Mathematical and Statistical Insights in Evaluating State Dependent Effectiveness of HIV Prevention Interventions

by

Yuqin Zhao

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Approved November 2014 by the
Graduate Supervisory Committee:

Yang Kuang, Chair
Jesse Taylor
Dieter Armbruster
Wenbo Tang
Yun Kang

ARIZONA STATE UNIVERSITY

December 2014

ABSTRACT

Pre-Exposure Prophylaxis (PrEP) is any medical or public health procedure used before exposure to the disease causing agent, its purpose is to prevent, rather than treat or cure a disease. Most commonly, PrEP refers to an experimental HIV-prevention strategy that would use antiretrovirals to protect HIV-negative people from HIV infection. A deterministic mathematical model of HIV transmission is developed to evaluate the public-health impact of oral PrEP interventions, and to compare PrEP effectiveness with respect to different evaluation methods. The effects of demographic, behavioral, and epidemic parameters on the PrEP impact are studied in a multivariate sensitivity analysis. Most of the published models on HIV intervention impact assume that the number of individuals joining the sexually active population per year is constant or proportional to the total population. In the second part of this study, three models are presented and analyzed to study the PrEP intervention, with constant, linear, and logistic recruitment rates. How different demographic assumptions can affect the evaluation of PrEP is studied. When provided with data, often least square fitting or similar approaches can be used to determine a single set of approximated parameter values that make the model fit the data best. However, least square fitting only provides point estimates and does not provide information on how strongly the data supports these particular estimates. Therefore, in the third part of this study, Bayesian parameter estimation is applied on fitting ODE model to the related HIV data. Starting with a set of prior distributions for the parameters as initial guess, Bayes' formula can be applied to obtain a set of posterior distributions for the parameters which makes the model fit the observed data best. Evaluating the posterior distribution often requires the integration of high-dimensional functions, which is usually difficult to calculate numerically. Therefore, the Markov chain Monte Carlo (MCMC) method is used to approximate the posterior distribution.

ACKNOWLEDGEMENTS

I would like to thank my advisor Professor Yang Kuang for his encouragement, insightful guidance, and career suggestions. I also want to thank Professor Jay Taylor for helping me with the Bayesian parameter estimation and his patience and kindness. I really appreciate all the advises from my collaborator Dr. Dobromir Dimitrov.

Thanks to my family for their emotional support, which makes my life and study more enjoyable.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Chapter 1

INTRODUCTION

Mathematical models have been used extensively to simulate HIV transmission and to study the use of male circumcision, antiretroviral therapy (ART), microbicides, and pre-exposure prophylaxis (PrEP) for HIV prevention. My study will focus on PrEP for HIV prevention.

PrEP is any medical or public health procedure used before exposure to the disease causing agent, its purpose is to prevent, rather than treat or cure a disease (such as malaria, and HIV). PrEP can also refer to the aggressive use of vaccination (such as for rabies). Most commonly, PrEP refers to an experimental HIV-prevention strategy that would use antiretrovirals to protect HIV-negative people from HIV infection. With an estimated 39.5 million people living with HIV worldwide and 4.3 million new infections per year (UNAIDS (2013)), many people are at risk for HIV infection through sexual transmission. Therefore preemptive measures must be taken to prevent further dissemination. PrEP provides a promising prevention strategy for further HIV transmission.

In 2010, evidence from two different randomized clinical trials (Grant *et al.* (2010); Karim *et al.* (2010)) suggested that PrEP products based on antiretroviral drug Tenofovir taken orally as a pill (oral PrEP) or applied topically in the form of gel (vaginal microbicides (VMB)) can help prevent HIV. First in South Africa, the CAPRISA 004 trial demonstrated a 39% (95%CI, 6% to 60%) overall decrease in HIV incidence among women in the VMB (gel) arm of the trial who were advised to use the product before and after each sex act (Karim *et al.* (2010)). Later, the Global iPrEx trial of a daily use of a combination of two oral antiretroviral drugs, emtricitabine and

tenofovir disoproxil fumarate, demonstrated a 44% efficacy (95%CI, 15% to 63%) reduction in the incidence of HIV, among men-who-have sex with men (MSM) (Grant *et al.* (2010)).

The search for a safe and effective HIV vaccine is still ongoing. There is no HIV vaccine and it is not expected in a near future, which makes the study of PrEP important and practical. In 2012, one product (oral Truvada) was approved for PrEP use by FDA in United States, and recommended for use in South Africa. And there has been a broad discussion on what will be population-level benefits from wide-scale PrEP use in high prevalence settings.

Mathematical models have been used to simulate HIV transmission and to study the use of chemoprophylaxis among MSM (Desai *et al.* (2008); Supervie *et al.* (2010)). Deterministic mathematical models of HIV heterosexual transmission stratified by gender have been analyzed in Abbas *et al.* (2007); Dimitrov *et al.* (2011, 2010); Pretorius *et al.* (2010); Vissers *et al.* (2008); Wilson *et al.* (2008). Often a single intervention outcome based on cumulative number or fractions of infections prevented, on reduction in HIV prevalence or incidence has been used to evaluate the effectiveness of PrEP interventions. These indicators express a wide variation over time and often disagree in their forecast on the success of the intervention (Dimitrov *et al.* (2010); Pretorius *et al.* (2010)). Therefore, the conclusions of many modeling studies are significantly influenced by the choice of the evaluation method and the period of evaluation. In particular, it has been pointed out that indicators based on prevented infections tend to show mixed results over time due to their sensitivity to changes in population dynamics (Dimitrov *et al.* (2011)).

In the first part of my study (Zhao *et al.* (2013)), I develop a deterministic mathematical model of HIV transmission to evaluate the public-health impact of oral PrEP interventions, compare PrEP effectiveness with respect to different evaluation meth-

ods, and analyze its dynamics over time. Four traditional evaluation methods are compared, including relative reduction in HIV prevalence and incidence which avoid the ambiguity associated with commonly used indicators based on the absolute number of prevented infections. Two additional methods are considered which estimate the burden of the epidemic to the public-health system. I then investigate the short term and long term behavior of these indicators and the effects of key parameters on the expected benefits from PrEP use. The effects of demographic, behavioral, and epidemic parameters on the PrEP impact are studied in a multivariate sensitivity analysis.

