortality. The idea is that a high prevalence of HIV or HIV-associated mortality will result in behavior change, when people notice the burden of the epidemic. Both approaches are merely devices to address the artificially high prevalence produced by compartmental models, and to fit prevalence data retrospectively.

We use the former approach in this chapter, directing a fraction of newborns into a category where they will not be at risk of HIV infection. This fraction is adjusted to achieve the desired prevalence curve. We note however that the device is more questionable in this setting, where we are interested in the details of HIV dynamics, than in a setting where the influence of HIV on the TB epidemic is studied.

Transmission within discordant relationships. We define $\vartheta_k(a, b)$ to be the rate of infection within a discordant relationship where the woman is aged b and the man is aged a. In principle this rate depends on many factors: on whether the infective in the relationship is a woman or a man, on the duration of his/her infectious period, on the duration of the relationship, frequency of sexual contacts within the relationship, and many more.

There is evidence that HIV infectiousness is high during the first few weeks of infection and much lower subsequently. To keep things simple we use an average infectiousness. If the

average duration of an infectious period is T_i, and the infectiousness as a time since infection is $\vartheta_k(a,b,\tau)$, then the time-average of the infectiousness during an infectious period is [59]:

$$\vartheta_k(a,b) = \frac{1}{T_i}\int_0^{T_i} \vartheta_m(a,b,\tau)\,d\tau\,.$$

In the simulations below, we assume the average time of escaping infection within a discordant relationship is constant and is approximately 2 years: $\vartheta_k = 1/2$ per year. In practice, this rate is thought to be different for male-to-female and female-to-male transmission, but for simplicity we use the same rate in our simulations. Adjustment of this value must take relationship dissociation rate into account (see relationship duration above), and this rate should be appreciably lower than the infection rate within discordant relationships.

Risk of contracting HIV outside of steady relationships. An infection risk outside of (steady) relationships can be introduced by 'allowing' susceptibles and infectives to make random contacts. The force of infection for a man aged a can be modeled as [60],[13, Ch. 11.3,13.3],[12]:

$$\phi_2(t,a) \quad = \quad \beta(a)\,\frac{\int_0^\infty \beta(b)\,f(a,b)\,p_{11}(t,b)\,db}{\int_0^\infty \beta(b')f(a,b')\big[p_{01}(t,b)+p_{11}(t,b')\big]\,db'}\,, \tag{6.2}$$

where $\beta(a)$ is a measure of the risk of contracting HIV at age a, $b \mapsto f(a,b)$ is the age distribution of partners of a person aged a. It is an effective rate which includes a partner turnover rate and transmission probability per casual relationship. An age-dependent partner preference is built into $f(a,b)$. This preference for casual partners could in reality be different to the age-dependent preference for steady relationships, although in this discussion we assumed they are identical.

The denominator in Eq. (6.2) accounts for proportionate mixing:

$$\frac{\beta(a)\,\beta(b)\,f(a,b)\,p_{11}(t,b)}{\int_0^\infty \beta(b')\,f(a,b')\,p_{11}(t,b')\,db'}\,, \tag{6.3}$$

for the rate at which men aged a meet women aged b during risky contacts. This, multiplied with the probability that a random encounter is with an infective, $p_{11}(t,b)/\big[p_{01}(t,b)+p_{11}(t,b)\big]$, gives the rate at which susceptible men aged a make infectious contacts with infective woman aged b. Integrating this over the ages of all women gives Eq. (6.2) for the force of infection experienced by men aged a [56].

Fig. 6.2(c) depicts the following risk function $\beta(a)$:

$$\beta(a) = \begin{cases} 0 & \text{if } a < 15 \\ 5\times 10^{-5}\times 12^2\times(a-15)^2 e^{-0.24(a-15)} & \text{if } a \geq 15 \end{cases}\,,$$

A time dependence could be built into this function to balance demand and availability of partners [44]. Risk function $\beta(a)$ is an effective contact rate, and includes a transmission probability per contact. In this case we assume that the transmission probability is independent of age, and we absorb it in $\beta(a)$.

Our formulation assumes that one contact is made with each casual partner, and leads us to define a casual relationship as a single union initiated by a sexual contact. On the other hand, a steady relationship allows for repeated contacts between partners.

HIV survival. We assume an average of 10 years of survival with HIV (Fig 6.2(d)). To avoid complicating the model, we assume that

$$\mu_{dk} = \mu_k + 1/10 \text{ per year}\,.$$

6.2 Simulation results

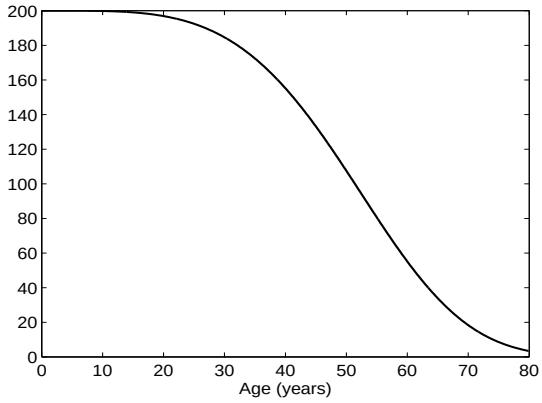
We use a deterministic finite difference scheme, updated in yearly intervals, to simulate population dynamics. Different types of dynamics were introduced in the following sequence: basic demography, relationship dynamics, followed by HIV infection.

Setting up the model. The population is first put into a demographical steady state consistent with the prevailing demographic rates, as set out in Sect. 6.1.1. The timescale associated with this steady state is of the order of a few lifetimes.

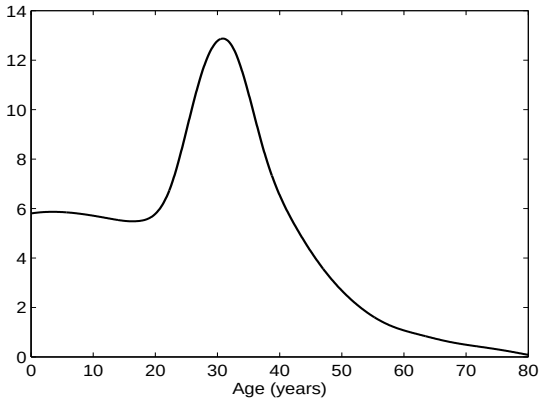
Relationship dynamics are then introduced and the model is put into a new steady state consistent with rates associated with relationship dynamics, as set out in Sect. 6.1.1. The timescale associated with this steady state is of the order of the duration of a relationship. The partner separation rate is assumed to be constant and large giving relationships which are short in duration.

At this point the epidemic is introduced by adding an infected individual into the (eligible) population, and run until a new steady state (in the sense of stable age distributions) is reached. The probability of establishing a major epidemic depends sensitively on the denominator used in the partnering function, the age category into which the disease is introduced, the average waiting time to enter steady relationships, and the level of risk allowed for casual sex.

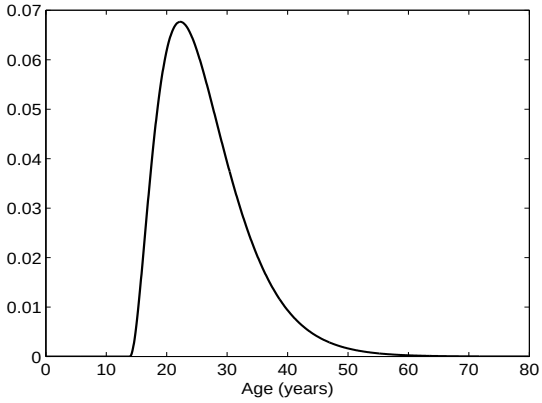
It is not possible to obtain a persistent epidemic in a model with serial monogamy only (i.e. without any random contacts), when the denominator of the partnering function includes all eligibles, as in Eq. (6.1). As depicted in Fig. C.1(d) in appendix B, this choice of denominator results in long waiting times between relationships. This, together with serial monogamy, offers a substantial 'protective effect' and the epidemic dies out. Relaxing partner preference (by increasing the variance of age separation in relationships) and increasing



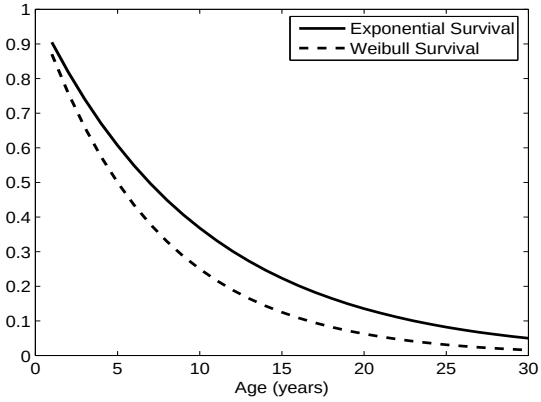
(a) age distribution of hypothetical population



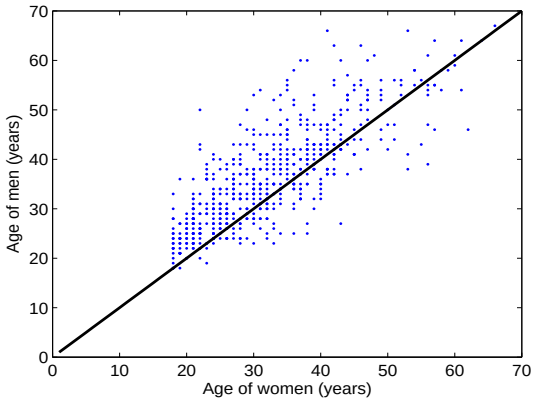
(b) age distribution of Masiphumelele



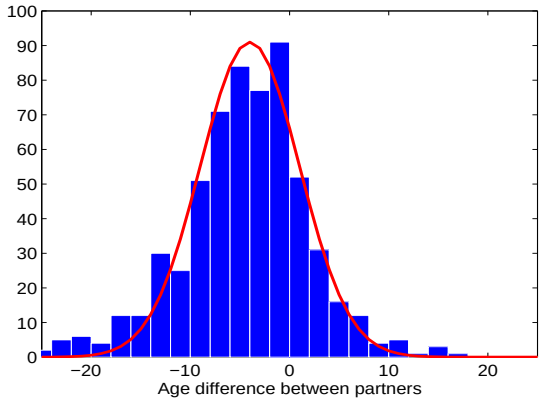
(c) risk of contracting HIV as a function of age



(d) survival function for HIV_+ individuals



(e) age difference between partners in Guguleto.



(f) distribution of age difference

Figure 6.2: Age-dependent parameters in the age-structured partnering model. (e) displays the ages of 1200 (600 couples) of men and women who were interviewed in Guguletu [62]. The solid line depicts equal ages for men and women in a relationship. (e) approximates the distribution of age differences between partners as Gaussian.

relationship dissociation rates, both help extend the duration of the epidemic, but it still dies out.

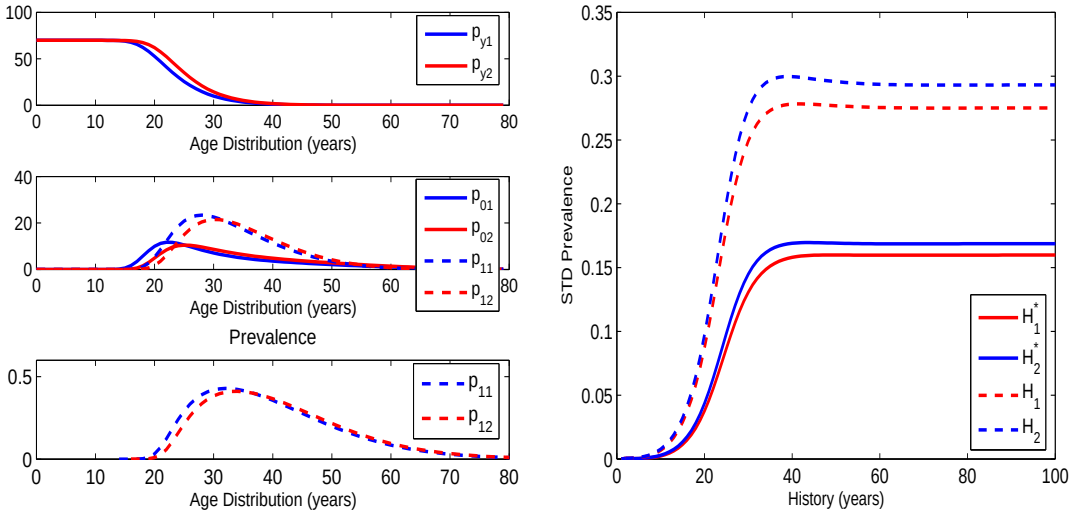
One way of establishing a major epidemic is to reduce the waiting time between relationships by applying a weighting function to the eligibles in the denominator. A possible age-dependent measure is depicted in Fig. 6.2(c). Numerical simulation confirms that a major epidemic can be established when a appropriately weighted denominator is used in the partnering function (6.1). The results are not realistic and are omitted.

A more realistic way of establishing a major epidemic, which adds interesting complexity to the model, is to include a mechanism for risky behavior. We used the simple risk function as depicted in Fig. 6.2(c). Various reasonable functions may be proposed, but it is not easy to decide which one to use in this particular model.

Stable age structure. Fig. 6.3(a) displays the stable age distributions obtained by running the model for a few hundred years. The top figure displays the age distribution of eligibles, with women making an earlier sexual debut than men. The middle figure shows the age distribution of susceptibles and infective eligibles. Note the shift to the left of the age distributions of infected women, as well as the higher peak in the age distribution of infected women. These are both the result of women making an earlier sexual debut. The result that female-to-male prevalence ratios can be created in a model without differential susceptibility between males and females was previously mentioned in [50]. The difference between the peaks of the age distribution of infected men and women can be enhanced by making women face a higher age-dependent risk of contracting HIV (see Fig. 6.2(c)).

Equal risk for men and women. In Fig. 6.3(a,bottom) we show the prevalence of disease as a function of age and gender. Note that the prevalence of disease is higher for women and peaks at a lower age than for men. In this simulation we apply the same risk function for contracting HIV to men and women. The shift between the prevalence of men and women is the result of women making their sexual debut at a younger age, as depicted in Fig. 6.3(a,top). It is not necessary to model a higher risk of infection for women to obtain this result. Increasing the risk women have of contracting HIV enhances the shift in prevalence as well as enhancing the difference in the peak.

Sampling a general population. To relate the simulation results to data from a real community, like Masiphumelele, we sample from the simulation results from the hypothetical population using the age distribution of Masiphumelele, which is depicted in Fig. 6.2(b). Fig. 6.3(b) shows the prevalence trend produced by the model (in red). Viewing the township as a sample from this wider community gives the prevalence trend pertaining to the township (in blue). This procedure can be motivated on the basis of the township not being closed, mixing also within the wider community. This sampling procedure is more effective than extrapolating results obtained by modeling the unusual township to a wider population.



(a) steady state age distributions (b) HIV prevalence in two communities

Figure 6.3: Simulation of an HIV epidemic. (a) Steady-state age distributions. (b) Prevalence trends. H_k (solid line) is the prevalence of women ($k = 1$) and men ($k = 2$) in a hypothetical population. The same data in Masiphumelele is shown in the dashed line.

6.3 Age separation in partner choice and a threshold condition for HIV invasion

Age separation in relationships plays a critical role in the spread and persistence of sexually transmitted diseases [118, p.562]. It is possible to study the role of age separation within relationships in a linearized version of a two-sex model that includes age and time since infection. Using the results from appendix B.2, we first build a next-generation matrix (NGM) for an age-structured epidemic model, and then study its spectral radius, R_0, as we vary the age separation between partners in our model (by varying the standard deviation of an assumed normally distributed age difference between partners). A helpful example can be found in [59], where the basic reproductive number in pair formation models was studied. In this section we extend this method to model infection due to random contacts and contacts within relationships. We also explore how effectively our model describes and fits reality, a concern which was not evident in [59]. In an another paper [58], the same author studied an iterative marriage model, modeling state transitions as a semi-Markov process. We use a similar idea to model individuals participating in sequential relationships.

Consider the following definition for the next-generation operator (see appendix B.2 and Sect. 4.4):

$$(\mathbb{K}\varphi)(a) = \int_{D(\alpha)} K(a,\alpha)\,\varphi(\alpha)\,d\alpha\,,$$

where $\varphi(\alpha)$ represents an infective distributed over some state with domain $D(\alpha)$, by a

distribution φ. We identify six infectious states for each age group, three each for woman and men: eligible and not in relationships, in a relationship with a susceptible of any age, and being in a relationship with a susceptible of a particular age. The kernel K of the operator $\mathbb{K}$ can be written as:

$$K(a,\alpha) = \int_0^\infty \mathcal{A}(a,\theta,\tau)\,\mathcal{C}(\theta,\alpha,\tau)\,d\tau\,,$$

where $\mathcal{A}(a,\theta,\tau)$ models infectiousness towards susceptibles in state a of an infective in state θ, τ years into the infectious period. $\mathcal{C}(\theta,\alpha,\tau)$ is the conditional probability that the infective is in state θ at time τ, given that he was in state α when his infectious period started after he was introduced into a population in a disease-free steady state. $\mathcal{C}(\theta,\alpha,\tau)$ is defined by the following equation:

$$\frac{d}{d\tau}\mathcal{C}(\theta,\alpha,\tau) = \mathcal{B}(\theta,\alpha,\tau)\,\mathcal{C}(\theta,\alpha,\tau),\ \mathcal{C}(\theta,\alpha,0) = I\,, \tag{6.4}$$

where $\mathcal{B}(\theta,\alpha,\tau)$ is a time-dependent transition matrix acting on a state probability with an initial value $\mathcal{C}(0) = I$, with I being the identity matrix. The transition rates in $\mathcal{B}$ are duration (of state)dependent, and the individuals transitions are modeled as a semi-Markov process [58]. Since $\mathcal{A}(a,\theta,\tau)$ represents the number of infections made during a unit of time since infection and $T(\theta,\alpha) = \int_0^\infty \mathcal{C}(\theta,\alpha,\tau)\,d\tau$ is the total time spent in state θ, it is clear that Eq. (6.4) accounts for the total number of infections during an infectious period, which started in state α [16].

In this formulation, state refers to gender, being eligible for or partnered in relationships, and age: a discrete and finite set of states. Using this approach, the next-generation operator $\mathbb{K}$ is reformulated as a NGM. The spectral radius (R_0) of the kernel of this operator, the matrix K, determines the stability properties of this system: when $R_0 > 1$ the disease-free steady state is unstable, and a major epidemic will occur. We did not perform a mathematical study of the stability properties of this system (demography, random sexual mixing, and steady relationship) but point out that Inaba [60] addresses the stability of the endemic and the disease free steady states in a similar situation.

Consider first the matrix $\mathcal{B}(i,j,\tau)$, governing transitions between being partnered and eligible, for an infectious period starting at state j at time $\tau = 0$. $\mathcal{B}(i,j,\tau)$ is a 6×6 matrix for an infective of a given age and has two non-zero 3×3 blocks: one in the upper left section and one in the lower right section of the matrix. These sub-matrices are denoted by $\mathcal{B}_1(i,j,\tau)$ and $\mathcal{B}_2(i,j,\tau)$ respectively. They model transitions of an infective between eligible and partnered states, for an infectious period starting at age j which act toward susceptible partners aged i. More formally:

$$\mathcal{B}(i,j,\tau) \;=\; \begin{pmatrix} \mathcal{B}_1(i,j,\tau) & 0 \\ 0 & \mathcal{B}_2(i,j,\tau) \end{pmatrix},$$

where

$$\mathcal{B}_1(i,j,\tau) = \begin{pmatrix} -\mu_{d1}(j+\tau) - \mathcal{R}_1(j+\tau) & \bar{\mu}_{d2} + \sigma & \bar{\mu}_{d2} + \sigma \\ \mathcal{R}_1(j+\tau) & -\mu_{d1}(j+\tau) - \bar{\mu}_{d2} - \sigma - \vartheta & 0 \\ 0 & \vartheta & -\mu_{d1}(j+\tau) - \bar{\mu}_{d2} - \sigma \end{pmatrix}.$$

We note that:

- The age variable is discretized. In the simulations which follow we distinguish $a = 1:80$ age categories.
- $\mathcal{B}_1(i,j,\tau)$ represents how the transition matrix depends on men aged i, τ years after she became infected at age j. Transitions from one state to another are captured by off-diagonal elements of the matrix $\mathcal{B}_1(i,j,\tau)$.
- $\mu_1(i+\tau)$ and $\mu_{d1}(i+\tau)$ denote disease-free and disease-related mortality, respectively. We assume that the survival probability includes a factor $e^{-\int_j^{j+\tau}\mu_1(a)\,d\tau}$ for natural mortality and a disease-related factor $e^{-\int_j^{j+\tau}\mu_d(a)\,d\tau}\; e^{-\mu_d\,\tau}$. The simplest assumption to make is that disease related mortality is independent of disease duration. This means that the infective faces disease-related mortality immediately after the infectious periods starts. This gives a slightly smaller R_0 than using a Weibull survival functions gives (see Fig.6.2(d)).
- $\mathcal{B}_1(i,j,\tau)_{1,1}$, representing position $(1,1)$ of the transition matrix $\mathcal{B}_1(i,j,\tau)$, is the rate at which a woman leaves the state of being eligible. This can happen due to mortality $\mu_{d1}(j+\tau)$ or due to

  $$\mathcal{R}_1(j+\tau) = \frac{1}{\Omega_e}\sum_v \kappa_1(v, j+\tau)\,\hat{p}_{02}(t,k)\,,$$

  which is the rate at which she may enter a relationship with a susceptible man of any age, where Ω_e is the size of the total eligible pool, as given by Eq. (6.1). $\sum_v$ is a sum over all age categories. This model tracks continuous distributions of individuals. In particular, a fraction of the infected woman is partnered with each susceptible man. The only way to track each individual in each relationship individually, is by means of individual based models, which is the subject of Ch. 7.
- $\mathcal{B}_1(i,j,\tau)_{2,2}$ is the rate at which a woman aged $(j+\tau)$ may leave a steady relationship

with a susceptible partner of any age: she faces mortality $\mu_f(j+\tau)$, and a relationship disassociation rate σ, while the man faces an average mortality of:

$$\bar{\mu}_{d2} = \sum_v \mu_{d2}(v)\, \mathbb{P}_1(v, j+\tau) \tag{6.5}$$

where $\mathbb{P}_1(v, j+\tau)$ is the probability that a woman aged $(j+\tau)$ is partnered with man aged v. This probability can be computed from the relationship rate at the disease free steady state. The probability of finding an infected woman aged $(j+\tau)$ in a relationship with a susceptible man, after a partnering event, is given by:

$$\frac{\Psi\big(t, p_{02}(i), p_{11}(j+\tau)\big)}{\sum_v \Psi\big(t, p_{02}(v), p_{11}(j+\tau)\big)}\,. \tag{6.6}$$

However, this level of detail will not matter if the relationship dissociation rate σ is much larger than the mortality rate, in which case σ will dominate the relationship dissociation rate.