Since HIV is acquired predominately through sexual contacts, it is usually modeled as a sexually transmitted disease, ignoring the other routes of transmission. In this study, sexually active heterosexual population aged from 15 to 49 year are considered, as assumed in many published papers. Most of the published models on HIV intervention impact I found assume that the number of individuals joining the sexually active population per year is constant (constant entrance) or proportional to the total population (linear entrance). In sexually active population with no migration, constant entrance implies that the same number of people reaches sexual maturity annually. This may be an acceptable approximation over some short time period, but not reasonable as the simulation period increases. The progression of HIV infection from acquisition to symptoms of AIDS and then death is incredibly slow compared with many other fatal diseases, and it can be further delayed by ART. Therefore, a meaningful impact of prevention intervention should be expected over several decades, which explains simulation periods of 20-50 years used in mathematical models. In the second part of this study, I will present and analyze three models to study the PrEP intervention, with constant, linear, and logistic recruitment rates. Except the recruitment rate, all the assumptions for the three models stay the same.

How different demographic assumptions can affect the evaluation of PrEP will be studied.

In the third part of this study, I will apply Bayesian inference on parameter estimation on the basic HIV model without PrEP intervention. The least square fitting (or similar approaches) only gives a single set of approximated parameter values that make the model fit the data best. It only provides point estimates and does not provide information on how strongly the data supports these particular estimates. Therefore, a second approach I will try is the Bayesian parameter estimation on fitting ODE model to the related HIV data. Starting with a set of prior distributions for the parameters as initial guess, we can apply Bayes' formula to obtain a set of posterior distributions for the parameters which make the model fit the observed data best. The posterior distributions provide the probability for each parameter to equal any possible value.

Chapter 2

EFFECTIVENESS INDICATORS FOR HIV INTERVENTIONS

Details of the work in this section can be found in Zhao *et al.* (2013).

2.1 Model Description

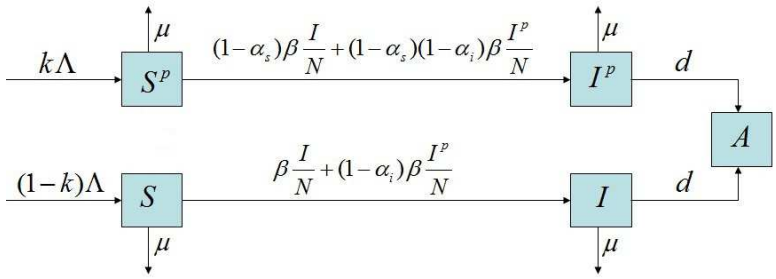


Figure 2.1: Flow diagram of a PrEP intervention for the model (1) formulation.

In my models (see Figure 2.1) the population is divided into two major classes, PrEP users (superscript p, $S^p + I^p$) and those that do not use PrEP ($S + I$), and further stratified according to their HIV status into susceptible (S, S^p) and infected (I, I^p). Individuals who develop AIDS are accumulated in non-sexually active class A. Individuals join the community (reaching sexual maturity) and departure from the sexually active population at constant rates (Λ, μ). A proportion k of the new recruits start using PrEP. PrEP users are assumed to strictly follow the prescribed regimens. The model which assumes that PrEP reduces both susceptibility and infectiousness of the users ("dual-protection" model) is formulated by the following system of differential equations:

$$
\begin{aligned}
\frac{dS^p}{dt} &= k\Lambda - (1-\alpha_s)\beta\frac{S^pI}{N} - (1-\alpha_s)(1-\alpha_i)\beta\frac{I^pS^p}{N} - \mu S^p \\
\frac{dS}{dt} &= (1-k)\Lambda - \beta\frac{IS}{N} - (1-\alpha_i)\beta\frac{I^pS}{N} - \mu S \\
\frac{dI^p}{dt} &= (1-\alpha_s)\beta\frac{IS^p}{N} + (1-\alpha_s)(1-\alpha_i)\beta\frac{I^pS^p}{N} - (\mu+d)I^p \\
\frac{dI}{dt} &= \beta\frac{IS}{N} + (1-\alpha_i)\beta\frac{I^pS}{N} - (\mu+d)I.
\end{aligned}
\tag{2.1}
$$

Table 2.1: Parameter description and baseline values

Par.	Description	Value	Ref.
d	HIV carrier's annual rate of progression to AIDS	0.1302	fitted sta (2012)
Λ	Annual rate at which individuals become sexually active	38094	calc. sta (2012)
$\frac{1}{\mu}$	Time (in years) to remain sexually active	$\frac{1}{0.0250}$	fitted sta (2012)
b_a	HIV acquisition risk per act	0.0030	fitted sta (2012)
n	Number of sexual acts per year per individual	65.8494	fitted sta (2012)
β	Cumulative HIV-acquisition risk	$\beta(n, b_a)$	calc.
$N(0)$	Initial sexually active population	10^6	assumed
P	Initial HIV prevalence	0.166	sta (2012)
k_1	Initial PrEP coverage	0.2	assumed
θ	Reduction in the initial fraction of HIV positive individuals as a result of pre-enrollment screening	0.5	assumed
k	Proportion of the new recruits that start using PrEP	$k = k_1$	assumed
α_s	Efficacy of PrEP in reducing susceptibility of PrEP users	0.5	assumed
α_i	Efficacy of PrEP in reducing infectiousness of PrEP users	0.5	assumed

Since the differential equations for these four compartments are independent from the AIDS class (A), it is not included in the ODE system. Here $N = S^p + S + I^p + I$ represents the sexually active population and α_s (α_i) measures the efficacy of PrEP in reducing susceptibility (infectiousness) of PrEP users. The cumulative HIV acquisition risk per year β is calculated based on the HIV risk per act (b_a) with a HIV-positive partner and the average number of sex acts per year (n):

$$\beta = 1 - (1 - b_a)^n.$$

Cumulative acquisition risk (β) is an increasing function with respect to HIV-acquisition risk per act b_a and average number of sexual acts per year n.

PrEP is introduced at time ($t = 0$) in a population with $N(0) = 1,000,000$ and HIV-prevalence (P). It is assumed that PrEP is initially adopted by a fraction k_1 of the individuals and that the initial fraction of HIV-positive individuals is reduced by θ as a result of pre-enrollment HIV screening:

$$\begin{aligned} S^p(0) &= k_1(1-P)N(0) \\ S(0) &= (1-k_1)(1-P)N(0) \\ I^p(0) &= (1-\theta)k_1PN(0) \\ I(0) &= (1-(1-\theta)k_1)PN(0), \end{aligned}$$

The initial distribution of the epidemic classes may not be critical for the asymptotic behavior of the system but it is essential for the impact indicators calculated over fixed periods of time after the start of the intervention. To isolate the impact of the choice of the evaluation method I simplify the intervention schedule and assume instantaneous uptake of PrEP at predetermined level ($k = k_1$).

Clinical trials in HIV prevention are usually designed to evaluate the efficacy of the tested products in reducing susceptibility (α_s) only. Although a reduction in

infectiousness (α_i) is plausible it is not verifiable since all HIV positive participants are immediately withdrawn from the product. Moreover, the biomedical products currently in testing, are based on antiretroviral drugs and are not recommended for use by infected individuals due to the risk of drug resistance development. Therefore, the majority of the modelers assume uni-directional PrEP protection ($\alpha_i = 0$) which means that using PrEP has no effect on the infectiousness or that infected individuals do not take PrEP anymore. This scenario may also represent the idea of control of the PrEP usage by the HIV-positive individuals since fast removal of the infected users from PrEP is the equivalent of setting $\alpha_i = 0$. To address that possibility I consider a "single-protection" model in which the variable I^p is removed from the baseline model as follows:

$$\begin{aligned} \frac{dS^p}{dt} &= k\Lambda - (1-\alpha_s)\beta\frac{S^pI}{N} - \mu S^p \\ \frac{dS}{dt} &= (1-k)\Lambda - \beta\frac{SI}{N} - \mu S \\ \frac{dI}{dt} &= \beta\frac{SI}{N} + (1-\alpha_s)\beta\frac{S^pI}{N} - (\mu+d)I \end{aligned} \tag{2.2}$$

with initial conditions:

$$\begin{aligned} S^p(0) &= k_1(1-P)N(0) \\ S(0) &= (1-k_1)(1-P)N(0) \\ I(0) &= PN(0). \end{aligned}$$

In my analysis HIV epidemics are simulated in presence and in absence of PrEP. If PrEP is not available the "no intervention" model is reduced to the following system:

$$\begin{aligned} \frac{dS}{dt} &= \Lambda - \beta\frac{SI}{N} - \mu S \\ \frac{dI}{dt} &= \beta\frac{SI}{N} - (\mu+d)I \end{aligned} \tag{2.3}$$

with $S(0) = (1-P)N(0)$ and $I(0) = PN(0)$.

2.1.1 Modeling Assumptions

Several important assumptions are incorporated into the model:

- The HIV prevalence in the whole population is representative for the HIV prevalence among each gender, i.e., the chance to have a HIV-positive partner is proportional to the total HIV prevalence. Transmission and susceptibility do not differ between the sexes, either with or without PrEP.
- Individuals are assumed to have a fixed number of sex acts per year.
- Sexual behavior of an individual does not change if he/she starts using PrEP but sexual activity stops once AIDS is developed.
- The use of PrEP reduces both HIV susceptibility and infectiousness (independently) and by this reduces the HIV acquisition risk per sex act.
- Perfect adherence to PrEP is assumed: individuals who start using PrEP continue to follow the prescribed regimen indefinitely. However, the scenario with no reduction of infectiousness due to PrEP ($\alpha_i = 0$) is equivalent to immediate withdrawal from PrEP after HIV acquisition.
- The use of other HIV prevention measures including condom use, male circumcision, and ARV treatments are not considered separately in my model. Their effects on HIV transmission are aggregated in the HIV acquisition risk per act.

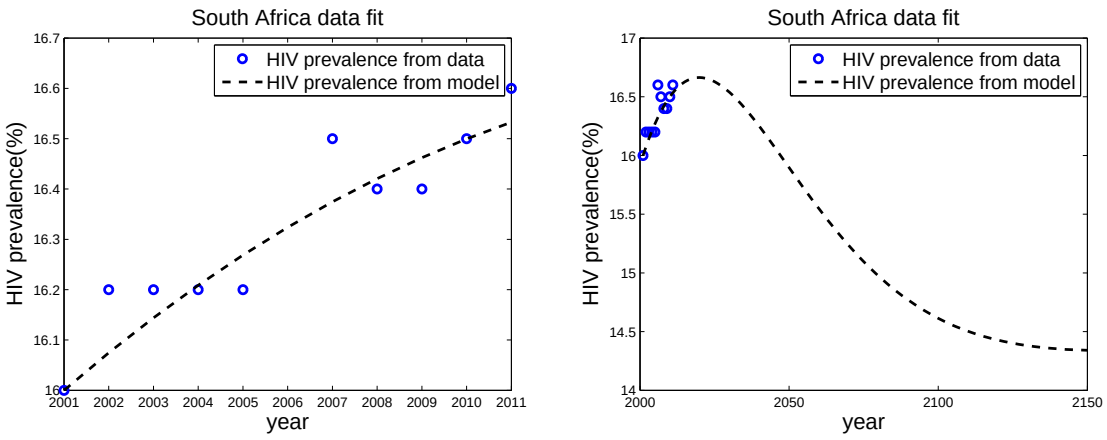
2.1.2 Model Parametrization

To parameterize the model in the scenario without PrEP, I used demographic and HIV prevalence data representative for the sexually active population (15-49 years

old) in South Africa for the period between 2001 and 2011 provided by the Statistical Institute of South Africa (sta (2012)) .

First, I estimate the recruitment rate in the sexually active population (Λ). Calculations are based on the approximated number of 15-year olds ($\frac{\text{population aged 15 to 19}}{5} = \frac{5175400}{5} = 1035100$) and the total population size (27,172,400) aged 15 to 49 , in year 2011. In the model without PrEP, initial total sexually active population is assumed to be $N = 10^6$. Therefore, I scale the estimated entrance rate to obtain the recruitment of the sexually active population (Λ) in the model: $\Lambda = 10^6 \cdot \frac{1035100}{27172400} \approx 38094$ which willed be used in the epidemic simulations.

Next, I fit the projected HIV prevalence $\frac{I}{S+I}$ by the model without PrEP to the 2001-2011 prevalence data from South Africa (sta (2012)). I use the Matlab built-in function 'fminsearch' to do the data fitting, with error measurement $\frac{\sum_{i=1}^{n} |ps_i - p_i|}{n}$, where ps_i represents the HIV prevalence from model simulation, p_i represents the HIV prevalence from data, and n represents the number of data points. Starting with initial parameter values borrowed from published studies: $b_a = 0.0038$(Boily *et al.* (2009)), $n = 80$(Kalichman *et al.* (2009); Wawer *et al.* (2005)), $\mu = 1/35$(UNAIDS (2009)), and $d = 1/10$(Morgan *et al.* (2002); Porter and Zaba (2004)), I obtain the following parameter set which fits best the prevalence data from year 2001 to year 2011: $b_a = 0.0030$, $n = 65.8494$, $\mu = 0.0250$, and $d = 0.1302$ (with error of data fitting=0.0737). Figure 2.2 shows the HIV prevalence data and the best-fitting estimates obtained by the "no intervention" model for the period 2001-2011 (Figure 2.2(a)) as well as its long-term projections (Figure 2.2(b)).