She enters the state of being in relationship with an infective at rate ϑ, which is the rate at which she infects her uninfected steady partner.

- $\mathcal{B}_1(i,j,\tau)_{2,2}$ is the rate at which a woman aged $(j+\tau)$ may leave a steady relationship with an infected partner.
- Men undergo a similar sequence of transitions. Transitions between males and females are not possible, explaining the blocks of zero in the matrix:

$$\mathcal{B}_2(i,j,\tau) = \begin{pmatrix} -\mu_{d2}(j+\tau) - \mathcal{R}_2(j+\tau) & \bar{\mu}_1 + \sigma & \bar{\mu}_2 + \sigma \\ \mathcal{R}_2(j+\tau) & -\mu_{d2}(j+\tau) - \bar{\mu}_1 - \sigma - \vartheta & 0 \\ 0 & \vartheta & -\mu_{d2}(j+\tau) - \bar{\mu}_1 - \sigma \end{pmatrix}.$$

The matrix $\mathcal{A}(i,j,\tau)$, of infection rates, is given by:

$$\mathcal{A}(i,j,\tau) = \begin{pmatrix} 0 & 0 & 0 & A_{1,4} & 0 & 0 \\ 0 & 0 & 0 & 0 & A_{2,5} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ A_{4,1} & 0 & 0 & 0 & 0 & 0 \\ 0 & A_{5,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{6.7}$$

where:

$$
\begin{aligned}
A_{1,4} &= \frac{\beta(i)}{C_1(i)}\,\beta(j+\tau)\,f(i,j+\tau)\,\hat{p}_{01}(t,i)\,,\\
A_{2,5} &= \vartheta\left[\kappa_1(i,j+\tau)\,\hat{p}_{02}(t,k)\right],\\
A_{4,1} &= \frac{\beta(i)}{C_2(i)}\,\beta(i)\,\beta(j+\tau)\,f(i,j+\tau)\,\hat{p}_{02}(t,i)\\
A_{5,2} &= \vartheta\left[\kappa_1(j+\tau,i)\,\hat{p}_{01}(t,k)\right],
\end{aligned}
\tag{6.8}
$$

where the normalization factors in $A(1,4)$ and $A(4,1)$ are given by

$$
\begin{aligned}
C_1(i) &= \sum_v \beta(v)\,f(i,v)\,\hat{p}_{01}(t,v)\,,\\
C_2(i) &= \sum_v \beta(v)\,f(i,v)\,\hat{p}_{02}(t,v)\,.
\end{aligned}
$$

We note that:

- $A_{1,4}$ and $A_{2,5}$ represents male-to-female transmission and $A_{4,1}$ and $A_{5,2}$ represents female-to-male transmission.
- $\hat{p}_{01}(t,i)$ and $\hat{p}_{02}(t,i)$ are the number of susceptible eligible women and men at the demographical steady state, the state into which HIV is introduced.
- The matrix $\mathcal{A}(i,j,\tau)$ captures the rate at which a woman aged j infects men of age i. In the partnered state for women, the rate $A_{5,2}$ is given by the transmission rate in a steady relationship, ϑ multiplied by $\kappa_1(i,j+\tau)\hat{p}_{02}(t,k)$, which is the rate at which a woman aged $(j+\tau)$ partners to a man aged i.

  When calculating R_0 in epidemic models, it is usually assumed that susceptibles remain unchanged throughout the duration of the infectious period of the initial case, i.e. no time dependence is included for the density of susceptibles. To do so seems acceptable if the susceptible pool is large (a necessary assumption in R_0 calculations) when the initial infective is introduced, but infection within a steady relationship is an entirely different process. However, the susceptible is clearly not susceptible for the duration of the entire relationship and the transition matrix $\mathcal{B}_1(i,j,\tau)$ accounts for the fraction of relationship time that the susceptible escapes infection.
- In the eligible state for women the rate of infecting men aged i, $M_{4,1}$, includes the following factors: $\beta(i)$ for the rate at which men aged i make risky contacts (see Fig. 6.2(c)). $\beta(j+\tau)$ for the rate at which women aged $j+\tau$ make risky contacts. $f(i,j+\tau)$ modeling the probability of a casual contact between an infective aged $j+\tau$ and a susceptible aged i. The normalization factor accounts for all contacts between susceptibles aged i with partners of all ages.

6.3.1 Numerical investigation into the properties of the NGM

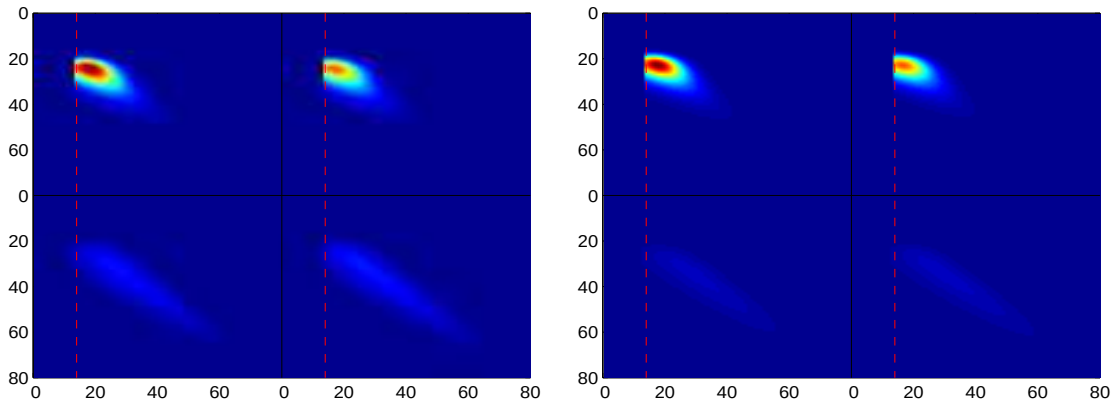
We can approximate a solution for $\mathcal{C}(\theta, \alpha, \tau)$ in Eq. (6.4) by means of the discretized transition matrix $N(\tau_n)$, which remains constant over each interval $\Delta\tau$ of a discretized time since infection:

$$\mathcal{C}_k(\tau_n) = e^{\mathcal{B}_k(\tau_n)\Delta\tau}\, e^{\mathcal{B}_k(\tau_{n-1})\Delta\tau}\, \cdots\, e^{\mathcal{B}_k(\tau_1)\Delta\tau}\, \mathcal{C}_k(0)\,,$$

where, $\mathcal{C}(0)$ is the 6×6 identity matrix (dropping reference to θ and α). The NGM is given by:

$$\begin{aligned} K(x,y) \;=\; & \sum_{k=1,2} \mathcal{A}_k(\tau_n)\, e^{\mathcal{B}_k(\tau_n)\Delta\tau}\, \mathcal{C}_k(\tau_{n-1}) + \mathcal{A}_k(\tau_{n-1})\, e^{\mathcal{B}_k(\tau_{n-1})\Delta\tau}\, \mathcal{C}_k(\tau_{n-2}) + \cdots \\ & + \mathcal{A}_k(\tau_1)\, e^{\mathcal{B}_k(\tau_1)\Delta\tau}\, \mathcal{C}_k(0)\,. \end{aligned}$$

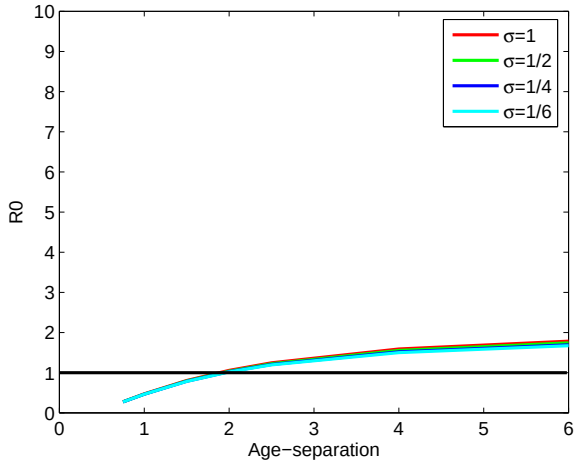
A more direct way of finding $\mathcal{C}(\theta, \alpha, \tau)$ is to solve the ODE (6.4). However, the approach of a product of year-on-year exponential matrices fits conveniently within our finite difference solution, which is also updated at yearly intervals.



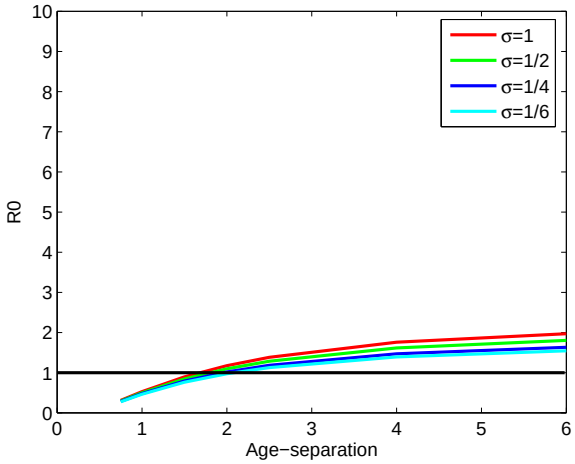
(a) infection of women by men (b) infection of men by women

Figure 6.4: (a) The top left block represents the number of eligible women infected by eligible men who themselves became infected when eligible. The number of infections caused drops from red toward blue. The top right block represents the number of eligible women infected by eligible men who became infected when they were in a relationship. The bottom right block represents the number of partnered women infected by men who became infected when they were in a relationship. (b) A similar block representing infection of men. It is slightly offset from the block depicted in (a), due to the different rate at which men and woman mature into eligibility.

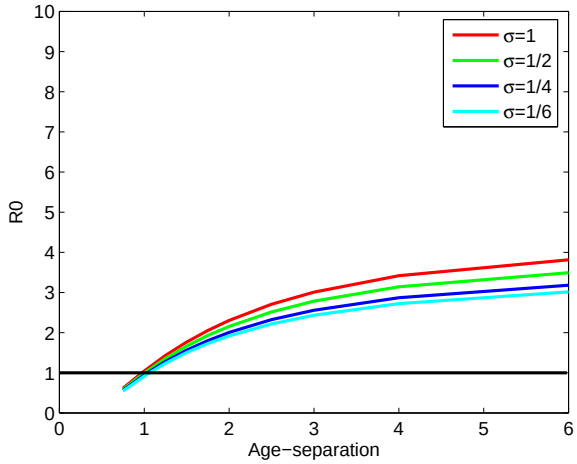
Structure of the NGM. Fig. 6.4(a) shows the expected number of women infected by men. The top left block represents the number of eligible women infected by eligible men who themselves became infected when eligible. The top right block represents the



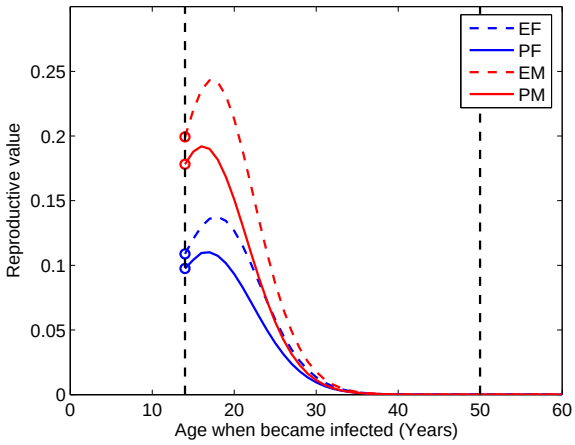
(a) infection dominated by random mixing



(b) infection of steady partners plays a role



(c) increased random contact rate



(d) reproductive value of HIV

Figure 6.5: (a) The spectral radius of M, i.e. R_0, as a function of the standard deviation of an assumed normally distributed partner choice function (see Fig. 6.2). The function is plotted for $\sigma = 1, 1/2, 1/4, 1/6$ in red, green, blue and cyan respectively. For R_0 to be greater than one requires a age separation of roughly 1.5-2.5 years. (b) The result is sensitive to the relationship dissociation rate, the relationship formation rate and to the infection rate within relationships. (c) The same relationship between R_0 and age variance in partner choice when the risk of contracting HIV through risky contacts is increased by a factor of 1.3. A smaller variance in age difference between partners can sustain an HIV epidemic. (d) The left eigenvector of R_0, i.e. the basic reproductive value, as a function of age at infection. It shows the relatively smaller contribution infectives make when they are infected at an older age.

number of eligible women infected by eligible men who became infected when they were in a relationship. The bottom right block represents the number of partnered women infected by men who became infected when they were in a relationship. Regions to the left of the dotted line in each block are zero, since children don't participate in the epidemic.

A similar block representing the infection of men is displayed in Fig. 6.4(b). It differs slightly from the block depicted in (a), due to the different rate at which men and woman mature into eligibility. Note that our model parameters and the conclusions we have drawn represent a generalized situation, and would need to be adjusted to fit specific situations.

Relationship between R_0 and variance in age difference in partner choice. The NGM represents the number of infected cases of different types (age, gender and being eligible or in a relationship) caused by infectious cases of the same types: position (i, j) in this matrix is the number of cases of type i caused by an infective who became infected when in type j. The largest eigenvalue, called the spectral radius, gives R_0 for this model.

We assume susceptibles to be (and to remain) in a steady state when constructing this matrix, so the R_0 we obtain is an upper estimate of its true value. This is the usual approach in disease invasion studies, which use R_0 as an invasion criteria. It could be instructive to use simulation results to supply time-dependent susceptibility counts to the matrix, and to repeat the study presented in this chapter for an advanced epidemic. It may be that these results depend on the stage of the epidemic.

In Fig. 6.5 (a and b) four R_0 curves are plotted as a function of standard deviation of the normal distribution for partner choice (from here onward simply referred to as age separation) between partners, with the average relationship lasting $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{6}$ in red, green, blue and cyan respectively. In Fig. 6.5(a) random mixing, between eligibles, dominates and the relationship separation has almost no influence on R_0. In this scenario we used parameter values of $\kappa = 5\,/\text{yr}$ and $\gamma = 0.5\,/\text{yr}$ respectively. The relationship state can be made more critical by increasing the partnership formation rate and by increasing the infection rate within relationships. If we use parameter values of $\kappa = 50\,/\text{yr}$ and $\gamma = 2\,/\text{yr}$, then infection within discordant couples happens at a rate comparable to infections outside of steady relationships and R_0 increases with an increase in the separation rate. This scenario is shown in Fig. 6.5(b). The overall infection rate tends to one dominated by infections due to random risky contacts when the separation rate is very high. In Fig. 6.5(c) the rate of risky contacts outside of steady relationships is increased by a factor of 1.3, leading to age separation playing a less critical role in disease spread.

An investigation (not displayed) indicates that a very similar R_0 curve can be obtained when the age of sexual debut is 14 years for both men and women. The R_0 curve is only slightly shifted, with 1.5 to 2.5 years still being a good estimate for the smallest age-difference in relationships that would sustain an epidemic. The result is not strongly correlated to the age difference at sexual debut. A similar finding was reported in [50], who found that delaying sexual debut results in only a small reduction in lifetime risk of infection.

It is often said that steady relationships protects partners from HIV infection and that reducing the length of relationships should increase R_0, but this truism is not *a priori* true. It is possible that increasing the rate of relationship separation can decrease R_0, as reported by [59, Fig. 2]. One critical detail is the infection rate within a discordant relationship: If the infection rate is slow then longer-lasting relationships can actually increase R_0. Detailed information about the age-dependent characteristics of both random mixing and discordant relationships are required before the infection process can be completely understood.

Reproductive value. Fisher considered that each female child was given a 'loan' at birth, a loan that could only be repaid by her future offspring (see appendix B.3). He then constructed a function of age, called the reproductive value, representing the fraction of her debt repaid by a certain age: it is exactly 1 at birth and then increases as the individual faces mortality while making no contribution to the population size. As the individual becomes sexually active, the reproductive value increases to a point where the woman starts repaying her debt. In growing populations, her debt will be repaid before the oldest age of childbearing. Examples of demographic reproductive curves are shows in Keyfitz, et al. [65, Ch. 8.1, p.188].

The left eigenfunction of the NGM operator $\mathbb{K}$, or the left eigenvalue of the NGM matrix K, is analogous to the fraction of debt paid back by an infective during his/her life. Our NGM matrix is structured according to heterogeneity in our population, and we can use it to compare the contribution different types of infectives are making to the future infected population.

Fig. 6.5(d) shows the 'infectious reproductive' values of our four infectious types as a function of age. The shape of the reproductive and depend strongly on the shape of the 'risk' function for contracting HIV (Fig. 6.2(c)). This risk curve resembles that of the fertility of women as a function of age (Fig. 5.3(b)). Justification of the use of this curve can be made on the same grounds as basing population-level HIV trends using data from antenatal clinics, as we did in Ch. 5.

6.4 Concurrent relationships and the criticality of age separation

It is thought that concurrent partnerships played an crucial role in the emergence and continue to play a role in the persistence of the HIV epidemic [113]. However, the details of the role of concurrency in the spread of sexually transmitted diseases has thus far not been reliably determined, and the extent of its impact on the spread of HIV has been questioned [82]. One counter-argument hinges on the observation that HIV doubling times are relatively long: 8-16 months is reported in [78, Tab. 3]. If concurrency were a driving factor in the spread of an HIV epidemic in a sexual network, the observed doubling time would be much shorter [117]. The argument is that HIV has a high initial (a few weeks) and much lower

subsequent (a few years) infectiousness. Concurrent partners are therefore more likely to be infected in the early part of an infectious period.

Concurrency and risk heterogeneity partly determine how interventions should be structured. Early interventions that were informed by risk-structured models and targeted risk, sexual behavior and mixing patterns, were found to be ineffective if they did not also reduce concurrency [39]. Although this cannot be a decisive argument regarding the importance of concurrency in the spreading of HIV, it does appear that concurrency and risk heterogeneity processes produced highly complex sexual networks which can be said to exacerbate transmission of the disease.

Understanding how sexual networks are formed is therefore vital to modeling the HIV epidemic. The effect of concurrency could be modelled by allowing individuals to have casual short term relationships, regardless of whether they are active in a current relationship. The result is a network consisting of edges which can be thought of as always active, and edges which define a network in terms of all possible partnerships that may form in a period of time (one year, five years or a lifetime) [38]. This approach is not expected to impact much on the critical role payed by variance in age difference between partners, since the influence of an infective in a steady relationship closely follows that of one confined to mixing outside of steady relationships. The difference is determined by the length of an average relationship. The following approach sheds more light on the impact of concurrency on the threshold behavior.

Using a modified next-generation matrix to model concurrency

In order to help focus our discussion we remove explicit steady relationship dynamics from the model presented in Sect. 6.1. What remains is a process through which individuals become infected through proportionate random mixing. Now suppose that in addition to the random contact process described in Sect. 6.1.1, all individuals take part in another type of contact process, at a slower rate.

This contact process leads to events which cause an infective to infect not only a randomly encountered susceptible, but his/her susceptible partner. This introduces concurrent triads (clusters of three) into the population, where an infectious event infects everyone in the particular cluster. For clarity we use small clusters, where two susceptibles can be infected by an infective, but the method can be generalized to included larger clusters. Using the proportionate mixing approach of Sect.6.1.1, the force of infection faced by men aged a becomes:

$$\begin{aligned}\phi_2(t,a) &= \frac{\beta(a)}{C_2(t,a)}\int_0^\infty \beta(b)f(a,b)p_{11}(t,b)\,db \\ &+\frac{\beta^*(a)}{C_2(t,a)}\int_0^\infty\int_0^\infty \beta^*(b)\beta^*(a')f(a,b)f(b,a')\left[p_{11}(t,b)\right]\left[p_{02}(t,a')+p_{12}(t,a')\right]db\,da' \\ &+\frac{\beta^*(a)}{C_2(t,a)}\int_0^\infty\int_0^\infty \beta^*(b)\beta^*(a')f(a,b)f(b,a')\left[p_{01}(t,b)\right]\left[p_{12}(t,a')\right]db\,da' \\ &+\frac{\beta^*(a)}{C_2(t,a)}\int_0^\infty\int_0^\infty \beta^*(b)\beta^*(a')f(a,b)f(b,a')\left[p_{11}(t,b)\right]\left[p_{11}(t,a')\right]db\,da'\,,\end{aligned}$$

where

- The first term in $\phi_2(t,a)$ represents proportionate contacts between a man and a woman. The second term is the risk a man faces when 'joining' a pair where the woman is susceptible and the man either susceptible or infected. The third term is the risk faced when joining an infected man and a susceptible woman. The last term is the risk he faces when joining two infected women. In this case, the man forms a link between the two women, and the term is not meant to model direct contact between the two women.

- C_2(t,a) is a normalization factor for contacts of men aged a, at time t:

$$\begin{aligned}C_2(t,a) &= \int_0^\infty \beta(b')f(a,b')\,p_{11}(t,b')\,db' \\ &+\int_0^\infty\int_0^\infty \beta^*(b')\beta^*(a')f(a,b')f(b',a')\,P_1(t,b')\,P_2(t,a')\,db'da' \\ &+\int_0^\infty\int_0^\infty \beta^*(b')\beta^*(a')f(a,b')f(b',a')\,P_1(t,b')\,P_1(t,a')\,db'da'\,,\end{aligned}$$

  where $P_1(t,b')=\left[p_{01}(t,b')+p_{11}(t,b')\right]$ and $P_2(t,a')=\left[p_{02}(t,a')+p_{12}(t,a')\right]$. The first term in the normalization factor accounts for 'usual' random mixing through the first term. The second term models the number of ways in which a susceptible man can be infected as a consequence of pairing up with all possible pairs of women and men. The second term models the number of ways at which a susceptible man can be infected as a consequence of pairing up with all possible pairs of women. An assumption is made that the probability of making contact with 'dyads' is independent of the probability of making contacts with 'monads'. It is assumed that these contacts lead to infection, and the order in which infection happens in a 'triad', with two susceptibles and one infective is ignored. $\beta^*(a')$ is the effective rate at which all susceptibles in a triad gets infected.