(a) (b)

Figure 2.2: (a) HIV prevalence among sexually active population in South Africa for the period 2001- 2011 from data and fitted with the "no intervention" model; (b) Long-term projections of the HIV prevalence based on fitted "no intervention" model.

2.2 Effectiveness Indicators

The impact of PrEP in my analysis is evaluated by the quantitative indicators described in Table 2.2. The first four indicators are widely used in modeling studies to evaluate the impact of interventions over fixed periods $[0, T]$. The cumulative and the fractional indicators measure the intervention effectiveness based on the infections prevented in scenarios with PrEP compared to scenarios without PrEP. The prevalence and incidence indicators measure the reduction of the projected HIV prevalence and incidence due to PrEP. I propose the last two evaluation methods based on the reduction of the number of infected individuals as they are closely related to the economic burden of the HIV epidemic on the public health system at community and state level since the money allocated for HIV treatment is proportional to the size of the infected population.

Predictions of mathematical models based on quantitative indicators are often

Table 2.2: Indicator description

Indicator	Name	Description
$C_I(T)$	Cumulative indicator	Cumulative number of infections prevented over the period $[0, T]$ due to the usage of PrEP
$F_I(T)$	Fractional indicator	Fraction of infections prevented over the period $[0, T]$ due to the usage of PrEP
$P_I(T)$	Prevalence indicator	Reduction in HIV-prevalence at time $t = T$ due to the usage of PrEP
$aI_I(T)$	Incidence indicator	Reduction in the annual HIV incidence at time $t = T$ due to the usage of PrEP
$\hat{C}_I(T)$	Reduction indicator	Reduction in the projected number of infections at time $t = T$ due to the usage of PrEP
$\hat{F}_I(T)$	Fractional reduction indicator	Fraction of the number of infections reduced at time $t = T$ reduced due to the usage of PrEP

used to estimate the effectiveness of novel interventions and to compare the expected benefits from different prevention options. The analytical conclusions in favor of specific option are usually based on evaluations of the indicators over a few fixed periods of intervention time, most likely 10 years but almost certainly between 5 and 30 years. However, all indicators vary over time and may express different preferences when used to decide between comparable prevention programs. The idea is illustrated with a comparison of the indicator dynamics for two hypothetical PrEP interventions. Intervention 1 assumes no control of the PrEP use by HIV-positive individuals ($\theta = 0$) and 50% PrEP efficacy in reducing both susceptibility and infectiousness ($\alpha_s = \alpha_i = 0.5$) while Intervention 2 requires a negative HIV test as a condition for prescribing PrEP

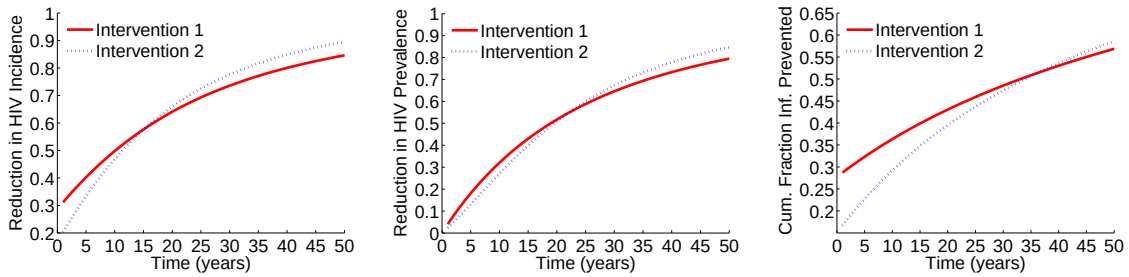


Figure 2.3: Comparison of the indicators projections for two PrEP interventions over 50-year period. Intervention 1 assumes that $\theta = 0$ and $\alpha_s = \alpha_i = 0.5$. Intervention 2 assumes that $\theta = 1$, $\alpha_s = 0.5$ and $\alpha_i = 0.9$. All other parameters are fixed on their baseline parameter values from Table 2.1.

($\theta = 1$) and better PrEP efficacy in reducing infectiousness ($\alpha_s = 0.5$, $\alpha_i = 0.9$). Each of the incidence, prevalence and fractional indicators shows increasing effectiveness of both interventions over 50 years after initiation of PrEP (Figure 2.3) with more benefits attributed to Intervention 1 initially but higher impact of Intervention 2 in a long-term. However, they disagree on the timing when the advantage of Intervention 1 ends. For instance, a preference to Intervention 2 is given after 17 years of PrEP use if based on reduction in HIV incidence and after 22 years if based on reduction in HIV prevalence. The public-health impact of Intervention 1 measured in terms of cumulative fraction of prevented infections remains higher compared to Intervention 2 for up to 32 years which is substantially longer than the evaluation periods used in the majority of the quantitative analyses. Therefore, if PrEP is evaluated over periods between 17 and 32 years the choice of quantitative indicator is critical. Now take a closer look at the key drivers of those discrepancies in the indicators' dynamics.

2.2.1 Indicator Expressions

To utilize the calculation of the cumulative indicators it is necessary to keep track of the cumulative number of new infections. For this reason I add two equations to the "dual-protection" model (2.1):

$$\begin{array}{rcl} \frac{d(I^p_{New})}{dt} & = & (1-\alpha_s)\beta\frac{S^pI}{N} + (1-\alpha_s)(1-\alpha_i)\beta\frac{S^pI^p}{N} \\ \frac{d(I_{New})}{dt} & = & \beta\frac{SI}{N} + (1-\alpha_i)\beta\frac{SI^p}{N}, \end{array} \tag{2.4}$$

and add an equation to the "single-protection" model (2.2):

$$\frac{d(I_{New})}{dt} = \beta\frac{SI}{N} + (1-\alpha_s)\beta\frac{S^pI}{N} \tag{2.5}$$

with initial conditions $I_{New}(0) = I^p_{New}(0) = 0$, and $I_{New}(0) = 0$ respectively. These new variables (I^p_{New} and I_{New}) represent cumulative HIV infections in PrEP users and non-users, respectively.

If PrEP is not available, the "no intervention" model becomes:

$$\begin{array}{rcl} \frac{dS}{dt} & = & \Lambda - \beta\frac{SI}{N} - \mu S \\ \frac{dI}{dt} & = & \beta\frac{SI}{N} - (\mu+d)I \\ \frac{d(I_{New})}{dt} & = & \beta\frac{SI}{N} \end{array} \tag{2.6}$$

with initial conditions $S(0) = (1-P)N(0)$, $I(0) = PN(0)$, and $I_{New}(0) = 0$.