- The rate at which triads form is given by a function $\beta^*(a,b',a')$. It is assumed that $\beta^*(a,b',a')$ factorizes into independent partner turnover rates: $\beta^*(a,b',a') =$

$\beta^*(a,b')\beta^*(b',a')$. The rate at which individuals make this contact (i.e. forming small clusters) is assumed to be lower than ordinary random contacts: $\beta^*(a) < \beta(a)$.

- The probability that a man-woman-man pairing will form is given by a function $f(a,b',a')$. It is assumed that $f(a,b',a')$ factorizes into independent partner choices: $f(a,b',a') = f(a,b')f(b',a')$.

- This formulation offers flexibility in including higher order clusters. The general formula for the normalization factor is[1]:

$$\sum_{k=1}^{S} \int_0^\infty \cdots \int_0^\infty \beta^*(a_1)\cdots\beta^*(a_k) f(a_1,a_2)\cdots f(a_{k-1},a_k) \times \left[P(t,a_1)P(t,a_2)\cdots P(t,a_k)\right] da_1 \cdots da_k \,,$$

  where S is the order of the largest cluster included. Note that the subscripts of for men and women are omitted, to help illustrate this general form.

The basic reproduction number R_0 Concurrency, in the form of triads involving an infective and two susceptibles, is included in the contact matrix M as follows:

$$A_{i,j} = \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix} \tag{6.9}$$

where

- $A_{i,j}$ is the number of infections of suceptibles aged i, caused by infectives who got infected when aged j. The normalization factor is C_1 (resp. C_2) when $i = 1$ (resp. $i = 2$).

[1] The formula proposed for the normalization factor is inspired by the so-called cluster expansion to model interacting gases in statistical mechanics [88, Ch. 9]

•

$$
\begin{aligned}
A(1,2) &= \frac{\beta(i)}{C_1(i)}\beta(j+\tau)f(j+\tau,i)\,\hat{p}_{01}(t,i)\,, \\
&\quad + \frac{\beta^*(j+\tau)}{C_1(i)}\sum_k \beta^*(i)\beta^*(k)f(j+\tau,i)f(k,i)\,\hat{p}_{01}(t,i)\,\hat{p}_{02}(t,k)\,, \\
A(2,2) &= \frac{\beta^*(j+\tau)}{C_2(i)}\sum_k \beta^*(i)\beta(k)f(j+\tau,k)f(i,k)\,\hat{p}_{01}(t,k)\,\hat{p}_{02}(t,i) \\
A(2,1) &= \frac{\beta(i)}{C_2(i)}\beta(j+\tau)f(i,j+\tau)\,\hat{p}_{02}(t,i)\,, \\
&\quad + \frac{\beta^*(j+\tau)}{C_2(i)}\sum_k \beta^*(i)\beta^*(k)f(i,j+\tau)f(i,k)\,\hat{p}_{02}(t,k)\,\hat{p}_{01}(t,i)\,, \\
A(1,1) &= \frac{\beta^*(j+\tau)}{C_1(i)}\sum_k \beta^*(i)\beta^*(k)f(j+\tau,i)f(k,i)\,\hat{p}_{02}(t,i)\,\hat{p}_{01}(t,k)\,.
\end{aligned}
$$

• The normalization factors $C_1(i)$ and $C_2(i)$, at time t, are given by:

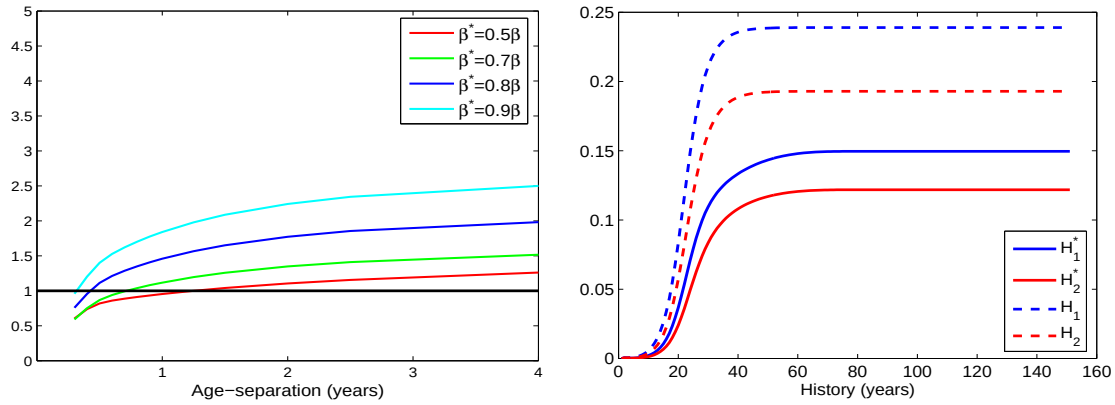
$$
\begin{aligned}
C_1(i) &= \sum_v \beta(v)f(v,j+\tau)\big[\hat{p}_{01}(t,v)\big] \\
&\quad + \sum_{v,w} \beta^*(v)\beta^*(w)f(v,j+\tau)f(v,w)\big[\hat{p}_{02}(t,v)\big]\big[\hat{p}_{01}(t,w)+\hat{p}_{02}(t,w)\big]\,, \\
C_2(i) &= \sum_v \beta(v)f(j+\tau,v)\,[\hat{p}_{02}(t,v)] \\
&\quad + \sum_{v,w} \beta^*(v)\beta^*(w)f(j+\tau,v)f(v,w)\,\big[\hat{p}_{01}(t,v)\big]\big[\hat{p}_{01}(t,w)+\hat{p}_{02}(t,w)\big]\,.
\end{aligned}
$$

• $A(2,2)$ is the rate at which susceptible men aged i are infected by infective men who got infected at age j, through women of all age categories, who act as 'vectors' in this process. Similarly, $A(1,1)$ represents women infecting women through men. The second contributions to $A(1,2)$ and $A(2,1)$ are the rates at which women (resp. men) infect men (resp. women) indirectly through a third partner.

The simulation results presented in Fig. 6.6(a) are surprising. Our first conclusion, shown in Fig. 6.6(a), concerns the effect of concurrency on the critical role played by age separation is explored by setting $\beta^* = 0.5\beta,\ 0.7\beta,\ 0.8\beta$ and 0.9β respectively. It appears that increased numbers of clusters diminish the critical influence of age separation in the spread of the epidemic. This result was previously found in a Reed-Frost model with 'tunable' clustering [23].

Secondly, it is interesting to note that concurrency can be an important factor even when infectiousness is not modelled as dependent on time since infection. The result shows that concurrency can be an important factor even in models with a long infectious period.

Our third and most surprising conclusion concerns the impact of concurrency on the doubling time of HIV. Fig. 6.6(b) is a simulation result showing prevalence as a function of age. Here it is assumed that $\beta^* = \beta$. It could be anticipated that including concurrency would drastically increase the doubling time and final prevalence of the epidemic, but our results suggest otherwise. Prevalence as function of time is still reasonable, and can be adjusted to fit an observed curve, even when the formation of triads happens at the same rate as that of dyads. The normalization factor has moderated the impact of concurrency on the prevalence curve.



(a) concurrency and age separation (b) realistic prevalence trend

Figure 6.6: (a) Comparing the relative influence of concurrency, as approximated by the formation of triads between individuals. The relative rate of currency, β_2 is varied as a fraction of β_1: $\beta_2 = 0.5\beta_1,\ 0.7\beta_1,\ 0.8\beta_1$ and 0.9β. (b) Realistic doubling time and prevalence curves in a model where 30% of individuals participate in concurrent relationships. H_k (solid line) is the prevalence of women ($k = 1$) and men ($k = 2$) in a hypothetical population. The same data in Masiphumelele is shown in the dashed line.

The influence of concurrency on R_0 and doubling times is not as straightforward as it seems and must be explored in the context of many possible transitions and histories which can be played out in an individual's lifetime. The value of the NGM is that it takes all possible transitions, as given by a transition matrix, into account.

6.5 Conclusions

Attempts have previously been made to model the impact of intergenerational sex on the spread of HIV. An interesting study is that of Hallett, et al. [50] which developed a mathematical model for the heterosexual spread of HIV in order to study the population-level impact of behavioral interventions, focusing on reducing cross-generational sex and delaying sexual debut. Their conclusions were surprising. It was found that endemic prevalence is higher with cross-generational sex, and that female-to-male prevalence ratios can be created in a model without differential susceptibility between males and females. In all scenarios investigated, they found that delaying sexual debut results in only a small reduction in

lifetime risk of infection. More surprisingly, they found that reducing cross-generational sex has a limited effect on lifetime risk of infection if not accompanied by a reduction in the number of risky contacts.

Our more subtle approach to modeling age difference and HIV spread indicates that, in fact, a critical level of intergenerational sex is required to establish a major HIV epidemic in a heterosexual population. Building on the framework used in previous chapters, we formulated a next-generation operator and estimate R_0, the epidemic threshold criterion. We showed that a critical level of age separation, given by roughly 1.5-2.5 years standard deviation of a normally distributed (with zero mean) partner choice, is necessary for a major HIV epidemic. The result is insensitive to the difference in the average age of sexual debut for both men and women.

The result is however sensitive to the details of partnership dynamics. For example, increasing the rate at which random contacts occur reduces the criticality of the role played by intergenerational sex in establishing a major epidemic. Another critical aspect of the model is the infection rate within discordant couples. To measure the relative impact of different infectious types (classified according to age, gender, relationship type) we proposed the use of the left eigenvector associated with the largest eigenvalue.

The influence of concurrency was modeled by a modified proportionate mixing scheme. We first simplified the model by not including steady relationship dynamics explicitly. Then, in additional to random casual contacts, we allowed individuals to meet pairs of individuals, thus creating more routes of infection. We propose the use of a normalization factor which can in principle be generalized to handle large groups of individuals. We found that adding 'triads' (groups of three) already increases the connectivity of the sexual network to such an extent that the critical role played by age separation is diminished, although in many communities it remains critical enough to warrant attention.

There is an interplay between age separation and concurrency: if a person has more than one partner, it is likely that they will not be the same age. Concurrency can therefore lead to variance in the age difference between partners [34]. Behavioral interventions should aim to discourage both intergenerational sex and high partner turnover rates in order to help control the epidemic.

Chapter 7

Deterministic and individual-based STD models

We develop simulation and analytical tools to study variation and fluctuations in age-structured partnering models. We noted (in Ch. 3) that macroscopic equations are approximations to population models when there are so many individuals involved that fluctuations can be neglected. To better understand the nature of fluctuations in partnering models, we investigate two different techniques. One is a Gillespie-type stochastic simulation technique, which focuses on individuals [45]. The other is an analytical technique, which goes beyond the macroscopic approach, but does not model all the information contained in the microscopic model. Instead, the method gives an intermediate description, modeling the fluctuations in the system by means of a system of differential equations [103].

Information regarding life events at the individual level can be averaged to obtain a description of dynamics at the population level, a description which may shed light on the complexities of sexual networks and the spread of epidemics. During his lifetime an individual experiences a number of events: birth, maturing, relationship participation possibly a number of times and finally death, each of which can be seen as putting the individual in a certain state. When an individual life cycle can be described as a Markov chain, we can compute and assign probabilities to a sequence of life events. The method we use 'sums over all possible individual histories'.

In epidemiological models we can apply the method to the linear and early epidemic, when susceptibles are not yet appreciably depleted. This would allow us to compute distribution for statistics and variables in models which can be compared to real-world data. When the stochastic process is non-linear, as many interesting epidemic models are, then the Laplace technique becomes considerably more difficult. For non-linear models, we developed a Gillespie-type simulation platform, as well as a master equation approach.

We show how a Fokker-Planck equation (FPE) can be derived to model fluctuations in non-linear age-structured populations. The method is based on that of van Kampen [107, Chp 10] and models the fluctuations around macroscopic or mean-field values as proportional

to the square root of the population size [89, Chp 8.4]. The method relies on an expansion technique in which only the first two terms are required to formulate the FPE. Initial effort spent in deriving a master equation and expanding it in terms of a parameter proportional to the inverse of the size of the population, is rewarded by having differential equations for the variance and co-variance of macroscopic variables.

7.1 Linear dynamic models and the Markov approach

Connections between the many ways of studying problems of linear population dynamics are summarized in appendix B.3. When the population-level process is given a stochastic individual-level description and the resulting master equation is linear, linear operator theory comes into play in a powerful way. The linear theory discussed in this section gives a complete description of a linear system. We see that the resolvent of an operator describes the probabilities of all possible transitions through states, because it is essentially a series expansion of the projection operator, where each term in the series corresponds to successive iterates of the projection.

Let us assume a Markov process for the transition of an individual between an arbitrary but finite set of states, obeying the following master equation [107, Chp 7]:

$$\dot{P}_i(t) = \sum_j \{\gamma_{ij} P_j(t) - \gamma_{ji} P_i(t)\}\,, \tag{7.1}$$

where γ are constant rates at which random jumps between states occur. Suppose one knows a sequence of times at which jumps took place, as well as the levels $(i_0, i_2, \cdots)$ occupied before these jumps:

$$0 < \tau_{1|_{i_0}} < \tau_{2|_{i_1}} < ... < \tau_{s|_{i_{s-1}}} < t_{|_{i_s}}$$

The probability for any level i to survive for a time τ is $e^{-\gamma_i \tau}$ where $\gamma_i = \sum \gamma_{ji}$. The probability for a jump to occur between τ and $(\tau + d\tau)$ is:

$$\gamma_{ji}\, e^{-\gamma_i \tau} d\tau\,,$$

hence the probability density for the particular realization is:

$$\begin{aligned}&\gamma_{i_1 i_0} \exp[-\gamma_{i_0}\tau_1]\gamma_{i_2,i_1} \exp[-\gamma_{i_1}(\tau_2 - \tau_1)] \times \\ &\times \gamma_{i_s i_{s-1}} \exp[-\gamma_{i_{s-1}}(\tau_s - \tau_{s-1})] \exp[-\gamma_{i_s}(t - \tau_s)]\,.\end{aligned}$$

The solution of Eq. (7.1) is obtained by integrating this realization with respect to arbitrary sequences of jump events, summing over all possible sequences and initial conditions. The

derivation is done in detail by [107, Chp 7].

The highlights are as follows. First regard P_i as components of a vector P and then define matrices:

$$\begin{aligned} X_{ij}(t) &= \gamma_{ij} e^{-\gamma_j t} \\ Y_{ij}(t) &= \delta_{ij} e^{-\gamma_j t}\,, \end{aligned}$$

to obtain:

$$\begin{aligned} \vec{P}(t) = \Bigg[Y(t) + \sum_{s=1}^{\infty} \int_0^t d\tau_s \int_0^{\tau_s} d\tau_{s-1} \int_0^{\tau_{s-1}} d\tau_{s-2} ... \int_0^{\tau_2} d\tau_1 \\ \times Y(t-\tau_s) X(\tau_s - \tau_{s-1}) X(\tau_{s-1} - \tau_{s-2}) ... X(\tau_1) \Bigg] \cdot \vec{P}(0)\,. \end{aligned} \tag{7.2}$$

Introducing the Laplace transform changes the convolutions into products of matrices X and Y:

$$\hat{P}(\lambda) = \left[\hat{Y}(\lambda) + \sum_{s=1}^{\infty} \hat{Y}(\lambda) \{X(\lambda)\}^s \right] \cdot \hat{P}(0)\,.$$

The solution of (7.1) is thus expressed as a sum over histories (i.e. a sum over all s and intermediate times $\tau_1, ..., \tau_s$), and is finally obtained by means of the inverse Laplace transform:

$$\vec{P}(t) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{t\lambda} \hat{Y}(\lambda) \{1 - \hat{X}(\lambda)\}^{-1}\, d\lambda \cdot \vec{P}(0)\,, \tag{7.3}$$

where c is taken to the right of all the singularities of the integrand. The asymptotic behavior is determined by the pole of the integrand with the largest real part (compare with appendix B.3).

The Markov process given by Eq. (7.1) can of course be studied by means of a transition matrix:

$$\dot{P}_i(t) = \mathbb{W} P_i(t)\,, \tag{7.4}$$

where $\mathbb{W}_{ij} = \gamma_{ij} - \delta_{ij} \sum_{j'} \gamma_{j'i}$. The poles of the integrand in Eq. (7.3) are the eigenvalues of $\mathbb{W}$ and the two approaches give equivalent descriptions of the asymptotic behavior of the system. Furthermore, $\mathbb{W}$ can be expressed in terms of its resolvent [91, Chp 6]:

$$e^{t\mathbb{W}} = \frac{1}{2\pi i} \oint_{|\lambda|=a} e^{t\lambda} R(\lambda, \mathbb{W})\, d\lambda\,.$$

where a is greater than all the eigenvalues of $\mathbb{W}$ and the resolvent of $\mathbb{W}$ is

$$R(\lambda, \mathbb{W}) \quad = \quad \hat{Y}(\lambda)\{1 - \hat{X}(\lambda)\}^{-1}\,.$$

This short excursion shows how the Laplace transform can be used to study the asymptotic properties of a Markov process. Some non-Markov processes can also be studied using this technique. For example, an age structured HIV model is studied by Inaba [60]. After transforming the problem to an integral equation, the Laplace transform is used to find the functional form of the next-generation operator. The spectrum of this operator is then analyzed using Perron Frobenius theory and the existence of an eigenvalue with a dominant real part is proved. In appendix B.3 we study the one dimensional renewal equation of demography, using the same approach.

The Laplace approach leads to further insights. The framework in which the probability distribution in Eq. (7.3) is defined allows one to compute probability generating functions, which can be used to calculate the average time spent in each state, the correlation between time spent in different states, and the number of times each state is entered.

There is an equivalence between the exponential of the state transition matrix and the inverse of the resolvent of the state transition matrix. It shows that solving for the distribution of states in a system by means of an exponential of the state transition matrix over a time interval, takes all possible histories over the time interval into account.

7.2 Simulating age-independent partnering dynamics

The Gillespie simulation method, originally used to study reacting chemical systems' [45], is used to study partnering and infection models at an individual level. The algorithm provides the following input to a simulation algorithm: (1) when will the next reaction happen, and (2) what type of reaction will it be? Estimates for these two values (a continuous-time estimate in the first and a reactant type in the second) are calculated by taking all current reactants and types of interactions into account. The method is modified here and built into a framework which can be used to track individuals in a partnering model. The method is first applied to a Markov model where all rates are independent of age.

The method is refined to maintain a list of person objects (see Tab. 7.1) in an individual based stochastic model. Each person object in the list contains enough information to characterize a particular individual uniquely. During the simulation, events are scheduled which update characteristics (or 'fields') of individuals in the list. Tracking the list of person objects over time allows us to construct a realization of the stochastic evolution of the system. From this list we extract interesting statistics regarding the underlying process.

A list of relationship objects is also maintained. Each relationship object contains information about the man and woman participating in the relationship: what their identities are, when the relationship formed, what their ages were when the relationship formed, and so on. This allows us to construct a realization of the stochastic network object which un-

derlies the partnering model. This object is analyzed statistically to gain more insight into partnering models.

A simple macroscopic partnering model with exponential lifetime distribution, constant rates for linear processes (mortality, maturing from youth into eligibility) and a simple harmonic mean partnering function, is given by the following system of coupled ODEs:

$$\begin{aligned}
\frac{d}{dt}M_y(t) &= B - \mu_m\, M_y(t) - \eta_m\, M_y(t)\,, \\
\frac{d}{dt}F_y(t) &= B - \mu_f\, F_y(t) - \eta_f\, F_y(t)\,, \\
\frac{d}{dt}\,M_e(t) &= -\mu_m\, M_e(t) + \eta_m\, M_y(t) + \big[\sigma + u_f\big]\, C(t) - K\frac{M_e(t)F_e(t)}{M_e(t)+M_e(t)}\,, \\
\frac{d}{dt}F_e(t) &= -\mu_f\, F_e(t) + \eta_f\, F_y(t) + \big[\sigma + u_m\big]\, C(t) - K\frac{M_e(t)F_e(t)}{M_e(t)+F_e(t)}\,, \\
\frac{d}{dt}C(t) &= -\big[\sigma + \mu_m + \mu_f\big]\, C(t) + K\frac{M_e(t)F_e(t)}{M_e(t)+F_e(t)}\,.
\end{aligned}$$

where

- The following types of individuals are distinguished: M_y, F_y: young men and women. M_e, F_e: eligible men and women. $C = \{M_p, F_p\}$: partnered men and women.
- B: constant birth rate, which can be viewed as an exogenous Poisson process.
- μ: mortality, η: maturity, σ: partnership dissociation rate.
- $K\frac{M_e(t)F_e(t)}{M_e(t)+F_e(t)}$: partnership formation rate.

The main point of the Gillespie algorithm is to simulate the jump events which describe the system at a macroscopic level. It is constructive to classify the type of events in this system. Events which create individuals can be endogenous (i.e. individuals in the population giving birth) or exogenous (i.e. immigration). Some events move or transfer individuals between types, such as maturing into eligibility, or becoming partnered. Others remove individuals from the system, such as death or emigration.

The system can equivalently be described in terms of competing events. We identify the following time-inhomogeneous Poisson processes, identifying also their intensities at time t:

$$
\begin{aligned}
a_1 &= \eta\, M_y(t)\,, \\
a_2 &= K\,\frac{M_e(t)F_e(t)}{M_e(t)+F_e(t)}\,, \\
a_3 &= \sigma\, M_p(t)\,, \\
a_4 &= \mu_m\left[M_y(t)+M_e(t)+PM(t)\right], \\
a_5 &= \eta\, F_y(t)\,, \\
a_6 &= K\,\frac{M_e(t)F_e(t)}{M_e(t)+F_e(t)}\,, \\
a_7 &= \sigma\, F_p(t)\,, \\
a_8 &= \mu_m\left[F_y(t)+F_e(t)+F_p(t)\right], \\
a_9 &= B\,.
\end{aligned}
\tag{7.5}
$$

The intensities listed in Eq. (7.5) are used to compute the time to the first event and its expected type. Define a_0 as the intensity of the Poisson process corresponding to the first event:

$$
a_0 = \sum_{i=1,3,4,5,7,8,9,(2,6)} a_i\,. \tag{7.6}
$$

Draw two numbers r_1 and r_2 from the unit uniform distribution. The time to the next event is given by sampling the inverse CDF of an exponential distribution with rate a_0:

$$
\tau = \frac{1}{a_0} * \log\left(\frac{1}{r_1}\right). \tag{7.7}
$$

The event type is given by the smallest integer for which:

$$
\sum_{v=1}^{u-1} a_v \le r_2\, a_0 \le \sum_{v=1}^{u-1} a_v\,. \tag{7.8}
$$

Note that in Eq.(7.6) we have grouped events 2 and 6, because a pair of individuals (one man, one woman) is randomly chosen for the partnering event. Also note that at each step the algorithm considers the sum of probabilities in which the system can arrive at each new state, making it an analogue of the sum over histories considered in Sect. 7.1 (as commented on in [106]).

Having found τ and u by this method, the time evolution of the system is straightforward: the event is applied to update the relevant sub-population and an age of τ is added to each living population member. Suppose that a partnering event ($u = 2$ or $u = 6$) is the next event. An eligible man and woman are drawn randomly and partnered, after which

the population counts M_e, F_e and M_p, F_p are updated. Using this method the individual histories of suitably defined person objects can be tracked.

Defining a person object. A person object comprises fields which encapsulate the properties of an individual: gender, age, alive (or dead), number of times being eligible, number of times being partnered, and so on.

Tab. 7.1 lists the fields of a person object:

- The Id number of the person object is used to identify the person object uniquely. It is natural to represent the population as a list of person objects. The Id number of a person can correspond to the person's position in this list.
- Registering the birth date and continuously updating the age of a person makes it possible to time stamp each life event. A calender year is the natural unit of time to use, and events are stated as rates per year. Events over a lifetime make up the life history of an individual. We model the histories of a large number of individuals, and average over these histories to obtain a description for the population.
- Fields corresponding to events which may happen many times in a lifetime (e.g. being eligible or being partnered) have counters which count the number of times the field is entered. The relationship counter, for example, can be used to obtain a distribution of the number of relationships experienced in a lifetime.

	Field	Description
1	Id number	unique identification
2	sex	1=man, 2=woman
3	birth date	calender time
4	age	chronological age
5	relationship status	1 when in relationship, 0 if not
6	total time in relationships	
7	young	1 when young, 0 after maturing
8	eligible	counts each entry into eligibility status
9	partnered	counts each entry into being in relationship status
10	not alive	0 while alive

Table 7.1: Data structure of a person object

Defining a relationship object. A relationship object can similarly be defined and used to record information about the network of relationships links. The utility of the relationship object becomes more apparent in Sect. 7.3, when we study the relationship network of age-structured partnering models. However, the age-independent model is ideal for discussing the architecture and design of our simulation approach. Having laid the ground work in a simplified model, we can focus on further design issues when studying a structured and more complex model.

Tab. 7.2 lists the fields of a relationship object:

- When a relationship formation event is scheduled, an eligible man and woman are chosen randomly. A relationship object is created to record vital information about each relationship.
- The relationship object is uniquely defined by the identity number of the man and woman participating in it.
- The ages of the participating man and woman are recorded at the instant when the relationship is formed. This information is used in Sect. 7.3 when we use the relationship object to extract relationship paths between men and women. We then study the age separation in these infectious links statistically.
- The number of relationships entered into by men and women is recorded. This statistic can be used to obtain an empirical distribution of the number of relationships individuals have during their life, and can be compared to behavioral data. A complete history of relationships can also be constructed, but real data on relationship histories is rarely available.
- The calender times of the relationship start and end events are recorded. This information is used to construct relationship paths that have correct time ordering. By this we mean that each link in the path must be in chronological order.

	Field Description
1	Id number of man
2	Id number of woman
3	age of man when relationship starts
4	age of woman when relationship starts
4	no. of relationship man had
5	no. of relationships woman had
6	calender time relationship starts
7	calender time relationship end

Table 7.2: Data structure of a relationship object

Simulation results. Fig. 7.1 presents results from a single simulation of a partnering model in a small population. Tab. 7.3 lists the parameter values used in the simulation. Fig. 7.1(a) shows a population history, with the partnering process initiated at year 100, after demographical equilibrium is reached. The ratios between the numbers of men and women in each category are determined by relationship formation (K) and dissociation (σ) rates.

To test basic functionality we verify that the simulation algorithm returns correct distributions over states, as determined by hazards stated in Tab. 7.3. Fig. 7.1(b), for example,

shows that the simulated age distribution is an exponential with a correct average of 50 years. Relationship duration is also simulated correctly, with a relationship lasting on average two years, as shown in Fig. 7.1(c). Fig. 7.1(d) shows the number of relationships men and women are expected to have during their lifetime. The right tails of these distributions are too long to be realistic. Constant hazards, producing exponentially distributed duration of escaping the hazard, lead to such unrealistic distributions.

Parameter	Value	Description
μ	1/50 per year	natural mortality rate
η_f	1/15 per year	rate at which young women mature into eligibility
η_m	1/20 per year	rate at which young men mature into eligibility
σ	1/2 per year	relationship dissociation rate
B	20 per year	birth rate

Table 7.3: Parameter values for a simplified partnering model

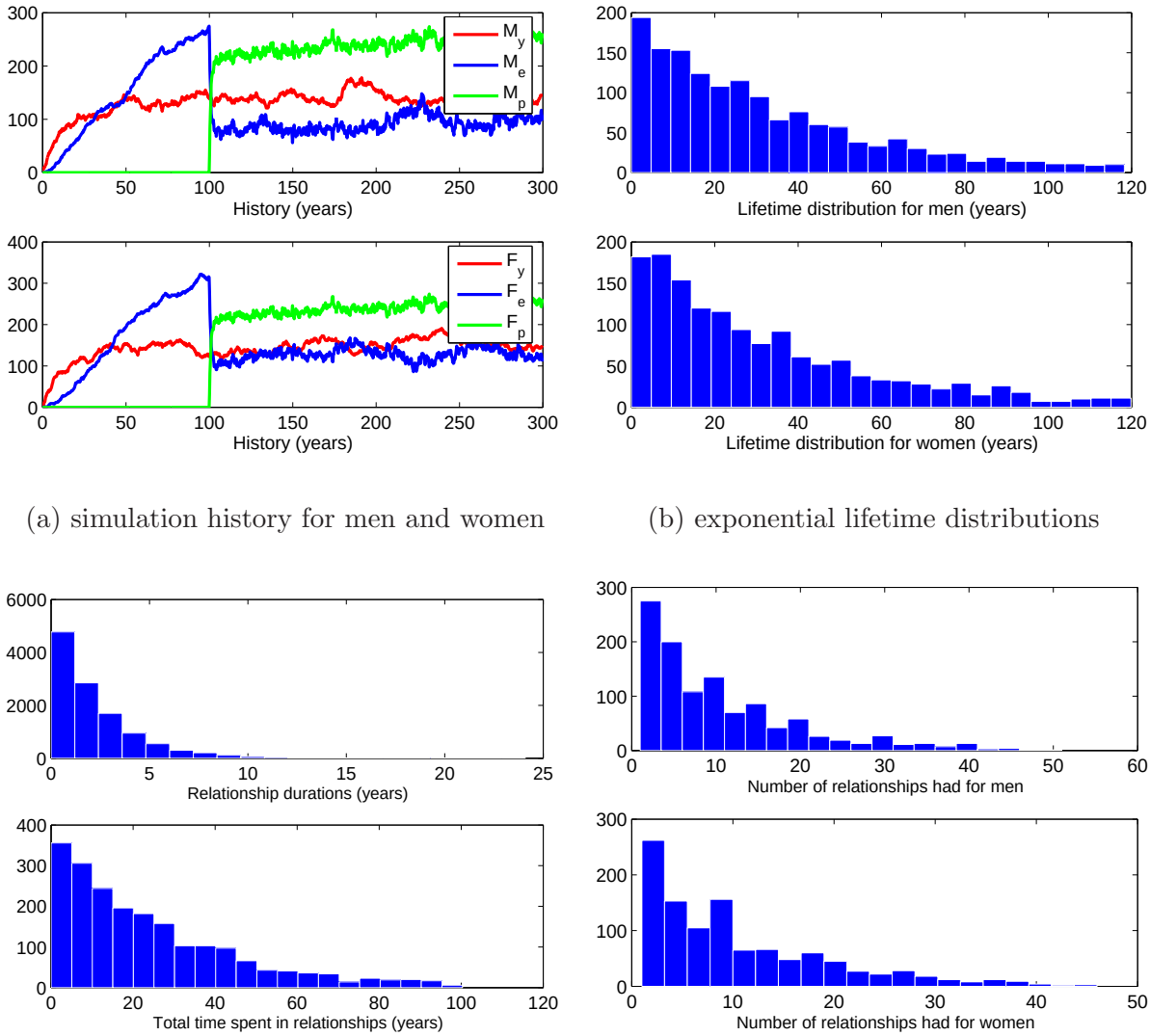
7.3 Simulating age-dependent partnering dynamics

A Gillespie-type method can be used to simulate the age-structured partnering SI models. To use this method, we have to keep track of all events in all age categories. To each age category, we attach event rates at the average age of the category. Aging events are not explicitly modeled. Instead, individual categories are updated each year. The partnering model we study in this section is the same as the basic partnering model discussed in appendix C. We model a type of infection process, by introducing an 'labeled' person and 'labeling' each person who have been in a relationship with a labeled person. We also allow individuals who are eligible for relationships to be 'labeled' according the same proportionate mixing process of Sect. 6.1.1.

Designing a simulation platform. The introduction of a continuous age variable brings with it a potentially infinite proliferation of possible events. To handle these events, we discretize the age variable into 80 one-year boxes. Tab. 7.4 is an index of all the possible events that can happen in our model. The table is a simple bookkeeping or indexing system, providing 80 event indices for each of the event types we use in our model.

The model is not a strict implementation of the SI model of Ch. 6; the process of infection is simplified. Here we simply label individuals who have been in a steady or casual relationship with a labeled individual. This can be seen as an infection process where being in a relationship with an infective (i.e. labeled) individual guarantees transmission. This mechanism is silent on the role played by time since infection on transmission probability, but focusses instead on how relationship formation facilitates disease spread.

It is helpful to present the outline of a Gillespie implementation for age-dependent partnering in pseudo programming language. The syntax of the 'language' we use in this section is self-evident. Regarding the meaning and calculation of each event, we note that:



(a) simulation history for men and women

(b) exponential lifetime distributions

(c) relationship durations

(d) lifetime number of relationships

Figure 7.1: (a) [top] Simulation history for men and women. [bottom] Population history for women. The partnering process was initiated at time $t = 100$. (b) [top] Exponential lifetime distribution for men. [bottom] Exponential lifetime distribution for women. (c) [top] Distribution of time spent in distinct relationships. [bottom] Distribution of total time spent in relationships. (d) [top] Distribution of number of relationships experienced by men. [bottom] Distribution of number of relationships experienced by women.

Event Type	Event Index	Description of state transition
1	1:80	young man becomes eligible
2	81:160	eligible man to steady relationship
3	161:240	eligible man to casual relationship
4	241:241	man from steady relationship to eligible
5	242:242	man from casual relationship to eligible
6	243:322	man in any category dies
7	323:402	young woman becomes eligible
8	403:482	eligible woman to steady relationship
9	483:562	eligible woman to casual relationship
10	563:563	woman from steady relationship to eligible
11	564:564	woman from casual relationship to eligible
12	565:644	woman in any category dies
13	645:645	a birth event

Table 7.4: Index to event types in an age-structured SI-type model.

- The way that events are calculated for women are naturally very similar to the way that the corresponding events are calculated for men. In what follows we list the event pairs, and we state a pseudo formula or in some cases pseudo algorithm to describe how these rates are calculated for men.

- Event $1, 7$: $a\big[\text{Index}(1)\big] = \eta_m \odot p_{ym}$, where p_{ym} is an array of young men in ϖ categories. Syntactically, the expression $\odot$ means a element-wise (or an age category-wise) multiplication of the two arguments on both sides of the operator.

- Event $2, 8$: In this function we have to calculate the age-dependent rates at which eligible men enter steady relationships. The entries into the matrix a, which stores the event rates according to the index tabulated in Tab. 7.4, are calculated in a loop of one calculation for each age category n:

$$\begin{aligned} \text{For n} &= 1:\varpi \\ \text{N}_{\text{pdf}} &= \text{Normal}_{\text{pdf}}\left(\text{n} - (1:\varpi), 0, 5\right) \\ \text{f}_{\text{M}}(1:\varpi) &= K * N_{pdf} \odot \frac{\text{p}_{\text{em}}(\text{n}) \odot \text{p}_{\text{em}}(1:\varpi)}{\sum_v \left[p_{em}(v) + p_{ef}(v)\right]} \\ \text{a(n)} &= \sum_{\text{v}} \text{f}_{\text{M}}(\text{v}) \\ \text{endfor} & \end{aligned} \tag{7.9}$$

where $\text{N}_{\text{pdf}} = \text{Normal}_{\text{pdf}}\left(\text{n} - (1:\varpi), 0, 5\right)$ is a normal probability distribution (with a mean of 0 and a standard deviation of 5 years) for the probability that a man aged n will choose to partner women aged $1:\varpi$. The partnering rate K provides a simple way of adjusting the intensity of the partnership formation process (see Sect. 6.1.1). $f_M(n)$ is the rate at which an eligible man aged n forms relationships.

- Event $3, 9$: $a\big[\text{Index}(3)\big] = \phi \odot \text{not labeled}\,(\text{p}_{\text{em}})$, calculates the hazard of being infected, for each eligible man who has not yet been labeled. $\phi(n)$ is an average age-dependent rate of getting labeled, calculated by the formula given by Eq. (6.2) for proportionate mixing.

- Event $4, 10$: $a\big[\text{Index}(3)\big] = \sum \sigma \odot p_c$, calculates the rate at which disassociation events happen to men in steady relationships. This is the overall rate at which relationships dissociate and an active relationships is chosen randomly when the event is processed.

- Event $5, 11$: $a\big[\text{Index}(5)\big] = 0$ is the rate at which casual relationships disassociate. We set it to zero, to comply with our previous definition of casual relationships being instantaneous. This poses no problem to the Gillespie algorithm. These events will simply never be chosen as the most likely next event.

- Event $6, 12$: $a\big[\text{Index}(6)\big] = \sum \mu_m \odot \text{M}_\text{L}$. Here $\text{M}_\text{L} = \text{Living}\,(p_{ym} + p_{em} + p_c)$ captures the age-dependent hazard of mortality for all living men: young, eligible or partnered.

- Event 13: $a\big[\text{Index}(13)\big] = \text{B}$ captures the probability of a birth event occurring in the next time interval. When the event is processed, a draw from a uniform unit interval can be used to decide if the newborn is male or female.

A large number of age-dependent events can be calculated and managed through an appropriate indexing or bookkeeping scheme. The next event type and the waiting time to this event are calculated as before by means of Eqs. (7.7) and (7.8). When the event type is chosen, the index is used to decide which subgroup and age category to affect. This is achieved through event processing subroutines, one corresponding to each event type.

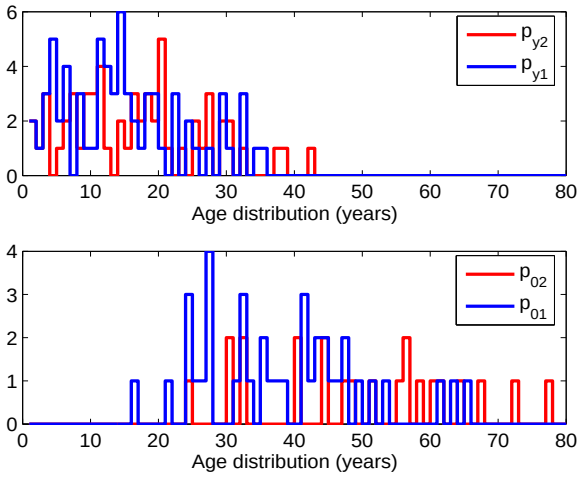
Processing of events entails choosing a random individual from the required age and population category. Once chosen, the state variables of each individual are updated: e.g. indicating that an individual is now eligible for relationships, or incrementing a counter recording how many times he or she has entered the partnered or eligible states. When these two counters are equal (and greater than zero) the individual is eligible for steady relationships. The event may be a death event, after which the individual will be flagged as dead and the individual's partners are returned to the eligible category.

Partnering and casual linking events are handled in a special way. To optimize run time, an average rate is computed at which individuals in each age group get partnered or linked through casual relationships. To find the correct (in an an average sense) partner for the partnering event, the average age-dependent rate must be disaggregated into the rates at which the individual can partner or link with partners in all age groups. A second random draw is made to decide which age group, on average, should be partnered or linked with the individual.

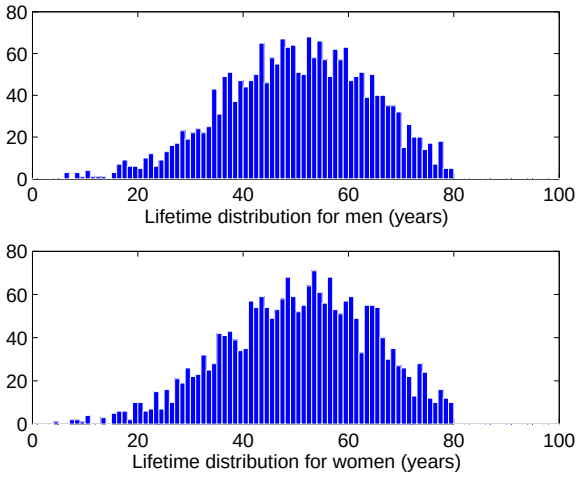
After processing each event the ages of all living individuals are incremented by the waiting time to the most recent event. The age categories of all subgroups are recalculated by applying a filter, one for each category, to the general population. Note the following

trade-off: the age categories should be small so that the 'aging error' is small. This will model aging correctly as a 'flow'. At the same time the categories must be large enough to contain enough individuals for competing events to be handled in a meaningful way [43, Chp. 8]. We encounter the same tradeoff in Sect. 7.4.

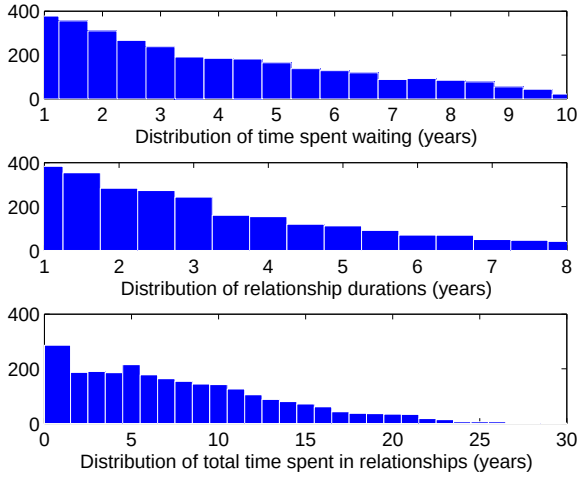
Simulation results. The following are some of the results of simulating an age-structured SI-type model Ch. 6. The parameter values are discussed in Sect. 6.1.1. In Fig. 7.2(a) we show the number of men (in red) and women (blue) in the categories of young (top) and eligible for relationships (bottom). From this we get an idea of fluctuations indifferent age categories. Fig. 7.2(b) depicts age distributions of men and women (excluding children) at steady state. Fig. 7.2(c) depicts empirical (histograms) relationships statistics, which can be used to help calibrate the model to real world data. Fig. 7.2(c,top) shows the average time a person waits to be linked in casual or steady relationship. Fig. 7.2(c, middle) depicts the average time spent in a steady relationship. Fig. 7.2(c, bottom) is a distribution of the total time spent in relationships in a typical lifetime. Fig. 7.2(d) depicts the lifetime number of relationships had by men (top) and women (bottom).



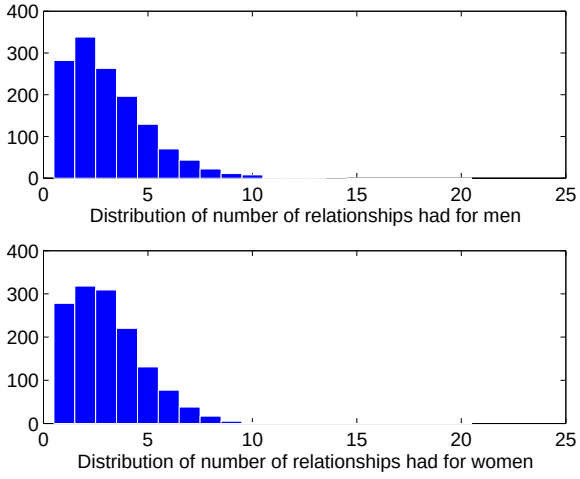
(a) age distributions



(b) lifetime distributions



(c) relationship statistics



(d) number of lifetime partners

Figure 7.2: Simulation results for age-independent partnering model.

7.4 The Fokker-Planck approximation to diffusion in two-sex models

Non-linear age-dependent models are difficult to simulate. Software used to simulate realizations take a long time to develop and individual runs are often time consuming. The Fokker-Planck approximation, the result of an expansion in inverse powers of the size of age bins, gives ODEs for the variance and co-variance in fluctuations. These equations are relatively straightforward to derive, once sufficient familiarity with the method is attained. Solving ODEs for the noise is much faster than simulating stochastic realizations using a Gillespie type method. Another advantage FPE approach is that makes connection between both stochastic and deterministic approaches to modeling non-linear population models.

Certain non-Markov processes, for example those describing age-structured population dynamics, can be modeled using a master equation. The approach is to make the problem essentially Markov, by including variables responsible for the non-Markov character, which in this case relate to the aging process. Building on an outline of the method given by van Kampen [104] and [107, Ch. 14], we discretize the age interval to create cohorts of individuals. There are N_λ members in each cohort, which is generically labeled λ. The purpose of this section is to model the joint distribution of all cohorts in the model, $P(\{N_\lambda\}, t)$, as a functional of the age-profile in the population.