I proceed with analysis of the behavior of the indicators assuming "dual protection". $[\ \ ]$ is used to denote variables from the model without PrEP (2.6) and $[\ \ ]_{DP}$ for variables from the "dual-protection" model with PrEP (2.4). Using these notations, the qualitative indicators have the following expressions:

$$
\begin{aligned}
C_I(T) &= \int_0^T ([\tfrac{d}{dt} I_{New}(t)] - [\tfrac{d}{dt} I^p_{New}(t) + \tfrac{d}{dt} I_{New}(t)]_{DP}) dt \\
F_I(T) &= \frac{\int_0^T ([\frac{d}{dt} I_{New}(t)] - [\frac{d}{dt} I^p_{New}(t) + \frac{d}{dt} I_{New}(t)]_{DP}) dt}{\int_0^T [\frac{d}{dt} I_{New}(t)] dt} \\
P_I(T) &= 1 - \frac{[\frac{I^p(T)+I(T)}{S^p(T)+S(T)+I^p(T)+I(T)}]_{DP}}{[\frac{I(T)}{S(T)+I(T)}]} \\
aI_I(T) &= 1 - \frac{[\frac{\int_T^{T+1} [\frac{d}{dt} I^p_{New}(t) + \frac{d}{dt} I_{New}(t)] dt}{S^p(T)+S(T)}]_{DP}}{[\frac{\int_T^{T+1} \frac{d}{dt} I_{New}(t) dt}{S(T)}]} \\
\hat{C}_I(T) &= [I(T)] - [I^p(T) + I(T)]_{DP} \\
\hat{F}_I(T) &= \frac{[I(T)] - [I^p(T)+I(T)]_{DP}}{[I(T)]} = 1 - \frac{[I^p(T)+I(T)]_{DP}}{[I(T)]}.
\end{aligned}
$$

Since integral evaluated on derivative function can be simplified, previous expressions of the indicators are equivalent to the following:

$$
\begin{aligned}
C_I(T) &= [I_{New}(T)] - [I^p_{New}(T) + I_{New}(T)]_{DP} \\
F_I(T) &= \frac{[I_{New}(T)] - [I^p_{New}(T) + I_{New}(T)]_{DP}}{[I_{New}(T)]} = 1 - \frac{[I^p_{New}(T) + I_{New}(T)]_{DP}}{[I_{New}(T)]} \\
P_I(T) &= 1 - \frac{[\frac{I^p(T)+I(T)}{S^p(T)+S(T)+I^p(T)+I(T)}]_{DP}}{[\frac{I(T)}{S(T)+I(T)}]} \\
aI_I(T) &= 1 - \frac{[\frac{I^p_{New}(T+1)+I_{New}(T+1)-(I^p_{New}(T)+I_{New}(T))}{S^p(T)+S(T)}]_{DP}}{[\frac{I_{New}(T+1)-I_{New}(T)}{S(T)}]} \\
\hat{C}_I(T) &= [I(T)] - [I^p(T) + I(T)]_{DP} \\
\hat{F}_I(T) &= \frac{[I(T)] - [I^p(T)+I(T)]_{DP}}{[I(T)]} = 1 - \frac{[I^p(T)+I(T)]_{DP}}{[I(T)]}.
\end{aligned}
$$

From these expressions it can be seen that the indicators F_I, P_I, aI_I, F_I and $\hat{F}_I$

are dimensionless and do not depend on the population size. The other two indicators C_I and $\hat{C}_I$ measure changes in population group sizes, and are not dimensionless.

2.2.2 Initial Dynamics of the Indicators

To understand the practical value of the qualitative indicators I examine their short, intermediate and long term dynamics. I begin with indicator approximations shortly after the start of the intervention. Using the initial conditions defined above, the following expressions associated with the initial indicators' behavior are obtained (details can be found in the Appendix of Zhao *et al.* (2013)):

$$\begin{aligned}
C_I &\approx [\alpha_s + (1-\theta)\alpha_i(1-\alpha_s k_1)]k_1\beta P(1-P)N(0)dt \\
F_I &\approx [\alpha_s + (1-\theta)\alpha_i(1-\alpha_s k_1)]k_1 \\
P_I &\approx [\alpha_s + (1-\theta)\alpha_i(1-\alpha_s k_1)]k_1\beta(1-P)dt \\
aI_I &\approx (1-\alpha_s k_1)[1-(1-\theta)\alpha_i k_1]\{1-(1-\alpha_s k_1)k_1[1-(1-\theta)\alpha_i k_1]\}k_1\beta P dt \\
\hat{C}_I &\approx [\alpha_s + (1-\theta)\alpha_i(1-\alpha_s k_1)]k_1\beta P(1-P)N(0)dt \\
\hat{F}_I &\approx [\alpha_s + (1-\theta)\alpha_i(1-\alpha_s k_1)]k_1\beta(1-P)dt.
\end{aligned}$$

Here I assume $dt = 1$ for the approximation for aI_I because the definition of the incidence indicator is on annual basis.

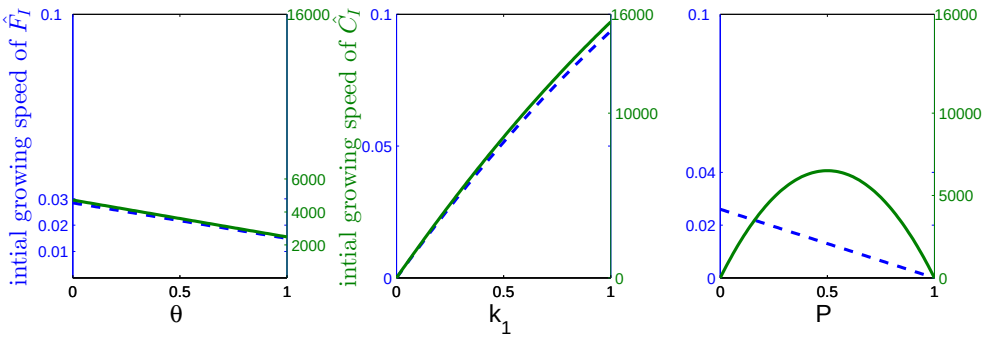
Note that the expression for the fractional indicator (F_I) depends only on the PrEP efficacy (α_s, α_i) and factors related to the implementation of the intervention at its start such as initial coverage (k_1) and the introductory control of the PrEP usage by HIV-positive individuals (θ) but not on the demographic, behavioral and epidemic parameters. Therefore, fractional indicator represents a metric of the "immediate impact of PrEP" on the HIV epidemic, which is independent of the specific population and the status of the HIV epidemic in it. This metric accounts for the effects of the reduced susceptibility (α_s) of the fraction k_1 of the population which initially uses PrEP combined with the reduced infectiousness (α_i) of a limited fraction $(1-\theta)k_1$ of