The model. The model we use is based on the one of Sect. 6.1, but simplified to help illustrate the expansion method. We remove explicit steady relationship dynamics and simplify the functional form of the force of infection to that of mass action. Disease status is given by subscripts 0-susceptible, 1-infectious. The subscript k is used to indicate gender: $k = 1$ for women and $k = 2$ for men. Let $p_{yk}(t, a)$ be the number of young women and men aged a. Let $p_{0k}(t, a)$ be the number of eligibles aged a who are susceptible to infection and $p_{1k}(t, a)$ the number of eligibles who are already infected. The mortality rate for individuals aged a is modeled by $\mu_k(a)$. Disease related mortality rate is given by $\mu_{dk}(a)$. The rate at which young individuals aged a mature into becoming eligible for relationships is given by $\eta_k(a)$. The rate at which eligible women aged a meet eligible men aged b is given by $\kappa(a, b)$. Individuals are born at a rate of B, and an equal fraction of men and women are born. The equations for the model are:

- For young men and women:

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) p_{yk}(t, a) = -\mu_k(a)\, p_{yk}(t, a) - \eta_k(a)\, p_{ky}(t, a)\,, \tag{7.10}$$

$$p_{yk}(t, 0) = \frac{B}{2}\,. \tag{7.11}$$

- For susceptible men and women:

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right)p_{0k}(t,a) \;=\; -\mu_k(a)\,p_{0k}(t,a)+\eta_k(a)\,p_{yk}(t,a)-\phi_k(a)\,p_{0,k}(t,a) \tag{7.12}$$

- For infected men and women:

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right)p_{1k}(t,a) \;=\; -\mu_{dk}(a)\,p_{1k}(t,a)+\phi_k(a)\,p_{0,k}(t,a)\,. \tag{7.13}$$

The forces of infection for women (ϕ_1) and men (ϕ_2) are:

$$\phi_1(t,a) \;=\; \beta(a)\int_0^\infty \beta(b)\,f(a,b)\,\frac{p_{12}(t,b)}{\left[p_{02}(t,b)+p_{12}(t,b)\right]}\,db\,, \tag{7.14}$$

$$\phi_2(t,a) \;=\; \beta(a)\int_0^\infty \beta(b)\,f(a,b)\,\frac{p_{11}(t,b)}{\left[p_{01}(t,b)+p_{11}(t,b)\right]}\,db\,. \tag{7.15}$$

The forces of infection ϕ_1 and ϕ_2 no longer model proportionate mixing, the type of mixing modeled by the force of infection defined in Sect. 6.1.1. Instead, we use a form of mass-action, including the probability that a contact is with an infective.

7.4.1 The master equation.

We start by deriving the master equation of the model outlined above:

$$\begin{aligned}
\frac{d}{dt}P\left(\{N_\lambda\},t\right) \;&=\; \sum_{i=1:5} T_i\,,\\
&=\; \sum_{\lambda_2,\lambda_1} B_{\lambda_2,\lambda_1}\left(E_{\lambda_2}^{-1}E_{\lambda_1}-1\right)N_{\lambda_1}\,P\\
&\quad+\sum_{\lambda_2,\lambda_1} \mu_{\lambda_2,\lambda_1}\left(E_{\lambda_2}^{-1}E_{\lambda_1}-1\right)N_{\lambda_1}\,P\\
&\quad+\sum_{\lambda_2,\lambda_1} \eta_{\lambda_2,\lambda_1}\left(E_{\lambda_2}^{-1}E_{\lambda_1}-1\right)N_{\lambda_1}\,P\\
&\quad+\Delta^{-1}\sum_{\lambda_2,\lambda_1} \gamma_{\lambda_2,\lambda_1}\left(E_{\lambda_2}^{-1}E_{\lambda_1}-1\right)N_{\lambda_1}P\\
&\quad+\sum_{\lambda_2,\lambda_1} I_{\lambda_2,\lambda_1}\,\phi_{\lambda_1}\left(E_{\lambda_2}^{-1}E_{\lambda_1}-1\right)N_{\lambda_1}\,P\,,
\end{aligned} \tag{7.16}$$

where:

- $P\left(\{N_\lambda\},t\right)$ is a joint probability distribution for the number of individuals in each category.

- $T_1 : T_5$ describe events which can be grouped into three categories, namely those corresponding to: 1) creation of individuals (birth, immigration), 2), transfer between cohorts (aging, maturity, infection) and 3) removal from the system (death, emigration). It is convenient to model all events in this model as transfer events, even when it does not seem appropriate for some processes. For example, birth can be modeled at transfer from an infinite resource of individuals into the population. Mortality can be modeled as transfer into a category keeping track of those who died. $I_{\lambda_2,\lambda_1} = 1$ when transfer from susceptible λ_1 to infected λ_2 is possible, and is zero otherwise.

- The use of just one label, namely λ, to model cohorts N_λ is for reasons of economy of notation. An extra label could be used to indicate whether the cohort represents men, women, susceptibles, infections, and so on. While formulating the master equation, these sub-types are handled by rate functions. The macroscopic equations for the system, as well as equations for the averages and correlation between fluctuations are derived shortly. In these derivations additional labels are introduced to represent different categories of individuals, such as those with a different disease status.

- The operators E_λ and E_λ^{-1} are 'raising' and 'lowering' operators respectively. They increment or decrement a function g (e.g. a population) by 1:

$$E(g) = g + 1\,,$$
$$E^{-1}(g) = g - 1\,.$$

  As an example of its use consider an application of the operator $\left(E_{\lambda_2}^{-1} E_{\lambda_1}\right)$:

$$\left(E_{\lambda_2}^{-1} E_{\lambda_1}\right) N_{\lambda_1} P\left(\{N_{\lambda_1}, N_{\lambda_2}, \cdots\}\right)$$
$$N_{\lambda_1+1} P\left(\{N_{\lambda_1+1}, N_{\lambda_2-1}, \cdots\}\right)$$

  The step operators obey a useful identity [107, Chp 6.3]:

$$\sum_{n=0}^{\infty} g(n)\, E\, f(n) = \sum_{n=1}^{\infty} f(n)\, E^{-1}\, g(n)\,, \tag{7.17}$$

  which we need in order to manipulate expressions featuring step operators.

The five terms in Eq. (7.16) are:

- T_1: This term models birth into cohort λ. The birth rate B_λ can be proportional to N_λ but here we use a constant birth rate. It is modeled as transfer from an infinite source. When deriving the approximating FPE associated with (7.16), it will be convenient to model all rates as transfers between suitable defined compartments.

- T_2: Age-dependent mortality is modeled by μ_λ.

- T_3: η_λ is the age-dependent rate at which young individuals mature into eligibility.
- T_4: This term models aging as a jump process. Individuals spend an average time of $\frac{\Delta}{\gamma}$ in each category. This approximation of the aging process is accurate when the length of the age interval is small. We have thus far discritized age into one year bins. Age bins of one year can only be used in this approach if each age bin contains many individuals. Hence, there is a compromise between making the age bins small enough to model aging accurately as a jump event (with an average waiting time), while at the same time having many individuals in the age bin.
- T_5: This term corresponds to transfers from susceptible cohorts (λ_1) to infective cohorts (λ_2) due to infection. The difficulty lies with the transfer coefficient, which is the rate at which a susceptible individuals become infected:

$$\phi_{\lambda_1} = \sum_{\lambda_3} \beta(\lambda_1)\beta(\lambda_3) f(\lambda_1, \lambda_3) \frac{\tilde{N}_{\lambda_3}}{N_{\lambda_3} + \tilde{N}_{\lambda_3}} .$$

where the number of infectives in age category λ is denoted by $\tilde{N}_\lambda$.

7.4.2 Expansion of the age-structured master equation.

Master equation (7.16) can be transformed into a Fokker-Planck equation, using an expansion technique. The variables N_λ, i.e. the number of individuals in cohort λ, is transformed to:

$$N_\lambda = \Delta\theta_\lambda + \Delta^{\frac{1}{2}} y_\lambda \,, \tag{7.18}$$

where θ_λ are time-dependent functions unique to each cohort λ. According to this assumption, N_λ is composed of a macroscopic part proportional to Δ and a fluctuating part, proportional to $\Delta^{\frac{1}{2}}$. It is assumed that the age interval Δ, which defines the age interval of each cohort, is small but not large enough to contain many population members, i.e. that each N_λ is large. The distribution $P(\{N_\lambda\}, t)$ is transformed into joint probability distribution for the fluctuations $\Pi(\{y_\lambda\}, t)$ [104, Sect. 8]:

$$P\left(\left\{\Delta\theta_\lambda + \Delta^{\frac{1}{2}} y_\lambda\right\}, t\right) = \Delta^{-\frac{\Omega}{2}} \Pi(\{y_\lambda\}, t) \,.$$

In this new description, the step operators used take on the form:

$$\begin{aligned} E_\lambda g\left(y_\lambda\right) &= g\left(y_\lambda + \Delta^{\frac{1}{2}}\right) \\ &= \left(1 + \Delta^{-\frac{1}{2}}\frac{\partial}{\partial y_\lambda} + \frac{\Delta^{-1}}{2!}\frac{\partial^2}{\partial y_\lambda^2} + ...\right) g(y_\lambda)\,, \\ E_\lambda^{-1} g\left(y_\lambda\right) &= g\left(y_\lambda - \Delta^{\frac{1}{2}}\right) \\ &= \left(1 - \Delta^{-\frac{1}{2}}\frac{\partial}{\partial y_\lambda} + \frac{\Delta^{-1}}{2!}\frac{\partial^2}{\partial y_\lambda^2} + ...\right) g(y_\lambda)\,. \end{aligned}$$

Using these notations, the master Eq. (7.16) can be transformed into:

$$\frac{\partial \Pi}{\partial t} - \Delta^{1/2}\sum_\lambda \frac{d\theta_\lambda}{dt}\frac{\partial \Pi}{\partial y_\lambda} = \sum_{i=1:5} T_i\,, \tag{7.19}$$

where

$$\begin{aligned} T_1 &= \sum_{\lambda_2,\lambda_1} B_{\lambda_2,\lambda_1}\left\{\Delta^{-1/2}\left(\frac{\partial}{\partial y_1} - \frac{\partial}{\partial y_2}\right) + \frac{1}{2}\Delta^{-1}\left(\frac{\partial}{\partial y_1} - \frac{\partial}{\partial y_2}\right)^2\right\}\left[\Delta\theta_{\lambda_1} + \Delta^{1/2} y_{\lambda_1}\right]\Pi\,, \\ T_2 &= \sum_{\lambda_2,\lambda_1} \mu_{\lambda_2,\lambda_1}\left\{\Delta^{-1/2}\left(\frac{\partial}{\partial y_1} - \frac{\partial}{\partial y_2}\right) + \frac{1}{2}\Delta^{-1}\left(\frac{\partial}{\partial y_1} - \frac{\partial}{\partial y_2}\right)^2\right\}\left[\Delta\theta_{\lambda_1} + \Delta^{1/2} y_{\lambda_1}\right]\Pi\,, \\ T_3 &= \sum_{\lambda_2,\lambda_1} \eta_{\lambda_2,\lambda_1}\left\{\Delta^{-1/2}\left(\frac{\partial}{\partial y_1} - \frac{\partial}{\partial y_2}\right) + \frac{1}{2}\Delta^{-1}\left(\frac{\partial}{\partial y_1} - \frac{\partial}{\partial y_2}\right)^2\right\}\left[\Delta\theta_{\lambda_1} + \Delta^{1/2} y_{\lambda_1}\right]\Pi\,, \\ T_4 &= \Delta^{-1}\sum_{\lambda_2,\lambda_1} \gamma_{\lambda_2,\lambda_1}\left\{\Delta^{-1/2}\left(\frac{\partial}{\partial y_1} - \frac{\partial}{\partial y_2}\right) + \frac{1}{2}\Delta^{-1}\left(\frac{\partial}{\partial y_1} - \frac{\partial}{\partial y_2}\right)^2\right\}\left[\Delta\theta_{\lambda_1} + \Delta^{1/2} y_{\lambda_1}\right]\Pi\,, \\ T_5 &= \sum_{\lambda_2,\lambda_1} I_{\lambda_2,\lambda_1}\sum_{\lambda_3}\beta(\lambda_1)\beta(\lambda_3) f(\lambda_1,\lambda_3)\left\{T_a + \Delta^{-\frac{1}{2}} T_b\right\}\times \\ &\quad \left\{\Delta^{-1/2}\left(\frac{\partial}{\partial y_{\lambda_1}} - \frac{\partial}{\partial y_{\lambda_2}}\right) + \frac{1}{2}\Delta^{-1}\left(\frac{\partial}{\partial y_{\lambda_1}} - \frac{\partial}{\partial y_{\lambda_2}}\right)^2\right\}\left[\Delta\theta_{\lambda_1} + \Delta^{1/2} y_{\lambda_1}\right]\Pi\,. \end{aligned}$$

The prevalence of infection in T_5 in cohort λ is written as:

$$\begin{aligned} \frac{\tilde{N}_{\lambda_3}}{N_{\lambda_3} + \tilde{N}_{\lambda_3}} &= \frac{\Delta\tilde{\theta}_{\lambda_3} + \Delta^{\frac{1}{2}}\tilde{y}_{\lambda_3}}{\Delta\theta_{\lambda_3} + \Delta^{\frac{1}{2}} y_{\lambda_3} + \Delta\tilde{\theta}_{\lambda_3} + \Delta^{\frac{1}{2}}\tilde{y}_{\lambda_3}} \\ &= \frac{\tilde{\theta}_{\lambda_3} + \Delta^{-\frac{1}{2}}\tilde{y}_{\lambda_3}}{\theta_{\lambda_3} + \tilde{\theta}_{\lambda_3}}\left[1 - \Delta^{-\frac{1}{2}}\left(\frac{y_{\lambda_3} + \tilde{y}_{\lambda_3}}{\theta_{\lambda_3} + \tilde{\theta}_{\lambda_3}}\right) + \Delta^{-1}\left(\frac{y_{\lambda_3} + \tilde{y}_{\lambda_3}}{\theta_{\lambda_3} + \tilde{\theta}_{\lambda_3}}\right)^2 + \cdots\right] \\ &= T_a + \Delta^{-\frac{1}{2}} T_b\,, \end{aligned}$$

where

$$\begin{aligned} T_a &= \frac{\tilde{\theta}_{\lambda_3}}{\theta_{\lambda_3}+\tilde{\theta}_{\lambda_3}} \\ T_b &= \frac{\theta_{\lambda_3}\tilde{y}_{\lambda_3}-\tilde{\theta}_{\lambda_3}y_{\lambda_3}}{\left(\theta_{\lambda_3}+\tilde{\theta}_{\lambda_3}\right)^2}\,, \end{aligned}$$

keeping only those terms that are lower in order than $\Delta^{-\frac{1}{2}}$, which are the only terms that contribute to the FPE derived from this master equation.

Macroscopic equation. The macroscopic equation associated with this system is obtained by equating terms of order $\Delta^{\frac{1}{2}}$ on the left and right hand side of Eq. (7.19):

$$\begin{aligned} -\frac{d}{dt}\theta^y_\lambda &= \frac{\gamma_{\lambda+1,\lambda}}{\Delta}\theta^y_\lambda - \frac{\gamma_{\lambda,\lambda-1}}{\Delta}\theta^y_{\lambda-1} - B_\lambda + \mu_\lambda\theta^y_\lambda + \eta_\lambda\theta^y_\lambda\,, \\ -\frac{d}{dt}\theta^e_\lambda &= \frac{\gamma_{\lambda+1,\lambda}}{\Delta}\theta^e_\lambda - \frac{\gamma_{\lambda,\lambda-1}}{\Delta}\theta^e_{\lambda-1} + \mu_\lambda\theta^e_\lambda - \eta_\lambda\theta^e_\lambda + \Lambda(\lambda)\theta^e_\lambda\,, \\ -\frac{d}{dt}\tilde{\theta}^e_\lambda &= \frac{\gamma_{\lambda+1,\lambda}}{\Delta}\theta^e_\lambda - \frac{\gamma_{\lambda,\lambda-1}}{\Delta}\theta^e_{\lambda-1} + \mu_\lambda\theta^e_\lambda - \Lambda(\lambda)\theta^e_\lambda\,, \end{aligned}$$

where $\Lambda(\lambda)=\sum_\kappa \beta(\lambda)\beta(\kappa)f(\lambda,\kappa)\frac{\tilde{\theta}^e_\kappa}{\theta^e_\kappa+\tilde{\theta}^e_\kappa}$ is the force of infection experienced by individuals in age category λ. In continuous notation, which is the limit where $\Delta\to 0$, these equations would give the following set of PDEs:

$$\begin{aligned} \frac{\partial}{\partial t}n^y(t,a) &= -\frac{\partial}{\partial a}n^y(t,a) - \mu(a)\,n(t,a) - \eta^y(a)\,n(t,a)\,, \\ \frac{\partial}{\partial t}n^e(t,a) &= -\frac{\partial}{\partial a}n^e(t,a) - \mu(a)\,n(t,a) + \eta^y(a)\,n(t,a) - \beta(a)\int_0^\infty \beta(b)f(a,b)\frac{n(t,a)\,\tilde{n}(t,b)}{n(t,b)+\tilde{n}(t,b)}\,db\,, \\ \frac{\partial}{\partial t}\tilde{n}^e(t,a) &= -\frac{\partial}{\partial a}\tilde{n}^e(t,a) - \mu(a)\,\tilde{n}(t,a) + \beta(a)\int_0^\infty \beta(b)f(a,b)\frac{n(t,a)\,\tilde{n}(t,b)}{n(t,b)+\tilde{n}(t,b)}\,db\,, \end{aligned}$$

where $n(a)=\frac{N_a}{\Delta}$ is the density of individuals in age interval a. The 'boundary' condition is $n^y(t,0)=B$. These equations are of naturally the same as the macroscopic Eqs. (7.10)-(7.15) we defined initially.

7.4.3 Fokker-Planck equation.

Equating terms of the order Δ^0 on the left and right hand side of Eq. (7.19) gives the multivariate Fokker-Planck equation:

$$
\begin{aligned}
\frac{\partial \Pi}{\partial t} \;=\; & \sum_{\lambda_2,\lambda_1} B_{\lambda_2,\lambda_1}\left(\frac{\partial}{\partial y_{\lambda_1}}-\frac{\partial}{\partial y_{\lambda_2}}\right) y_{\lambda_1}\Pi + \frac{1}{2}\sum_{\lambda_2,\lambda_1} B_{\lambda_2,\lambda_1}\left(\frac{\partial}{\partial y_{\lambda_1}}-\frac{\partial}{\partial y_{\lambda_2}}\right)^2 \theta_{\lambda_1}\Pi \\
& + \sum_{\lambda_2,\lambda_1} \mu_{\lambda_2,\lambda_1}\left(\frac{\partial}{\partial y_{\lambda_1}}-\frac{\partial}{\partial y_{\lambda_2}}\right) y_{\lambda_1}\Pi + \frac{1}{2}\sum_{\lambda_2,\lambda_1} \mu_{\lambda_2,\lambda_1}\left(\frac{\partial}{\partial y_{\lambda_1}}-\frac{\partial}{\partial y_{\lambda_2}}\right)^2 \theta_{\lambda_1}\Pi \\
& + \sum_{\lambda_2,\lambda_1} \eta_{\lambda_2,\lambda_1}\left(\frac{\partial}{\partial y_{\lambda_1}}-\frac{\partial}{\partial y_{\lambda_2}}\right) y_{\lambda_1}\Pi + \frac{1}{2}\sum_{\lambda_2,\lambda_1} \eta_{\lambda_2,\lambda_1}\left(\frac{\partial}{\partial y_{\lambda_1}}-\frac{\partial}{\partial y_{\lambda_2}}\right)^2 \theta_{\lambda_1}\Pi \\
& + \frac{1}{\Delta}\sum_{\lambda_2,\lambda_1} \gamma_{\lambda_2,\lambda_1}\left(\frac{\partial}{\partial y_{\lambda_1}}-\frac{\partial}{\partial y_{\lambda_2}}\right) y_{\lambda_1}\Pi + \frac{1}{2\Delta}\sum_{\lambda_2,\lambda_1} \gamma_{\lambda_2,\lambda_1}\left(\frac{\partial}{\partial y_{\lambda_1}}-\frac{\partial}{\partial y_{\lambda_2}}\right)^2 \theta_{\lambda_1}\Pi \\
& + \sum_{\lambda_2,\lambda_1}\left[I_{\lambda_2,\lambda_1}\sum_{\lambda_3}\beta(\lambda_1)\beta(\lambda_3)f(\lambda_1,\lambda_3)\,T_a\right]\left(\frac{\partial}{\partial y_{\lambda_1}}-\frac{\partial}{\partial y_{\lambda_2}}\right) y_{\lambda_1}\Pi \\
& + \sum_{\lambda_2,\lambda_1}\left[I_{\lambda_2,\lambda_1}\sum_{\lambda_3}\beta(\lambda_1)\beta(\lambda_3)f(\lambda_1,\lambda_3)\,T_b\right]\left(\frac{\partial}{\partial y_{\lambda_1}}-\frac{\partial}{\partial y_{\lambda_2}}\right) \theta_{\lambda_1}\Pi \\
& + \frac{1}{2}\sum_{\lambda_2,\lambda_1}\left[I_{\lambda_2,\lambda_1}\sum_{\lambda_3}\beta(\lambda_1)\beta(\lambda_3)f(\lambda_1,\lambda_3)\,T_a\right]\left(\frac{\partial}{\partial y_{\lambda_1}}-\frac{\partial}{\partial y_{\lambda_2}}\right)^2 \theta_{\lambda_1}\Pi\,. \qquad (7.20)
\end{aligned}
$$

This equation depends on time through θ_{λ_1}, T_a and T_b. The solution of the FPE is a multivariate Gaussian distribution, with the following mean and variance.

First moments. The equations for the first moments (averages) are obtained by multiplying (7.20) by y_α and integrating over all y. As an example of the algebra required consider [104, Sect. 5]:

$$
\sum_{\lambda_2,\lambda_1} \eta_{\lambda_2,\lambda_1}\left(\frac{\partial}{\partial y_{\lambda_2}}-\frac{\partial}{\partial y_{\lambda_1}}\right) y_{\lambda_1}\Pi\,,
$$

multiply with y_α, apply (7.17) in order obtain the same form as that given by Eq. (3.8):

$$
\sum_{\lambda_2,\lambda_1} \eta_{\lambda_2,\lambda_1}\left(\delta_{\lambda_2,\alpha}\right) y_{\lambda_1}\,\Pi - \sum_{\lambda_2,\lambda_1} \eta_{\lambda_2,\lambda_1}\left(\delta_{\lambda_1,\alpha}\right) y_{\lambda_1}\,\Pi\,,
$$

and integrate over y:

$$
\sum_{\lambda_1} \eta_{\alpha,\lambda_1}\langle y_{\lambda_1}\rangle - \sum_{\lambda_2} \eta_{\lambda_2,\alpha}\langle y_\alpha\rangle\,.
$$

Repeating this for the other terms shows that the average fluctuations $\langle y\rangle$ evolves according to:

$$\begin{aligned}\frac{d}{dt}\langle y_\alpha\rangle &= \sum_{\lambda_1}\left[B_{\alpha,\lambda_1}+\frac{\gamma_{\alpha,\lambda_1}}{\Delta}+\mu_{\alpha,\lambda_1}+\eta_{\alpha,\lambda_1}+I_{\alpha,\lambda_1}\Lambda^a(\lambda_1)\right]\langle y_{\lambda_1}\rangle \\ &\quad -\sum_{\lambda_2}\left[B_{\lambda_2,\alpha}+\frac{\gamma_{\lambda_2,\alpha}}{\Delta}+\mu_{\lambda_2,\alpha}+\eta_{\lambda_2,\alpha}+I_{\lambda_2,\alpha}\Lambda^a(\alpha)\right]\langle y_\alpha\rangle \\ &\quad +\sum_{\lambda_1}\left[I_{\alpha,\lambda_1}\,\Lambda^b(\lambda_1)\right]\theta_{\lambda_1}-\sum_{\lambda_2}\left[I_{\lambda_1,\alpha}\,\Lambda^b(\alpha)\right]\theta_\alpha\,, \end{aligned} \tag{7.21}$$

where

- The transfer rates, which define transfer matrices $\left(B,\frac{\gamma}{\Delta},\mu,\eta,\Lambda^a,\Lambda^b\right)$, are used to allow transfer from one compartment to another only when applicable. For example, aging, which is approximated by the transfer term $\frac{\gamma_{\nu,\lambda}}{\Delta}$, allows transfer only from λ to $\nu=\lambda+1$. Mortality is also modeled as an event which transfers individuals from one compartment to another, which means keeping track of those who are removed from the model.

- $\Lambda^a(\alpha)$ and $\Lambda^b(\alpha)$ are 'force of infection' terms, i.e. the rate of transfer from susceptible to infected:

$$\Lambda^a(\alpha) = \sum_\kappa \beta(\alpha)\beta(\kappa)f(\alpha,\kappa)\,\frac{\tilde\theta_\kappa}{\theta_\kappa+\tilde\theta_\kappa}\,,$$

and

$$\Lambda^b(\alpha) = \sum_\kappa \beta(\alpha)\beta(\kappa)f(\alpha,\kappa)\,\frac{\theta_\kappa\langle\tilde y_\kappa\rangle-\tilde\theta_\kappa\langle y_\kappa\rangle}{\left(\theta_\kappa+\tilde\theta_\kappa\right)^2}\,.$$

- Equation (7.21) for the averages in the fluctuations y can be written in matrix form:

$$\frac{d}{dt}\langle y_\alpha\rangle=\sum_\nu \mathbb{A}_{\alpha,\nu}\,\langle y_\nu\rangle+\sum_\nu \mathbb{C}_{\alpha,\nu}\,\theta_\nu \tag{7.22}$$

where, setting $\vartheta_{\alpha,\nu}=B_{\alpha,\nu}+\frac{\gamma_{\alpha,\nu}}{\Delta}+\mu_{\alpha,\nu}+\eta_{\alpha,\nu}+I_{\alpha,\nu}\Lambda^a(\alpha)$ and $\Lambda^b_{\alpha,\nu}=I_{\alpha,\nu}\Lambda^b(\alpha)$, the matrices $\mathbb{A}$ and $\mathbb{C}$ are given by:

$$\begin{aligned}\mathbb{A}_{\alpha,\nu} &= \vartheta_{\alpha,\nu}-\delta_{\alpha,\nu}\sum_\kappa\vartheta_{\kappa,\alpha}\,,\text{and}\\ \mathbb{C}_{\alpha,\nu} &= \Lambda^b_{\alpha,\nu}-\delta_{\alpha,\nu}\sum_\kappa\Lambda^b_{\kappa,\alpha}\,.\end{aligned}$$

Second moments. Equations for the second moments (correlations) are obtained by multiplying (7.20) $y_\alpha y_\zeta$, and integrating over all y. The first five lines of Eq. (7.20) are reduced as follows. Multiply (7.20) with $y_\alpha\, y_\zeta$ and apply (7.17):

$$\sum_{\lambda_2,\lambda_1} \vartheta_{\lambda_2,\lambda_1}\, y_{\lambda_1} \left(\frac{\partial}{\partial y_{\lambda_2}} - \frac{\partial}{\partial y_{\lambda_1}} \right) y_\alpha\, y_\zeta\, \Pi\,.$$

Now apply the product rule when differentiating using $\frac{\partial}{\partial y_{\lambda_1}}, \frac{\partial}{\partial y_{\lambda_2}}$ to find:

$$\begin{aligned} &\sum_{\lambda_2,\lambda_1} \vartheta_{\lambda_2,\lambda_1}\, y_{\lambda_1}\, (\delta_{\lambda_2,\alpha} - \delta_{\lambda_1,\alpha})\; y_\zeta + \\ &\sum_{\lambda_2,\lambda_1} \vartheta_{\lambda_2,\lambda_1}\, y_{\lambda_1}\, (\delta_{\lambda_2,\zeta} - \delta_{\lambda_1,\zeta})\; y_\alpha\,, \end{aligned}$$

giving:

$$\begin{aligned} &\sum_{\lambda_1} \vartheta_{\alpha,\lambda_1} \langle y_{\lambda_1} y_\zeta \rangle - \sum_{\lambda_2} \vartheta_{\lambda_2,\alpha} \langle y_\alpha y_\zeta \rangle + \\ &\sum_{\lambda_1} \vartheta_{\zeta,\lambda_1} \langle y_{\lambda_1} y_\alpha \rangle - \sum_{\lambda_2} \vartheta_{\lambda_2,\zeta} \langle y_\zeta y_\alpha \rangle\,. \end{aligned}$$

Line 6 of Eq. (7.20) introduces an unusual term (compare for example with Eq. (3.8)) in the equation for the second moments, as it did in Eq. (7.22) for the first moments. Multiply this line in the FPE with $y_\alpha\, y_\zeta$ and apply identity (7.17):

$$\begin{aligned} &\sum_{\lambda_1} I_{\alpha,\lambda_1}\, \Lambda^b(\lambda_1,\zeta)\, \theta_{\lambda_1} - \sum_{\lambda_2} I_{\lambda_2,\alpha}\, \Lambda^b(\alpha,\zeta)\, \theta_\alpha + \\ &\sum_{\lambda_2} I_{\zeta,\lambda_1}\, \Lambda^b(\lambda_1,\alpha)\, \theta_{\lambda_1} - \sum_{\lambda_2} I_{\lambda_2,\zeta}\, \Lambda^b(\zeta,\alpha)\, \theta_\zeta\,, \end{aligned} \tag{7.23}$$

where $\Lambda^b(\alpha,\zeta)$ is the matrix

$$\Lambda^b(\alpha,\zeta) \;=\; \sum_\kappa \beta(\alpha)\beta(\kappa) f(\alpha,\kappa)\, \frac{\theta_\kappa \langle \tilde{y}_\kappa y_\zeta \rangle - \tilde{\theta}_\kappa \langle y_\kappa y_\zeta \rangle}{\left(\theta_\kappa + \tilde{\theta}_\kappa\right)^2}\,. \tag{7.24}$$

The time evolution of the fluctuations are given by the following equation:

$$\begin{aligned}\frac{d}{dt}\langle y_\alpha y_\zeta\rangle &= \sum_\nu \mathbb{A}_{\alpha,\nu}\langle y_\nu y_\zeta\rangle + \sum_\nu \mathbb{A}_{\zeta,\nu}\langle y_\alpha y_\nu\rangle + \mathbb{B}_{\alpha,\zeta} \\ &\quad + \sum_\nu \mathbb{D}_{\alpha,\nu}\Lambda^b(\nu,\zeta) + \sum_\nu \mathbb{D}_{\zeta,\nu}\Lambda^b(\alpha,\nu)\,, \end{aligned} \tag{7.25}$$

which is a general form of a multivariate FPE. The matrix $\mathbb{A}$ is defined in Eq. (7.22), the symmetric matrix $\mathbb{B}$ is given by

$$\mathbb{B}_{\alpha,\nu} = \left[B_{\alpha,\nu} + \frac{\gamma_{\alpha,\nu}}{\Delta} + \mu_{\alpha,\nu} + \eta_{\alpha,\nu} + I_{\alpha,\nu}\Lambda^a(\alpha)\right]\theta_{\lambda_\alpha}\,, \tag{7.26}$$

and the matrix $\mathbb{D}$ by

$$\mathbb{D}_{\alpha,\nu} = I_{\alpha,\nu}\theta_\nu - \delta_{\alpha,\nu}\sum_\kappa I_{\kappa,\alpha}\,\theta_\alpha\,.$$

Although the second line in Eq. (7.25) is not standard for FPEs, its use will result in a symmetric variance-covariance matrix Y for the Gaussian distributed fluctuations y as long as the matrix $\Lambda^b(\alpha,\zeta)$ is symmetric. If it is not symmetric, then the assumptions that the noise follow a Gaussian distribution is violated. The Gaussian assumption would still hold if $\Lambda^b(\alpha,\zeta)$ defined in Eq. (7.24) is small, and it would be if θ_λ is large for all λ.

If we define matrices $Y_{\alpha,\zeta} = \langle y_\alpha y_\beta\rangle$ and $X_{\alpha,\zeta} = \Lambda^b(\alpha,\zeta)$ then Eq. (7.25) can be written in matrix form [103, Sect. XIV]:

$$\frac{d}{dt}Y = \mathbb{A}Y + Y\mathbb{A}^T + \mathbb{D}X + X\mathbb{D}^T + \mathbb{B}\,. \tag{7.27}$$

The problem of understanding the properties of fluctuations in non-linear age-structured STD model, assuming they follow a Gaussian distribution, is reduced to one of building matrices $\mathbb{A}, \mathbb{B}$ and $\mathbb{D}$. This form is particularly convenient for use in a numeric solution for Eq. (7.25).

7.4.4 Simulation results

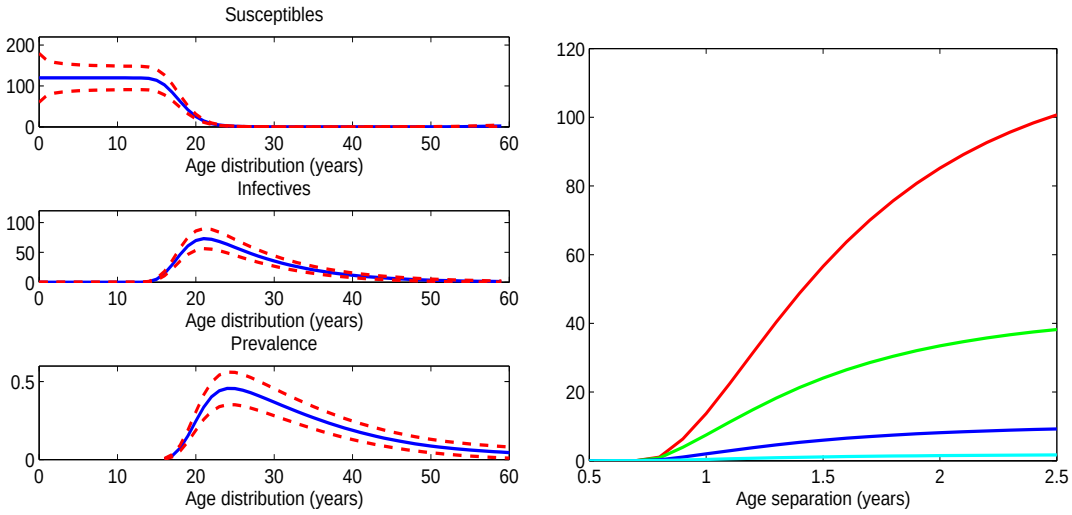
The Fokker-Planck approach is tested on a special case. We remove the process of first maturing before participating in random sexual contacts. This delay is already handled to some extent by the age-dependent risk function $\beta(a)$. We model only one gender. The rest of the parameters μ, β, f are defined in Sect. 6.1.1. A constant birth rate of $B = 200$ newborns per year is assumed.

The population is divided into one-year age boxes. This means that a parameter of $\Delta = 1$ is used in the expansion technique, and that there may not be a separation of scale between the number of population members and the size of its fluctuation, as assumed in

Eq. 7.18. However, an age interval of one year is large enough to include many population members. It is also long compared to the typical waiting time between different events in the model, which is of the order of a few days.

In Fig. 7.3(a) we show (in blue) total population numbers in each category at steady state. In red we show one standard deviation of the assumed Gaussian distributed steady-state fluctuations. Fig. 7.3(a)[bottom] shows the prevalence and its variation. It would be interesting to compare the Fokker-Planck predicted variance, with the variance between many stochastic realizations. The latter would require a considerable amount of time to generate, but such a comparison would help validate both approaches.

Fig. 7.3(b) shows the correlation between the numbers of 20-year-old infective cases and (25,30,35,40)-year-olds as a function of the standard deviation in the variance of the age difference between partners. It shows that correlation decreases with the age difference between parters, due to an assumed assortative partner-choice function. The effect of increasing the age difference between partners, is that it increases correlation among all age-categories. Therefore, this method provides further insight into the criticality of the age separation between partners in establishing a major epidemic.



(a) variation in age groups (b) correlation between age groups

Figure 7.3: (a) [top] number of susceptibles (in blue). [middle] number of infectives (in blue). [bottom] prevalence (in blue). One standard deviation around subpopulation in dashed red line. (b) correlation between infected 20-year-olds and infected 25- (in red), 30- (in green), 35- (in blue) and 40-year-olds (in cyan).

7.5 Conclusions

The master equation approach provides a flexible method for modeling fluctuations intrinsic to small populations. When the transition coefficients in the model (e.g. birth rate, mortality rate, relationship formation rate, and so on) are linear functions of the population state

variables, and when the underlying stochastic process is Markov, then the process can be studied by means of a Laplace transform technique.

We showed how the long-term behavior of the system is determined by the singularity (of the resolvent of the transition matrix) with the largest real part. The right eigenvector corresponding to this eigenvalue of the transition matrix is the long-term relative distribution of states. The exponential of an evolution or transition matrix, which we have used throughout, in fact takes into account all possible realizations or histories of events recognized in our models.

We adapted the Gillespie simulation approach to model age-structured partnering models, in order to validate some of the properties of the age-structured model of Ch. 6. We computed among other statistics, distributions for the time spent waiting between relationships, the lifetime-total of time spent in relationships as well as the number of lifetime steady relationships. The model of Ch. 6 is shown to be realistic, although that particular model is not calibrated to specific relationship data.

The Fokker-Planck approach provides an alternative method for studying fluctuations in age-structured models. The derivation of the Fokker-Planck equations is conceptually simple, although the algebra can be tedious. In addition, some terms in the derivation may break the symmetry requirements for Gaussian distributed fluctuations. These terms can safely be ignored if they are small, but when they are not, the variance around macroscopic values will not be symmetric. However, the method is systematic and leads to a Fokker-Planck equation defined in terms of a transition and diffusion matrix, which offers transparency and efficiency in estimating variance unequalled by stochastic simulation techniques.

Chapter 8

Conclusions

We started this work fascinated by the possibility that mathematical modeling could improve our understanding of HIV and TB epidemics in South Africa. Using a unique data set from a thoroughly studied HIV-TB epidemic in Masiphumelele (a township near Cape Town), we developed a simple but reliable model which not only confirmed our expectations but led to an unexpected result: ART treatment has the potential to control HIV transmission at a community level, indirectly also controlling the TB epidemic. To extend this analysis to HIV-TB epidemics at a national level, we developed models to account for demographical detail, adapting tools from fields as far apart as applied demography and statistical physics. The end result can best be described as a framework for integrating available epidemiological data, locally and nationally, with a toolkit of mathematical methods which can be continuously sharpened in accordance with the ever-changing picture of HIV and TB in South Africa.

We first developed what is arguably the simplest realistic model for HIV and TB. It recognizes 6 states: being HIV_- or HIV_+, and susceptible to TB or latently or actively infected with TB. Its 22 parameters were methodically obtained. The result was a model which fits the available data well, and which isolates HIV as a driving force behind increasing TB notification rates.

The sensitivity of the model was tested with respect to changes in critical parameters, i.e. those corresponding to non-linear terms in the model: TB transmission rates, reinfection parameters and parameters for HIV infection. The model showed no backward-bifurcation, which would require the reinfection rate to be higher than the primary infection rate. A bifurcation diagram of the steady-state TB notification and HIV prevalence rates shows how high levels of HIV can increase steady-state TB notification rates, even at fixed values for the TB transmission rate.

Masiphumele is of an intermediate size, 10,000: bigger than a household or a school where small-population effects are substantial, and smaller than a district or a country, where they can be ignored. We showed that small population effects are substantial in this model of 10,000 people, and would remain so up to a population of roughly 40,000 in size. Further, we showed that fluctuations grow (at what appears to be an exponential rate) during the

exponential growth phase of the TB epidemic. This is also the phase where macroscopic data were collected. Therefore, in addition to the various uncertainties already associated with the model and with observational data, we argue that uncertainty associated with finite population and transition effects should also be recognized when fitting deterministic models to available data.

We used the Fokker-Planck description of fluctuations to model the intensities of active TB events, for which we developed an accompanying point process theory. Using this theory, we computed two-time correlation functions to model the timescale of correlation between two successive active TB events. The exponentially distributed waiting times lead to longer than expected correlation times between active TB events. This improves upon the clinical definition of 'clusters' of active TB events, which currently relies only on DNA typing of TB strains, but lacks a temporal element.

Our analysis of a recent tuberculin skin test survey among schoolchildren in 2007 suggests that the annual risk of MTB infection in Masiphumelele is likely to be decreasing. The result, which can and should be verified using further MTB prevalence surveys, is unexpected for a community with an exploding number of possible sources of infection, in part due to rapid spread of HIV. We developed an age-structured HIV-TB model and used age-disaggregated HIV-TB data in Masiphumelele to help resolve a paradox regarding the role HIV_+ cases play in spreading TB. The data can be fitted reasonably well by assuming only rates in the HIV submodel to be age-dependent, while keeping rates in the TB submodel age-independent. The rise in the TB epidemic is clearly correlated to the rise in the HIV epidemic.

Is HIV therefore driving the TB epidemic? We constructed a hypothetical scenario where TB 'invades' a population in which HIV has already reached its endemic steady state. The reproductive value (i.e. the left eigenvector corresponding to the largest eigenvalue, R_0) of different TB cases suggests that HIV_+ cases are not in fact making a substantial contribution. This is due to their short infectious period. This conclusion hinges on the validity of the mass-action mixing assumption made in the model. Should this mixing assumption be verified empirically, our analysis suggests that simply scaling up DOTS (as opposed to restructuring DOTS by HIV disease staging criteria) should control the TB epidemic in Masiphumelele.

At this stage it is important to recognize the dichotomy that exists between the factors that drive the spread of the TB epidemic and the factors behind the high TB notification rate. Healthy individuals are responsible for the former, whereas immune-compromised individuals are responsible for the latter. The standard DOTS program would function well without augmentation or restructuring should an HIV intervention not only reduce the prevalence of HIV but also reconstitute the immune system and general health of individuals.

A universal test and treat strategy would do precisely that. We therefore developed a model to study HIV-TB dynamics at a national level, focusing on the impact of UTTS on HIV-TB epidemics. We assumed that the primary infection rate for HIV_+ cases on treatment is the same as that of HIV_- individuals. The effect of the intervention would be a

significant fall in the TB notification rate, back to levels of the pre-HIV era. Our models of HIV-TB are tailor-made to study a test trial of UTTS in a community like Masiphumelele, and to help guide its eventual deployment in South Africa as a whole.

The effect of generalized ART on the spread of HIV may be profound, but we did not neglect the potential impact of conventional interventions, such as discouraging intergenerational sex. We developed a model which combined basic demography with dynamics of steady relationships and risky random sexual contacts. The basic reproduction number is computed as a function of the standard deviation of the normally distributed partner choice. This shows that a critical level of variance in the age difference between partners of around 1.5-2.5 years is required to establish a major HIV epidemic. The relationship between R_0 and age separation depends strongly on the assumptions we make about relationship dynamics. For example, concurrency and/or an increased rate of random contacts would reduce the criticality of age separation.

A national UTTS program is likely to create renewed discussion around the issue of age variance between partners. It is well known that more women in need of ART receive it than men, since they are more frequently tested for HIV through antenatal programs. This is likely to have a significant effect on partner availability, which in turn might act in synergy with the effect of intergenerational sex. It could make young women even more vulnerable to infection. Understanding the subtle consequences of intergenerational sex is likely to be vital in creating an environment for an HIV-free generation. Our UTTS model for South Africa, which is based on data from antenatal clinics, currently assumes that women on ART do not change their sexual behavior. If data in this direction were to become available, it would be interesting to adapt our model in order to investigate this dynamic in more detail.

Our understanding of an epidemic at the population level can be improved through studying individual-based models. Micro-simulation techniques can shed light on relationship and infection dynamics in age-structured models. We cannot identify a microscopic system uniquely from a macroscopic description, because identity is lost in the averaging procedure. However, when the macroscopic description is translated into a collection of Poisson processes, we have a 'canonical' description at an individual level. Poisson processes allow us to build models mechanically and simulate population dynamics. Using a carefully designed bookkeeping device we can keep track of the histories of all the individuals in the model, and study the histories statistically by constructing empirical distributions (histograms).

The Fokker-Planck approach has thus far not been widely used to model fluctuations in epidemic models. We demonstrated its use in an age-independent model for TB and an age-structured model for HIV. This approach provides a handle on both deterministic and stochastic epidemic modeling. Put differently, it provides a direct link between dynamical assumptions in epidemic models and empirical data on infection events. An exiting and largely unexplored next step in epidemiological modeling would be validation of mixing assumptions against real data. The Fokker-Planck approach to modeling time-dependent

trends in individual infection events is ideally suited for this purpose.

The importance of taking epidemiological modeling from the realm of the purely theoretical and applying it to real-world situations cannot be overstated. Although this does limit the range of suitable mathematical tools, it is clear that even fairly simple models based on limited available data can lead to valuable insights. In this project we have cherry-picked the most useful mathematical tools and applied them to the most complete data in HIV and TB available in South Africa, in order to better understand HIV and TB epidemics both separately and in their complex interaction with each other. This will remain a work in progress for as long as these epidemics endure.