the infected population when in contact with partners unprotected by PrEP $(1-\alpha_s k_1)$. Clearly, if PrEP provides uni-directional protection ($\alpha_i = 0$) or none of the infected individuals is using PrEP ($\theta = 1$) then the "immediate impact of PrEP" is given by the product of PrEP efficacy and coverage ($\alpha_s k_1$). The initial behavior of all other indicators depend on the HIV prevalence (P) at the time of PrEP introduction as well as on the cumulative HIV-acquisition risk (β). Moreover, the cumulative (C_I) and reduction ($\hat{C}_I$) indicators also depend on the initial population size ($N(0)$), which is consistent with the fact that only indicators C_I and $\hat{C}_I$ measure changes on population group sizes, and are not dimensionless.

The initial rate of change of the indicators can be approximated as:

$$\begin{aligned}
{C_I}' &\approx [\alpha_s + (1-\theta)\alpha_i(1-\alpha_s k_1)]k_1\beta P(1-P)N(0) \\
{P_I}' &\approx [\alpha_s + (1-\theta)\alpha_i(1-\alpha_s k_1)]k_1\beta(1-P) \\
{aI_I}' &\approx (1-\alpha_s k_1)[1-(1-\theta)\alpha_i k_1]\{1-(1-\alpha_s k_1)k_1[1-(1-\theta)\alpha_i k_1]\}k_1\beta P \\
{\hat{C}_I}' &\approx [\alpha_s + (1-\theta)\alpha_i(1-\alpha_s k_1)]k_1\beta P(1-P)N(0) \\
{\hat{F}_I}' &\approx [\alpha_s + (1-\theta)\alpha_i(1-\alpha_s k_1)]k_1\beta(1-P).
\end{aligned}$$

Notice that initially ${C_I}' \approx {\hat{C}_I}'$ and ${P_I}' \approx {\hat{F}_I}'$. Now study the sensitivity of the initial rate of change of the reduction indicators ($\hat{C}_I$ and $\hat{F}_I$) to some of the intervention (θ, k_1) and epidemic (P) parameters by bifurcation simulations (Figure 2.4). These bifurcation parameters were chosen because they are easier to evaluate at community levels compared to HIV-acquisition risk and PrEP efficacy. The graphs in Figure 2.4, (a) and (b) demonstrate that the growth of both indicators accelerates if more people start on PrEP (larger k_1) but decelerates if the control of the PrEP usage by infected individuals is more effective (larger θ). The initial rate of change is more sensitive to k_1 than to θ but it is clear that the growth rate of both indicators at the time of PrEP introduction expresses qualitatively similar behavior with respect to the intervention parameters (θ, k_1). In contrast, the graphs presenting the dependence



(a) (b) (c)

Figure 2.4: Initial growth rate of reduction ($\hat{C}_I$) and fractional reduction ($\hat{F}_I$) indicators with respect to θ, k_1 and P. $\hat{C}_I$ is denoted by green solid line, while $\hat{F}_I$ is denoted by blue dashed line. All other parameters are fixed on their baseline parameter values from Table 2.1.

on the initial HIV prevalence show serious discrepancies (Figure 2.4(c)). The initial growth rate of the reduction indicator ($\hat{C}_I$) increases when the HIV prevalence ranges from 0 to 50% which includes all realistic values observed so far, particularly in Sub-Saharan Africa (Figure 2.4(c)). In comparison, the increase in HIV prevalence within the same range implies smaller growth rate of the fractional reduction indicator $\hat{F}_I$.

2.2.3 Asymptotic Behavior of the Indicators

In resource-constrained settings, it is unrealistic to expect the HIV epidemic will die out without additional intervention. Therefore, in the following, I assume that the basic reproduction number of the "no intervention" model is $R_0 = \frac{\beta}{\mu+d} > 1$. The asymptotic HIV prevalence in this case is given by: $[\frac{I}{S+I}] = 1 - \frac{\mu+d}{\beta} = 1 - \frac{1}{R_0}$.

I want to point out that if the PrEP intervention is strong enough to cause the eradication of HIV in the population, i.e., the HIV epidemic approaches the disease-

free equilibrium with the "dual-protection" model, then the asymptotic behavior of all indicators is well determined: i) the cumulative indicator will grow to infinity; ii) the reduction indicator will stabilize at $[I] = \frac{R_0-1}{\beta-d}\Lambda$; and iii) all other indicators will approach one. Unfortunately, PrEP intervention alone is unlikely to be sufficient to eradicate HIV. In that case it is shown that the asymptotic behavior of the indicators can be expressed in terms of the asymptotic proportion (p) of the HIV-positive subpopulation which have been infected while using PrEP (details are presented in Appendix of Zhao *et al.* (2013)), where p is defined as follows:

$$p = [\frac{I^p}{I^p + I}]_{DP}.$$

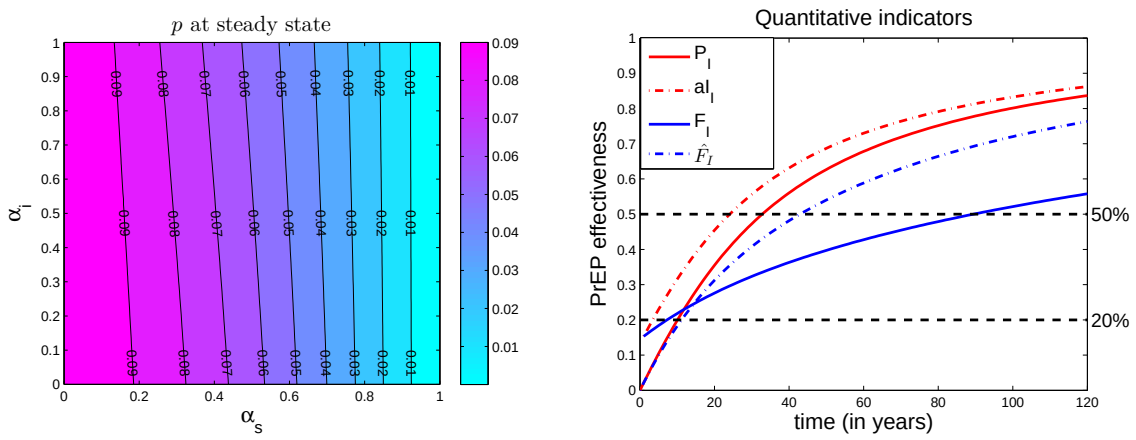
Expressions for the asymptotic values of four of the indicators:

$$\begin{aligned} P_I &= 1 - \frac{R_0 - \frac{1-\alpha_s(1-p)}{(1-\alpha_s)(1-\alpha_i p)}}{R_0 - 1} \\ aI_I &= 1 - \frac{R_0 \frac{(1-\alpha_s)(1-\alpha_i p)}{1-\alpha_s(1-p)} - 1}{R_0 - 1} \\ \hat{C}_I &= [\frac{R_0-1}{\beta-d} - \frac{R_0(1-\alpha_i p)(1-\alpha_s k)-1}{\beta(1-\alpha_i p)(1-\alpha_s p)-d}]\Lambda \\ \hat{F}_I &= 1 - \frac{R_0(1-\alpha_i p)(1-\alpha_s k)-1}{R_0-1}\frac{\beta-d}{\beta(1-\alpha_i p)(1-\alpha_s p)-d} \end{aligned}$$

and the asymptotic rate of growth of the cumulative indicator:

$$C_I{}' = (\mu + d)[\frac{R_0-1}{\beta-d} - \frac{R_0(1-\alpha_i p)(1-\alpha_s k)-1}{\beta(1-\alpha_i p)(1-\alpha_s p)-d}]\Lambda = (\mu + d)\hat{C}_I.$$

show that they are independent of the initial HIV prevalence (P) and the initial control on the PrEP use by HIV-positive individuals (θ) which have been of critical importance for the initial dynamics of the indicators. Cumulative indicators ($\hat{C}_I$ and C_I) depend indirectly on the population size (N) which determines the entry rate in the population (Λ). Notice that $C_I{}' = (\mu+d)\hat{C}_I$ so the value of the reduction indicator is proportional to the annual number of new infections prevented due to PrEP use in the long term. The rest of the indicators are not influenced by the population size (N). Although recruitment parameters such as k and Λ are not explicitly present in



(a) (b)

Figure 2.5: (a) Asymptotic values of the proportion of the HIV-positive individuals who have been infected while using PrEP (p) as a function of the PrEP efficacies (α_s and α_i) assuming 10 % PrEP coverage ($k = k_1 = 0.1$). (b) Long term dynamics of the quantitative indicators based on simulations with "dual-protection" and "no intervention" models using parameters from Table 2.1.

some of the expressions above they may affect the asymptotic proportion of PrEP users among infected sub-population (p).

Because of the algebraic complexity, I can not explicitly express p. Therefore, a good approximation of p is important for the evaluation of the asymptotic levels of the indicators.

Now study the variation of p when the reduction in susceptibility and infectiousness range from 0 to 100 % (Figure 2.5(a)) for intervention coverage ($k = 0.1$) which is not sufficient to eradicate HIV even if the PrEP protection against HIV is perfect ($\alpha_s = \alpha_i = 1$). It shows that p depends greatly on the reduction in susceptibility (α_s) and very little on the reduction in infectiousness (α_i). It is clear that the fraction of infections which occur when using PrEP (p) ranges from zero, in case that PrEP provides complete protection against HIV ($\alpha_s = 1$) and no PrEP users ever

get infected, to the level of the PrEP coverage (k) in case that PrEP is completely ineffective ($\alpha_s = 0$) and infections are proportionally distributed among PrEP users and non-users.

The next goal is to examine and compare the long-term behavior of the indicators (specifically those expressed as ratios) for a fixed PrEP intervention (Figure 2.5(b)). Although qualitatively similar the trajectories of the indicators show some important differences. First, some indicators such as the reduction in HIV prevalence and the reduction in the infected fraction start at zero while others such as the fraction of prevented infections and the reduction in HIV incidence initiate at positive values. Therefore, it is not surprising that the indicators reach a specific threshold of 20% at times varying from 3 to 11 years after the introduction of PrEP. The times needed to report 50% effectiveness are even farther apart. It takes the intervention 24 years and 33 years to reduce in half the expected HIV incidence and HIV prevalence, respectively. However, almost 90 years are necessary to reduce the cumulative number of new infections by 50%, i.e., such reduction is infeasible over traditionally used evaluation periods of up to 30 years.

2.2.4 Evaluation of the Public-health Impact of PrEP

The initial and asymptotic behavior of the indicators are useful in understanding what drives the observed differences in their projections. However, from a public health perspective it is more important to analyze the indicators values over more practical time intervals. Although no fixed standards exist, the majority of the quantitative studies assume that preventive interventions are implemented for 10 years when their effectiveness is evaluated. The same period is recommended by the World Health Organizations as an evaluation period when cost-effectiveness analyses are conducted (Edejer (2003)). Longer periods are investigated in few studies but always