Appendix A

List of abbreviations

Abbreviation	Description
AIDS	acquired immune deficiency syndrome
ART	Antiretroviral therapy
CD_4	T cell co-receptor
CSW	commercial sex workers
DNA	deoxyribonucleic acid
DOTS	Directly Observed Treatment Short-course
FPE	Fokker Planck equation
HIV	Human immunodeficiency virus
HIV_-	HIV negative
HIV_+	HIV positive
MTB	Mycobacterium tuberculosis
NGM	next-generation matrix
ODE	ordinary differential equation
PDE	partial differential equation
R_0	basic reproduction value

Abbreviation	Description
RNA	ribonucleic acid
STD	sexually transmitted disease
SI	Susceptible Infected
SIR	Susceptible Infected Removed
TB	active infection with MTB
TST	tuberculin skin test
UTTS	universal test and treat strategy
WHO	World Health Organization
Ch.	chapter
Fig.	figure
Sect.	section
Tab.	table

Table A.1: List of abbreviations continued.

Appendix B

An introduction to population dynamics and epidemiological modeling

In this chapter we present certain aspects of mathematical modeling which are central to this project, without going into too much technical detail. Recognizing and building relevant heterogeneity into epidemiological models has become a focus of research, especially from a theoretical point of view. The subject has remained largely theoretical because it is not always obvious that heterogeneity is required to fit data, particularly in large populations or in communities where exceptionally high levels of disease can mask the effect of heterogeneity. For example, the sheer magnitude of the TB epidemic in Masiphumelele may dwarf any age-specific difference in infection [92]. However, we found age-structure and age- mediated heterogeneity to be vitally important in the quest for a better understanding of the impact of HIV and TB epidemics in South Africa. We found the next-generation matrix framework adequate for the purpose of accommodating age structure, allowing us to calculate the basic reproduction number (R_0) in more complex and realistic models.

B.1 Compartmental models

When dealing with deterministic models it is helpful to identify compartments that would capture relevant structure. Invariably the structure is complex, requiring many compartments to capture the heterogeneities of the process. Corresponding to each compartment is a differential equation, usually a function of time only, where the rates of change of each of the compartments are expressed as rates of transfers in and out of compartments. A typical system of compartments and equations for an epidemic model could be:

$$\dot{x}(t) = (M + N - U)\ x(t) \tag{B.1}$$

where M_{ij} is the rate at which infectives of type j produce infectives of type i, N is a matrix of transfers between states and U a diagonal matrix capturing mortality. It is usually straightforward to solve this system of equations numerically, and it is often not helpful to develop methods to solve these equations analytically. Modelers are more interested in the stability properties of the disease-free or the endemic equilibrium point.

B.2 The basic reproduction number, R_0

In its simplest form in a conceptual model, R_0 for communicable diseases is typically defined as $R_0 = \kappa\,\vartheta\,\tau$, where κ is the rate at which infectious contacts are made, ϑ the probability of transmission upon contact and τ the duration of the infectious period. In models that recognize heterogeneity the statement regards a typical infective, and to define a typical infective one must average over the space on which it is 'distributed'.

The following is a useful algorithm defining the next-generation matrix and R_0 as developed in [36, Ch 5.2]. We make use of this formalism to study the impact of HIV on the spread of TB (Sect. 4.4) and to study the impact of age-dependent partner choice on disease invasion in a two-sex STD model (Sect. 6.3). A generic form of the NGM for age structured models of infection is derived in Sect. 5.1.

Let the $K(\xi, \eta)$ be the expected number of new infections of type ξ caused by an infective of type η. The individual types/states ξ, η are distributed over some domain $D(\eta)$. These individual states are meant to characterize each individual. A generation of infectives, distributed over a state in $D(\eta)$ according to some distribution $\phi(\eta)$, is defined as:

$$\int_{D(\eta)} \phi(\eta)\, d\eta \tag{B.2}$$

which is equal to the number of cases with states belonging to $D(\eta)$. The next-generation operator $\mathbb{K}$, an integral operator which "maps the current generation of susceptibles onto the next generation of such susceptibles" [36, Ch. 5], can be defined as:

$$(\mathbb{K}\phi)(\xi) = \int_{D(\eta)} K(\xi, \eta)\, \phi(\eta)\, d\eta \tag{B.3}$$

R_0 is now defined to be the spectral radius of the kernel K:

$$R_0 = \lim_{n\to\infty} \|K^n\|^{\frac{1}{n}} \tag{B.4}$$

A short review of the spectrum of a linear operator can be found in [91]. Using this formalism to study Eq. (B.1), for example, we see that the next-generation matrix is:

$$K = \int_0^\infty M\, e^{(N-U)\tau}\, d\tau = T\,(N-U)^{-1} \tag{B.5}$$

which is simply a product of the rate of producing infectives (M) multiplied by the duration of infection of the various infectious types, $(N-U)^{-1}$. Thus:

$$R_0 = \text{spectral radius}\left\{M\,(N-U)^{-1}\right\} \tag{B.6}$$

B.3 Mathematical demography

The linearized epidemic (i.e. the early epidemic) can usually be studied as a linear integral equation, which is analogous to the characteristic equation of demography ([15] and [36, Sect. 6.2]). The correspondence is between the time since infection in epidemic models and chronological age in demography. This means that some of the successful tools used in demography, such as the basic reproduction number and the reproduction value, are available as potential tools for epidemiology.

In this section we study a simple continuous-time demographical model for an age-structured population. This is a female-only model which contains the most important features of age-structured population systems. Using this linear model we will discuss a few familiar techniques for solving linear problems: obtaining solutions along characteristic lines, studying adjoining integral equations (the so-called renewal equation) and performing Laplace transforms.

Many of the principal ideas of population dynamics can be illustrated using the following model for age-dependent population growth:

$$\frac{\partial n(t,a)}{\partial t} \;=\; -\frac{\partial n(t,a)}{\partial a} - \mu(t,a)n(t,a)\,, \tag{B.7}$$

where

- $n(t,a)$ is the density of a population with respect to age a and time t.
- Age-dependent mortality is given by $\mu(a)$. The probability of survival in an age interval (a_1, a_2) is given by

$$l(a_1,a_2) = e^{-\int_{a_1}^{a_2}\mu(a')\,da'}\,. \tag{B.8}$$

  .

- Birth rate enters as a 'boundary condition':

$$n(t,0) = B(t) = \int_0^\infty \beta(a')n(t,a')\,da', t > 0\,. \tag{B.9}$$

- $\beta(a)$, the maternity function, is the expected number of offspring per individual aged a, in the next time interval $(t, t+dt)$.
- This is a model for a single-sex self-generating population. In demographical models it is mostly the female population that is used. This is partly because the states (mostly age) of women are more carefully documented in birth registers than those of men, and partly because women have a narrower fertility range (typically 15-50 years) [65].

The solution to (B.7) is given by [114]:

$$n(t,a) = \begin{cases} n(t-a,0)e^{-\int_0^a \mu(a')\,da'} & a < t \\ n(0,a-t)e^{-\int_{a-t}^a \mu(a')\,da'} & a \geq t \end{cases}\,. \tag{B.10}$$

It is seen that birth rate $B(t)$ satisfies the renewal equation [114]:

$$B(t) \;=\; \int_0^\infty b(a)B(t-a)\,da + B_0 \tag{B.11}$$

where $b(a) = \beta(a)l(a)$ is the maternity schedule. B_0 is the contribution to the birth rate from the population that existed at time $t = 0$. For simplicity it may be assumed that $B_0 = 0$, i.e that initial population given by the second integral in (B.11) has passed away.

Exponential population growth. Assuming exponential growth, Lotka showed that

$$n(t,a) \;=\; e^{rt}\mathbf{A}(a)\,, \tag{B.12}$$

and that boundary condition (B.9) implies:

$$1 \;=\; \int_0^\infty e^{-ra}\beta(a')l(a')\,da' = g(r)\,, \tag{B.13}$$

known as a characteristic equation for the asymptotic growth rate r. It has an infinite number of roots for a general maternity schedule $\beta(a)l(a)$ [65, Chp 7]. The asymptotic growth rate is determined by the unique positive real root, where uniqueness follows from the fact that $g(r)$ is a monotonically decreasing function of r. The number

$$R_0 = g(0) = \int_0^\infty m(a')l(a')\,da'\,, \tag{B.14}$$

is called the basic reproduction number of an individual. It is the expected number of offspring of the individual during her lifetime and is a critical parameter: whether population growth is positive, negative or zero (in which case the population size fluctuates stochastically), depends on whether $R_0 > 1$, $R_0 < 1$ and $R_0 = 1$ respectively.

Fisher had an insightful interpretation of (B.13). He thought that a female child is given, at birth, a loan which she must 'repay' by in turn giving birth to offspring. [65, Ch. 8.1]. The probability of giving birth in the next age interval da is $l(a)m(a)da$. The value of this offspring is discounted to $e^{-ra}l(a)m(a)da$. If the debt is 1 at birth, then:

$$1 = \int_{15}^{50} e^{-ra}l(a)m(a)\,da\,.$$

He then asks how much of the debt is outstanding at age $x < 50$. He shows that it is:

$$\begin{aligned} V(x) &= \int_x^{50} e^{-r(a-x)}\frac{l(a)}{l(x)}m(a)\,da \\ &= \frac{1}{e^{-rx}l(x)}\int_x^{50} e^{-ra}l(a)m(a)\,da\,, \end{aligned}$$

with $V(0) = 1$. Thus $V(x)$ can be seen as the discounted contribution an individual aged x makes to future generations. The reproduction value of the individual is proportional to $V(x)$. Its contribution, on average, to future births is given by $\frac{V(x)}{\chi}$, where χ is the mean age of giving birth:

$$\chi = \int_0^\infty a'm(a')l(a')\,da'\,.$$

Analysis of the birth rate. We now have a closer look at the birth rate, along the same lines as Hoppensteadt [55, Sect. 2.2], analyzing it by means of the Laplace transform. Define the Laplace transform:

$$\mathfrak{L}\,(B)\,(r) = \hat{B}(r) = \int_0^t e^{-rt}B(t)\,dt\,,$$

so that (B.11) becomes:

$$\begin{aligned}\hat{B}(r) &= \hat{B}_0(r) + \hat{b}(r)\hat{B}(r) \\ \hat{B}(r) &= \hat{B}_0(r) \Big/ \left[1 - \hat{b}(r)\right] .\end{aligned}$$

Hence the solution of the birth rate is given by:

$$B(t) = \frac{1}{2\pi i} \int_{r^*-i\infty}^{r^*-i\infty} e^{rt} \hat{B}_0(r) \left[1 - \hat{b}(r)\right]^{-1} dr \,. \tag{B.15}$$

The birth rate $B(t)$ can be written as:

$$B(t) = \sum_{i=1}^{\infty} c_i e^{r_i t} \,, \tag{B.16}$$

where $e^{r_i t}$ are solutions of the homogenous renewal equation (B.11). The coefficients are the residues of the integrand of equation (B.15) evaluated at the characteristic roots (eigenvalues) of (B.11). The pole with the largest real part determines the asymptotic or long-time behavior of the population age structure and the inverse Laplace transform is taken to the right of this value.

Appendix C

A PDE partnering model

The model presented here is an extension of a model discussed in Hoppensteadt [55, Chp 2]. It is used as a starting point to combine partnering dynamics with demography. The subscript k is used to indicate gender: $k = f$ for women and $k = m$ for men. Let $p_{yk}(t, a)$ be the number of young women and men aged a. Let $p_{ek}(t, a)$ be the number of women and men aged a who are eligible to participate in relationships. Individuals in relationships are denoted by $p_c(t, a, b)$ representing the number of relationships where the woman is aged a and the man is aged b. The mortality rate for individuals aged a is modeled by $\mu_k(a)$. $\eta_k(a)$ is the rate at which young individuals aged a mature into becoming eligible for relationships. The rate at which eligible woman aged a meet eligible men aged b is given by $\kappa(a, b)$. Individuals are born at a rate of B, and an equal fraction of men and women are born.

The model is an SIR-type model structured by age and gender which tracks relationships explicitly. With these notations, the equations are, for young men and women:

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) p_{yk}(t, a) \quad = \quad -\mu_k(a)\, p_{yk}(t, a) - \eta_k(a)\, p_{ky}(t, a)$$

For eligible men and women:

$$\begin{aligned}\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) p_{ek}(t, a) \quad = \quad & -\mu_k(a)\, p_{ek}(t, a) + \eta_k(a)\, p_{yk}(t, a) \\ & + \int_0^{\varpi} p_c(t, a, b)\big[\sigma(a, b) + \mu_{k^*}(b)\big]\, db \\ & - \int_0^{\varpi} \Psi(t, a, b, p_{e1}, p_{e2})\, db\end{aligned}$$

For couples:

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}+\frac{\partial}{\partial b}\right)p_c(t,a,b) = -[\sigma(a,b)+\mu_m(a)+\mu_f(b)]\,p_c(t,a,b) + \Psi(t,a,b) \quad \text{(C.1)}$$

with the following boundary conditions:

$$\begin{aligned} p_{m;y}(t,0) &= p_{f;y}(t,0) = \frac{B}{2} \\ p_c(t,0,b) &= p_c(t,a,0) = 0 \end{aligned}$$

and comments:

- $p_{yk}(t,a)$, $p_{ek}(t,a)$, $p_c(t,a,b)$ are non-negative functions defined for $t \geq 0$, $a \geq 0$, $b \geq 0$.
- All rates are defined as functions of age. Due to the scarcity of data it is often necessary to use constant (age-independent) parameter values.
- All processes (birth, mortality, maturity) enter the system of equation linearly. The essential non-linearity is due to pair formation $\Psi(t,a,b)$.
- Notice that relationship separation (due to death or divorce) adds only linear complexity to the two-sex model.
- Transition rates: μ- mortality, η- maturity, σ- relationship dissolution rate, β- birth rate, ϖ- oldest age we wish to consider.
- Note: $\mathbf{da} = da\,db$ is an age interval, small enough for little variation in conditions of the model, but large enough to contain a large number of individuals. This assumption makes it possible to model the process as continuous [91].
- We use the same rates for mortality, birth, relationship formation and dissociation as in Sect. 6.1.1. We use standard curve from demography to model the process of maturing into eligibility for relationships. The curve depicted in Fig. C.1(b) is $\eta(a) = 0.17e^{-4.411e^{0.309a}}$, where a is the youngest age of being eligible to participate in relationships. The values $a = 13$ and $a = 16$ are used for the first age of eligibility for men and women respectively. We use this function to model the process of maturing into participation in sexual relations, and in South Africa it is known that the age of first participation is generally younger for women than for men.

The two-sex model is a unresolved demographical model, despite its apparent simplicity. The main difficulty is associated with capturing a complicated 'marriage market', an unknown function of age-sex composition [94] in a simple closed form.

A widely used choice, considered to be the least flawed of the usual marriage functions [94], is the harmonic mean of eligible men and women [64],[95]:

$$\Psi(p_{m;e}, p_{f;e})(a,b) = \kappa(a,b)\left[\frac{p_{m;e}(t,a)p_{f;e}(t,b)}{p_{m;e}(t,a)+p_{f;e}(t,b)}\right]\Upsilon\left(\frac{p_{m;e}(t,a)}{p_{f;e}(t,b)}\right) \tag{C.2}$$

where $\kappa(a,b)$ is the rate at which couples form between a man aged a and a woman aged b, also called the coefficient of nuptiality. The function $\Upsilon\left(\frac{p_{m;e}}{p_{f;e}}\right)$ is meant to bias the harmonic mean of eligible men and women in order to make the marriage function proportional to whichever group is more abundant.

The harmonic mean pairing function assumes that pairs are formed within a 'marriage circle', and neglects competition for eligibles from other categories. A marriage function where pairing occurs within the whole eligible pool was proposed by [41]:

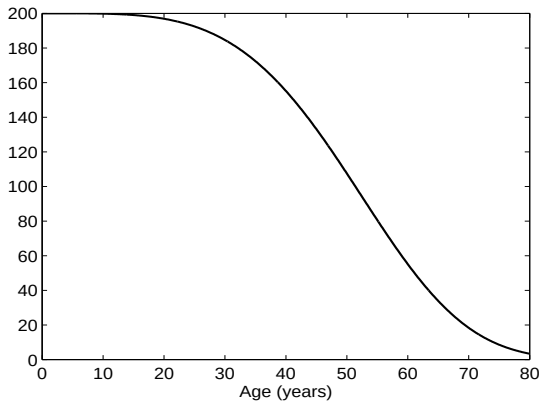
$$\Psi(p_{m;e}, p_{f;e})(a,b,t) = \kappa(a,b)\left[\frac{p_{m;e}(t,a)p_{f;e}(t,b)}{\int_0^\varpi p_{m;e}(t,a')\,da' + \int_0^\varpi p_{f;e}(t,b')\,db'}\right]$$

C.1 Simulation results

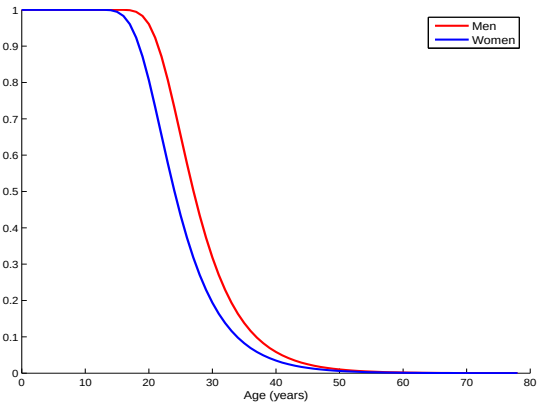
The use of a well-designed finite difference scheme opens up possibilities of studying partnering dynamics. However, each new dynamic is modeled by means of an extra dimension in the cohort matrices, so there is a practical limit to what can be achieved. For example, it is straightforward to track the total time spent in relationships, and hence to verify that the correct (e.g. a constant given) distribution of relationship durations is achieved (Fig. C.1(c)).

Subtle properties of partnering dynamics can be uncovered with a more elaborate finite difference scheme. In Fig. C.1(d) we show the age distributions of men who experienced $1, 2, ...$ relationships. The large age difference between the average age of these successive distributions, especially the gap between the first and second one, demonstrates what seems to be a fundamental limitation of using a generalized harmonic mean function to partner eligible men and woman: the waiting time between relationships becomes very long when all eligibles are included in the denominator of the harmonic mean function.

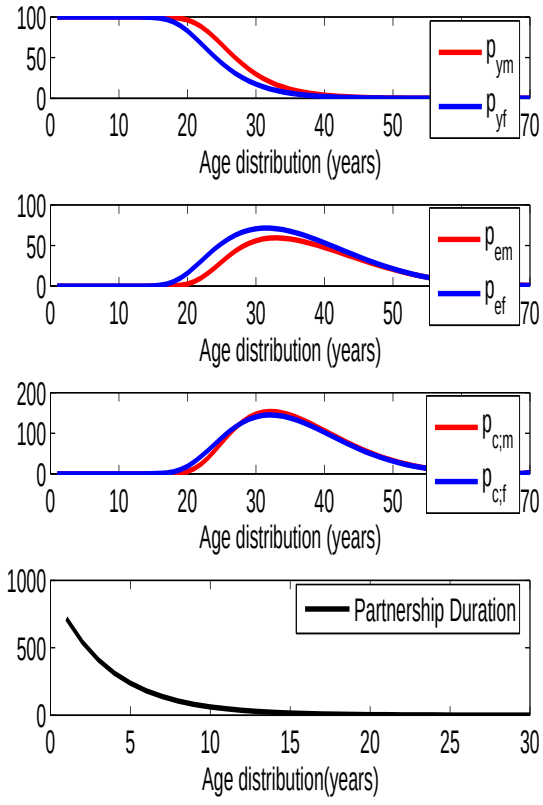
The implication of long waiting times between relationships, is that it is difficult to simulate a persistent epidemic in a two-sex population where relationships form according to a harmonic mean function of the form used in equation (C.2).



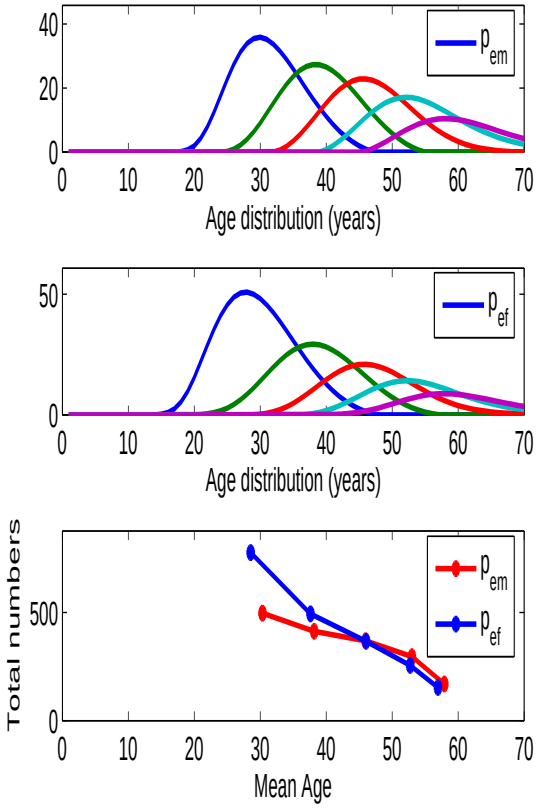
(a) mortality



(b) maturity



(c) age distribution



(d) ordinal relationship number

Figure C.1: (a) Survivor function, hazard function (mortality). (b) Distribution of time spent as young individuals. A standard curve, which is not based on a behavioral rational, is the exponential of an exponential curve, due to Coale: $\eta(a) = 0.17e^{-4.411e^{0.309a}}$ [57, p. 48]. (c) Age distribution of population state. Young women and men (top), eligible women and men (middle) and couples (bottom) (d) The age distribution of eligible men who have been in their first, second, ... relationship.

Bibliography

[1] *Health Systems Trust,* http://www.hst.org.za/healthstats/205/data, Tech. report.

[2] *UNData, Infant mortality rate.* http://data.un.org/Data.aspx?d=WHO&f=inID, Tech. report.

[3] *UNData, Under-5 mortality rate.* http://data.un.org/Data.aspx?d=WHO&f=inID, Tech. report.

[4] *Adult mortality (age 15-64) based on death notification data in South Africa*, Tech. report, Statistics South Africa, 1997-2004.

[5] *National HIV and Syphilis sero-prevalence survey of women attending public antenatal clinics in South Africa 2001*, Tech. report, Department of Health of South Africa, 2002.

[6] *National HIV and Syphilis antenatal sero-prevalence survey in South Africa 2003*, Tech. report, Department of Health of South Africa, 2004.

[7] *South African National HIV prevalence, HIV incidence, behaviour and communication survey, 2005. HSRC Press*, Tech. report, Nelson Mandela Foundation, 2005.

[8] *AIDS Epidemic Update*, Tech. report, UNAIDS, 2006.

[9] *World Health Organization: Global tuberculosis control: surveillance, planning, financing. WHO, Geneva*, (2007).

[10] *The national HIV and Syphilis prevalence survey South Africa 2007*, Tech. report, epartment of Health of South Africa, 2008.

[11] *World Health Organization: Global tuberculosis control – surveillance, planning, financing. WHO, Geneva*, Tech. report, 2009.

[12] L.J. Abu-Raddad, A.S. Magaret, C. Celum, A. Wald, I.M. Longini, S.G. Self, and L. Corey, *Genital herpes has played a more important role than any other sexually transmitted infection in driving HIV prevalence in Africa*, PLoS **3** (2008), no. 5, e2230.

[13] R.M. Anderson and R.M. May, *Infectious diseases of humans; Dynamics and control*, Oxford Science Publications, 1990.

[14] B. Auvert, D. Taljaard, E. Lagarde, J. Sobngwi-Tambekou, R. Sitta, and A. Puren, *Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: The ANRS 1265 trial*, PLoS Medicine **2** (2005), no. 11, e298.

[15] N. Bacaer and X. Abdurahman, *Resonance of the epidemic threshold in a periodic environment*, Journal of Mathematical Biology **57** (2008), no. 5, 649–1673.

[16] N. Bacaer, X. Abdurahman, and J. Ye, *Modeling the HIV/AIDS epidemics among injecting drug users and sex workers in Kunming, China*, Bulletin of Mathematical Biology **68** (2006), no. 3, 525–550.

[17] N. Bacaer, R. Ouifki, C.D. Pretorius, R. Wood, and B. Williams, *Modeling the joint epidemics of TB and HIV in a South African township*, Journal of Mathematical Biology **57** (2008), no. 4, 557–593.

[18] M. Badri, D. Wilson, and R. Wood, *Effect of highly active antiretroviral therapy on incidence of tuberculosis in South Africa: a cohort study*, Lancet **359** (2002), no. 9323, 2059–2064.

[19] F. Ball, D. Mollison, and G. Scalia-Tomba, *Epidemics with two levels of mixing*, The Annals of Applied Probability **7** (1997), no. 1, 46–89.

[20] N.G. Becker, *Analysis of infectious disease data*, 1st ed., Chapman and Hall, 1989.

[21] C.P. Bhunu, W. Garira, and Z. Mukandavire, *Modeling HIV/AIDS and tuberculosis coinfection*, Bulletin of Mathematical Biology **71** (2009), no. 7, 1745–1780.

[22] J. Bongaarts, *Late marriage and the HIV epidemic in sub-Saharan Africa*, Population Studies **61** (2007), no. 1, 73–83.

[23] T. Britton, M. Deijfen, A.N. Lageras, and M. Lindholm, *Epidemics on random graphs with tunable clustering*, Journal of Applied Probability **45** (2008), no. 3, 743–756.

[24] A. Buv, M. Caral, R.J. Hayes, B. Auvert, and et al., *Multicentre study on factors determining differences in rate of spread of HIV in sub-Saharan Africa: methods and prevalence of HIV infection*, AIDS **16** (2001), no. 15, 127–131.

[25] H. Carmichael, *An open systems approach to quantum optics: Lectures presented at the Universite Libre De Bruxelles October 28 to November 4*, Springer, 1991.

[26] J.B. Casterline, L. Williams, and P. McDonald, *The age difference between spouses: variations among developing countries*, Population Studies **40** (1986), no. 3, 353–374.

[27] W.Y. Chen and S. Bokka, *Stochastic modeling of nonlinear epidemiology*, Journal of Theoretical Biology **234** (2005), no. 4, 455–470.

[28] T. Cohen, M. Lipsitch, R.P. Walensky, and M. Murray, *Beneficial and perverse effects of isoniazid preventive therapy for latent tuberculosis infection in HIV-tuberculosis coinfected populations*, Proceedings of the National Academy of Sciences of the United States of America **103** (2006), no. 18, 7042–7047.

[29] E.L. Corbett, S. Charalambous, K. Fielding, T. Clayton, R.J. Hayes, K.M. De Cock, and G.J. Churchyard, *Stable incidence rates of tuberculosis (TB) among human immunodeficiency virus (HIV)-negative South African gold miners during a decade of epidemic HIV-associated TB*, American Journal of Infectious Diseases **188** (2003), no. 8, 1156–1163.

[30] E.L. Corbett, C.J. Watt, N. Walker, D. Maher, B.G. Williams, M.C. Raviglione, and C. Dye, *The growing burden of tuberculosis: global trends and interactions with the HIV epidemic*, Archives of Internal Medicine. **163** (2003), no. 9, 1009–1021.

[31] C.S.M. Currie, K. Floyd, B.G. Williams, and C. Dye, *Cost, affordability and cost-effectiveness of strategies to control tuberculosis in countries with high HIV prevalence*, BMC Public Health **57** (2000), no. 3, 235–247.

[32] C.L. Daley, P.M. Small, G.F. Schecter, G.K. Schoolnik, R.A. McAdam, W.R. Jacobs, and P.C. Hopewell, *An outbreak of tuberculosis with accelerated progression among persons infected with the human immunodeficiency virus: An analysis using restriction-fragment-length polymorphisms*, The New England Journal of Medicine **326** (1992), no. 2, 231–235.

[33] D.J. Daley and D. Vere-Jones, *An introduction to the theory of point processes: General theory and structure*, 1st ed., Springer, 2003.

[34] W. Delva, *Personal communication*, SACEMA.

[35] G. Di Perri, M. Cruciani, M.C. Danzi, R. Luzzati, G. De Checchi, M. Malena, and et al, *Nosocomial epidemic of active tuberculosis among HIV-infected patients*, Lancet **2** (1989), no. 8678–8679, 1502–1504.

[36] O. Diekmann and J.A.P. Heesterbeek, *Mathematical epidemiology of infectious diseases: Model building, analysis and interpretation*, 1st ed., Wiley, 2000.

[37] R. Dorrington, L. Johnson, D. Bradshaw, and T.-J. Danial, *The demographic impact of HIV/AIDS in South Africa; National and provincial indicators for 2006*, The Actuarial Society of South Africa, 2006.

[38] T.D. Eames and M.J. Keeling, *Monagamous networks and the spread of sexually transmitted diseases*, Mathematical Biosciences **189** (2004), no. 2, 115–130.

[39] H. Epstein, *The invisible cure, Africa the West and the fight against AIDS*, Viking, 2007.

[40] Castillo-Chavez C. Feng, Z. and A.F. Capurro, *A model for tuberculosis with exogenous reinfection*, Theoretical Population Biology **57** (2000), no. 3, 235–247.

[41] A.G. Fredericson, *A mathematical theory of age structure in sexual populations: random mating and monogamous marriage models*, Mathematical Biosciences **10** (1971), 117–143.

[42] K. French, S. Riley, and G. Garnett, *Simulations of the HIV epidemic in sub-Saharan Africa: sexual transmission versus transmission through unsafe medical injections*, Sexually Transmitted Diseases **33** (2006), no. 3, 127–134.

[43] C.W Gardiner, *Handbook of stochastic methods: for physics, chemistry and the natural sciences*, vol. 2, Springer, 1997.

[44] G.P. Garnett and R.M Anderson, *Balancing sexual partnerships in an age and activity stratified model of HIV transmission in heterosexual populations*, IMA Journal of Mathematics Applied in Medicine and Biology **11** (1994), no. 3, 161–192.

[45] D. Gillespie, *Exact stochastic simulation of coupled chemical reactions*, The Journal of Physical Chemistry **81** (1977), no. 25, 2340–2361.

[46] M.G.M. Gomes, A.O. Franco, M.C. Gomes, and G.F. Medley, *The reinfection threshold promotes variability in tuberculosis epidemiology and vaccine efficacy*, Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences **1539** (2004), no. 1539, 617–623.

[47] R.M. Granich, C.F. Gilks, C. Dye, K.M. De Cock, and B.G. Williams, *Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model*, Lancet **373** (2009), 48–57.

[48] D. Guwatudde, S.M. Debanne, M. Diaz, C. King, and C.C. Whalen, *A re-examination of the potential impact of preventive therapy on the public health problem of tuberculosis in contemporary sub-Saharan Africa*, Preventive Medicine **39** (2004), no. 4, 1036–1046.

[49] K.P Hadeler, *AIDS epidemiology: methodological issues*, ch. Structured population models for HIV infection: pairwise formation and non-constnt infectivity, pp. 156–173, Birkhauser, 1992.

[50] T.B. Hallet, S. Gregson, J.J.C Lewis, B.A. Lopman, and G.P. Garnett, *Behavioral change in generalized HIV epidemic: impact of reducing cross-generational sex and delaying sexual debut*, Sexually Transmitted Infections **83** (2007), 50–54.

[51] J. Hargrove, *Migration, mines and mores: the HIV epidemic in Southern Africa*, South African Journal of Science. (104), 53–61.

[52] A. Harrison, J. Cleland, and J. Frohlich, *Young people's sexual partnerships in KwaZulu-Natal, South Africa: patterns, contextual influences, and HIV risk*, Studies in Family Planning **39** (2008), no. 4, 295–308.

[53] H.W. Hethcote and P. van den Driessche, *Some epidemiological models with nonlinear incidence*, Journal of Mathematical Biology **29** (1991), no. 3, 271–287.

[54] T.D. Hollingsworth, R.M Anderson, and C. Fraser, *HIV-1 Transmission, by stage of infection*, The Journal of Infectious Diseases **198** (2008), no. 5, 687–693.

[55] F. Hoppensteadt, *Mathematical methods of population biology*, SIAM, 1976.

[56] J.H. Hyman and J. Li, *An intuitive formulation for the reproductive number for the spread of diseases in heterogeneous populations*, Mathematical Biosciences **167** (2000), no. 1, 65–86.

[57] M. Ianelli, M. Martcheva, and F.A. Milner, *Gender-structured population modeling: mathematical methods, numerics, and simulations*, Society for Industrial and Applied Mathematics, 2005.

[58] H. Inaba, *A mathematical model for human population reproduction by iterative marriage*, Tokyo : Institute for Population Problems, Ministry of Health and Welfare (1993).

[59] ———, *Calculating R0 for HIV infection via pair formation: Working paper series*, Tokyo : Institute for Population Problems, Ministry of Health and Welfare (1994).

[60] ———, *Endemic threshold results in an age-duration-structured population model for HIV infection*, Mathematical Biosciences **201** (2006), no. 1-2, 15–47.

[61] K. Jochelson, M. Mothibeli, and Leger J.P., *Human immunodeficiency virus and migrant labor in South Africa*, International Journal of Health servises **21** (1991), no. 1, 157–73.

[62] M. Kamupira, *A behavioral study in Guguletu.*, UCT, School of Public Health, unpublished work. (2009).

[63] I. Katz and D. Low-Beer, *Why has HIV stabilized in South Africa, yet not declined further? Age and sexual behavior patterns among youth*, Sexually Transmitted Diseases **35** (2008), no. 10, 837–842.

[64] N. Keyfitz, *The mathematics of sex and marriage*, Procedings of the sixth Berkely symposium on mathematical statistics and probability, vol. 4, 1971.

[65] N. Keyfitz and H. Caswell, *Applied mathematical demography*, 3rd ed., Springer, 2005.

[66] J.F.C. Kingman, *Poisson processes*, Oxford University Press, 1993.

[67] S.E. Kline, L.L. Hedemark, and S.F. Davies, *Outbreak of tuberculosis among regular patrons of a neighborhood bar*, The New England Journal of Medicine **333** (1995), no. 4, 222–227.

[68] S.D. Lawn, M. Badri, and R. Wood, *Tuberculosis among HIV-infected patients receiving HAART: long term incidence and risk factors in a South African cohort*, AIDS **19** (2005), no. 18, 2109–2116.

[69] S.D. Lawn, L.-G. Bekker, K. Middelkoop, L. Myer, and R. Wood, *Impact of HIV infection on the epidemiology of tuberculosis in a peri-urban community in South Africa: the need for age-specific interventions*, Clinical Infectious Diseases **42** (2006), no. 7, 1040–1047.

[70] S.D. Lawn, L.-G. Bekker, and R. Wood, *How effectively does HAART restore immune responses to Mycobacterium tuberculosis? Implications for tuberculosis control*, AIDS **19** (2005), no. 11, 1113–1124.

[71] S.D. Lawn, L. Myer, L.-G. Bekker, and R. Wood, *Burden of tuberculosis in an antiretroviral treatment programme in sub-Saharan Africa: impact on treatment outcomes and implications for tuberculosis control*, AIDS **20** (2006), no. 12, 1605–1612.

[72] S.D. Lawn, R.J. Wilkinson, M.C.I. Lipman, and R. Wood, *Immune reconstitution and "unmasking" of tuberculosis during antiretroviral therapy*, Journal of Respiratory and Critical Care Medicine **177** (2008), 680–685.

[73] S.D. Lawn and R. Wood, *The epidemic of HIV-associated tuberculosis in sub-Saharan Africa: does this also impact non-HIV-infected individuals?*, AIDS **20** (2006), no. 13, 1787–1788.

[74] _____, *Tuberculosis control in South Africa - will HAART help?*, South African Medical Journal **96** (2006), no. 6.

[75] _____, *When should antiretroviral treatment be started in patients with HIV-associated tuberculosis in South Africa?*, South African Medical Journal **97** (2007), 412–414.

[76] M. Lipsitch and M.B. Murray, *Multiple equilibria: tuberculosis transmission require unrealistic assumptions*, Theoretical Population Biology **63** (2003), no. 2, 169–170.

[77] T. Mah, *Concurrent sexual partnerships and HIV transmission in Khayelitsha, South Africa*, AIDS Prevention Research Project, Harvard School of Public Health (2008).

[78] R.M May, R.M. Anderson, and M.E. Irwin, *The transmission dynamics of human immunodeficiency virus (HIV)*, Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences **321** (1988), no. 1207, 565–607.

[79] J. McKane and T.J. Newman, *Stochastic models in population biology and their deterministic analogs*, Physical Review W **70** (2004), no. 4, 041902.

[80] K. Middelkoop, L.-G. Bekker, L. Myer, R. Dawson, and R. Wood, *Rates of tuberculosis transmission to children and adolescents in a community with a high prevalence of HIV infection among adults*, Clinical Infectious Diseases **47** (2008), 349–355.

[81] A. Miranda, M. Morgan, L. Jamal, and et al., *Impact of antiretroviral therapy on the incidence of tuberculosis: the Brazilian experience*, PLoS ONE **2** (2007), e826.

[82] V. Mishra and S. Bignami-Van Assche, *Concurrent sexual partnerships and HIV Infection: evidence from national population-based surveys*, Tech. Report 62, USAID, DHS Working Papers, 2009.

[83] S.M. Moghadas and M.E. Alexander, *Exogenous reinfection and resurgence of tuberculosis: A theoretical framework*, Journal of Biological Systems **12** (2004), no. 2, 231–247.

[84] S.M. Moghadas and A.B. Gumel, *Analysis of a model for transmission dynamics of TB*, Canadian Applied Math Quarterly **10** (2002), 411–428.

[85] L. Morison, H. A. Weiss, A. Buv, M. Caral, S.-C. Abega, F. Kaona, L. Kanhonou, J. Chege, R. J. Hayes, and et al., *Commercial sex and the spread of HIV in four cities in sub-Saharan Africa*, AIDS **15** (1001), 62–69.

[86] B.M. Murphy, B.H. Singer, and D. Kirschner, *On treatment of tuberculosis in heterogeneous populations*, Journal of Theoretical Biology **223** (2003), no. 4, 391–404.

[87] C.J.L. Murray, K. Styblo, and A. Rouillon, *Tuberculosis in developing countries: burden, intervention, and cost*, Bulletin of the International Union againts tuberculosis and Lung Disease **65** (1990), no. 1, 6–24.

[88] R.K. Pathria, *Statistical mechanics*, 2nd ed., Butterworth-Heinemann, 2006.

[89] M. Plischke and B. Bergersen, *Equilibrium statistical physics*, 3nd ed., World Scientific, 2006.

[90] L.E. Reichl, *A modern course in statistical physics*, 2nd ed., Wiley-Interscience, 1998.

[91] R.D. Richtmyer, *Priciples of advanced mathematical physics*, vol. I, Springer, 1978.

[92] H. Rieder, *On the risk of being and becoming infected with Mycobacterium tuberculosis*, Clinical Infectious Diseases **47** (2008), no. 3, 356–357.

[93] H. Risken and T. Frank, *The Fokker-Planck equation: Methods of solutions and applications*, 2nd ed., Springer, 1996.

[94] R. Schoen, *The harmonic mean as the basis of a realistic two-sex marriage model*, Demography **18** (1981), no. 1, 201–216.

[95] ———, *Modeling multigroup populations*, Plenum Press, 1988.

[96] P.A. Selwyn, D. Hartel, V.A. Lewis, E.E. Schoenbaum, S.H. Vermund, R.S. Klein, A.T. Walker, and G.H. Friedland, *A prospective study of the risk of tuberculosis among intravenous drug users with human immunodeficiency virus infection*, The New England Journal of Medicine **321** (1989), no. 18, 1268.

[97] P.A. Selwyn, B.M. Sckell, P. Alcabes, G.H. Friedland, R.S. Klein, and E.E. Schoenbaum, *High risk of ctive tuberculosis in HIV-infected drug users with cutaneous anergy*, JAMA, the Journal of the American Medical Association **298** (1992), no. 4, 504–509.

[98] O. Sharomi, C.N. Podder, and A.B. Gumel, *Mathematical analysis of the transmission dynamics of HIV/TB coinfection in the presence of treatment*, Mathematical Biosciences and Engineering **5** (2008), no. 1, 145–174.

[99] O. Shisana, T. Rehle, L. Simbayi, W. Parker, K. Zuma, A. Bhana, C. Connolly, and S. Jooste, *South African national HIV prevalence, HIV incidence, behaviour and communication survey*, HSRC, 2005.

[100] B.H. Singer and D.E. Kirschner, *Influence of backward bifurcation on interpretation of R_0 in a model of epidemic tuberculosis with reinfection*, Mathematical Biosciences and Engineering **1** (2004), no. 1, 81–93.

[101] P. Sonnenberg, J.R. Glynn, K. Fielding, J. Murray, P. Godfrey-Faussett, and S. Shearer, *How soon after infection with HIV does the risk of tuberculosis start to increase? A retrospective cohort study in South African gold miners*, Journal of Infectiuous Diseases **191** (2005), no. 2, 150–158.

[102] UNData, http://data.un.org/Data.aspx?d=GenderStat, Tech. report.

[103] N.G. van Kampen, *Advances in chemical physics*, vol. 34, ch. The expansion of the master equation, pp. 245–309.

[104] ———, *Topics in statistical mechanics and biophysics: A memorial to Julius L. Jackson.*, ch. Fluctuations in Continuous Systems, pp. 153–186, American Institute of Physics, 1976.

[105] ———, *Master equation treatment of aging populations*, Reports of mathematical physics **11** (1977), no. 1, 111–122.

[106] ———, *Composite stochastic processes*, Physica A **96** (1979), no. 3, 435–453.

[107] N.G. Van Kampen, *Stochastic processes in physics and chemistry*, 2nd ed., North Holland, 1992.

[108] S. Verver, R.M. Warren, Z. Munch, E. Vynnycky, PD. van Helden, M. Richardson, G.D. van der Spuy, D.A. Enarson, M.W. Borgdorff, M.A. Behr, and N. Beyers, *Transmission of tuberculosis in a high incidence urban community in South Africa*, International journal of epidemiology **33** (2004), no. 2, 351–357.

[109] E. Vynnycky and P. Fine, *The long-term dynamics of tuberculosis and other diseases with long serial intervals: implications of and for changing reproduction numbers*, Epidemiology and Infection **121** (1998), no. 2, 309–24.

[110] E. Vynnycky and P.E.M. Fine, *The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection*, Epidemiology and Infection **119** (1997), no. 2, 183–201.

[111] D.F Walls and G.J. Milburn, *Quantum optics*, Springer, 1995.

[112] W. Wang, *Epidemic models with nonlinear infection forces*, Mathematical Biosciences and Engeneering **3** (2006), 267–279.

[113] C.H. Watts and R. May, *The influence of concurrent partnerships on the dynamics of HIV/AIDS*, Mathematical Biosciences **108** (1992), no. 1, 89–104.

[114] G. Webb, *Theory of nonlinear age-dependent population dynamics*, CRC, 1985.

[115] C. Whalen, *Failure of directly observed treatment for tuberculosis in Africa: A call for new approaches*, Clinical Infectious Diseases **42** (2006), no. 7, 1048–1050.

[116] WHO, *World Health Organization: Global tuberculosis control – surveillance, planning, financing. WHO, Geneva*, Tech. report, 2007.

[117] B. Williams, *Personal communication.*, SACEMA.

[118] _____, *HIV/AIDS in South Africa*, 1st ed., ch. Models and trends, Cambridge University Press, 2006.

[119] B. Williams and D. Maher, *Tuberculosis fueled by HIV: putting out the flames*, American Journal of Respiratory and Critical Care Medicine **175** (2007), 6–7.

[120] B.G. Williams and C. Dye, *Antiretroviral drugs for tuberculosis control in the era of HIV/AIDS*, Science **301** (2003), no. 5639, 1535 – 1537.

[121] B.G. Williams and E. Gouws, *The epidemiology of human immunodeficiency virus in South Africa*, Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences **365** (2001), no. 1411, 1077–1086.

[122] B.G. Williams, J.O. Lloyd-Smith, E. Gouws, C. Hankins, W.M. Getz, J. Hargrove, I. de Zoysa, C. Dye, and B. Auvert, *The potential impact of male circumcision on HIV in sub-Saharan Africa*, PLoS Medicine **3** (2006), no. 7, e262.

[123] R. Wood, K. Middelkoop, L. Myer, A.D. Grant, A. Whitelaw, S.D. Lawn, G. Kaplan, R. Huebner, J. McIntyre, and L.-G. Bekker, *Undiagnosed tuberculosis in a community with high HIV-prevalence: implications for TB control*, American Journal of Respiratory and Critical Care Medicine. **175** (2007), 87–93.