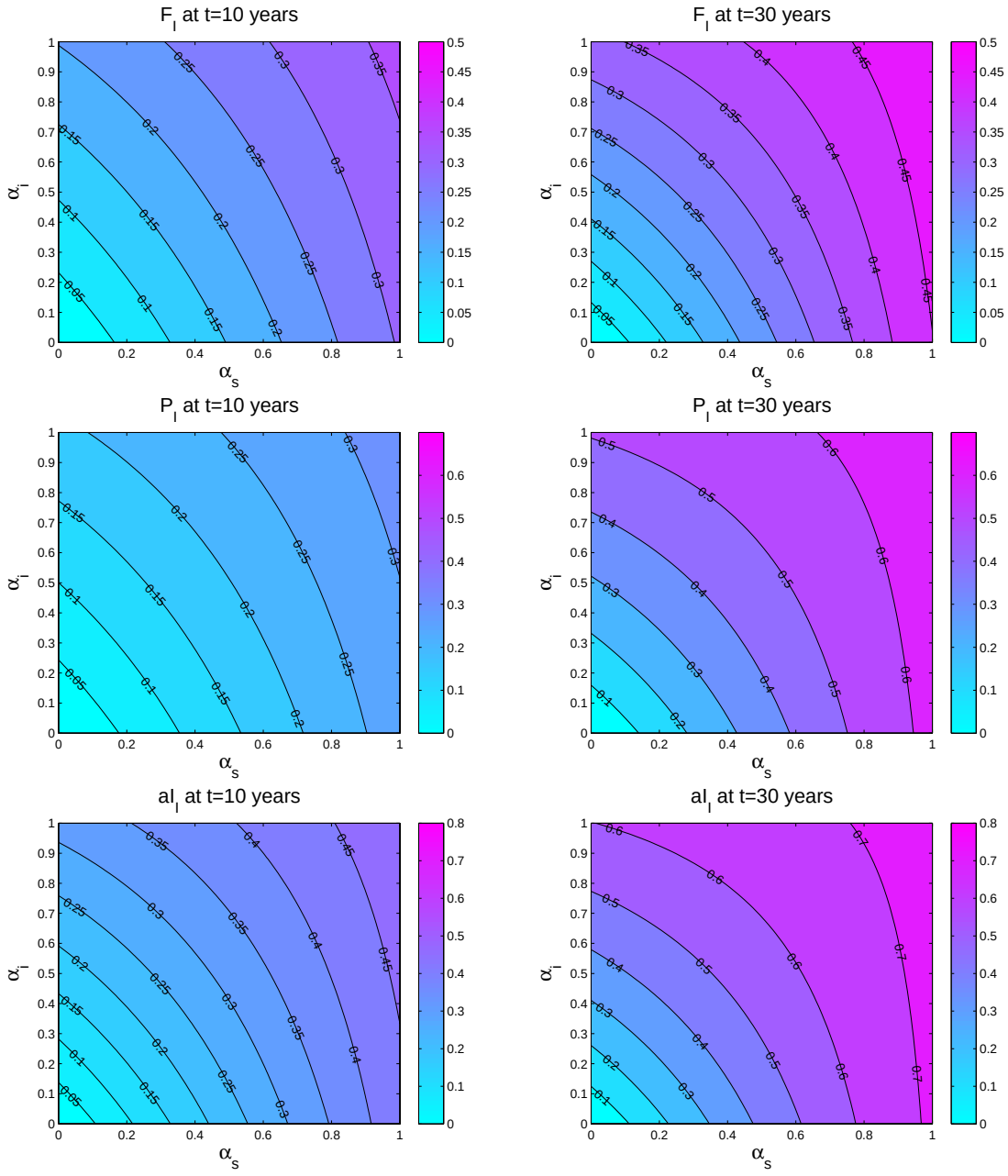


Figure 2.6: Contour plots of the indicators (F_I, P_I, and aI_I) over 10 and 30 years with respect to selected intervention parameters α_i and α_s. All other parameters are fixed at the baseline parameter values from Table 2.1.

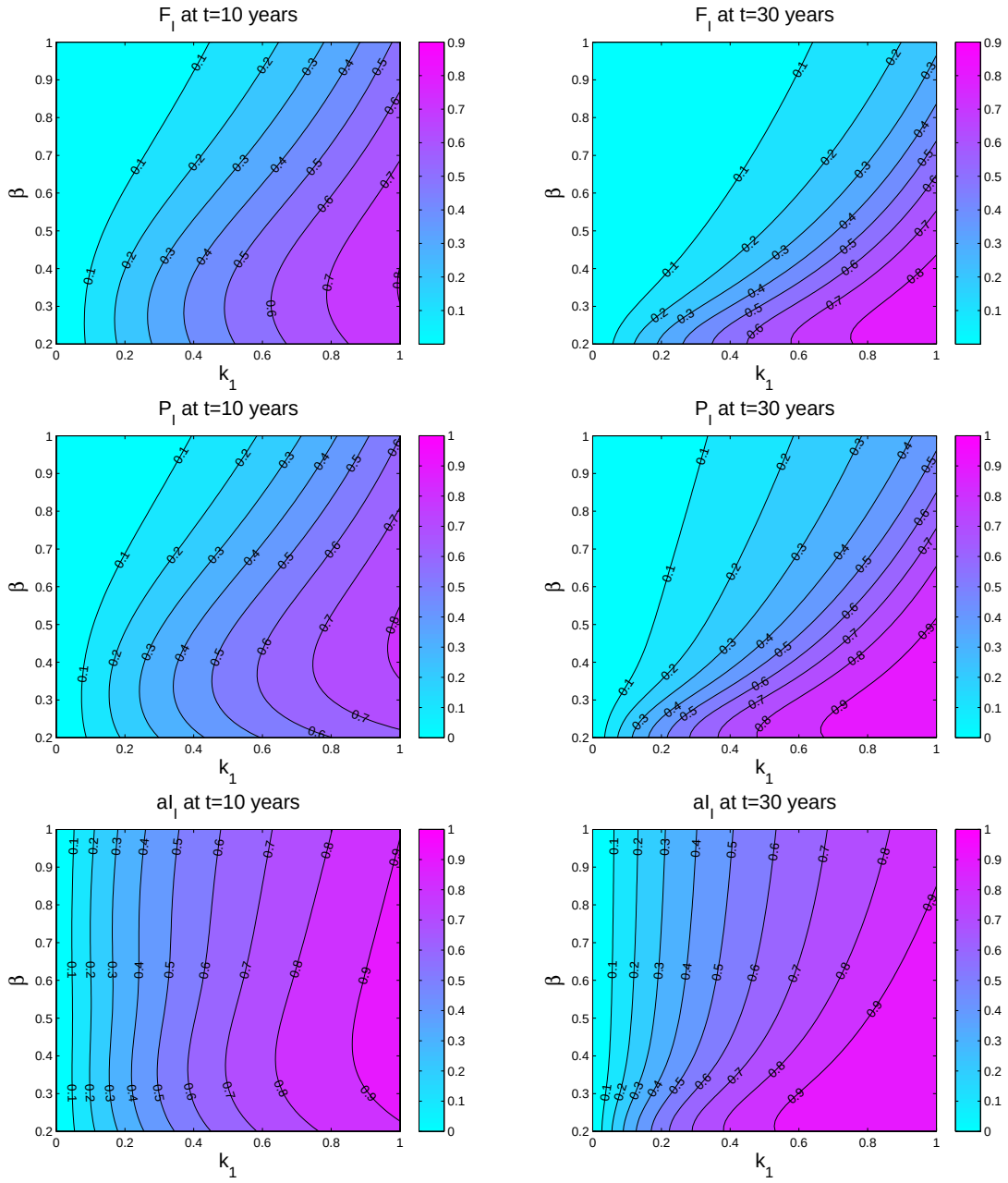


Figure 2.7: Contour plots of the indicators (F_I, P_I, and aI_I) over 10 and 30 years with respect to selected epidemic parameters β and k_1. All other parameters are fixed at the baseline parameter values from Table 2.1.

up to 30 years.

In this section it is explored the dependence of the indicator readings over 10 and 30 years on key epidemic and intervention parameters. Clearly, the impact of PrEP is positively correlated with both reductions in susceptibility (α_s) and infectiousness (α_i) regardless of what indicator is used to quantify it (Figure 2.6). The slopes of the contour plots in the PrEP efficacy parameter space show that if the susceptibility efficacy(α_s) is relatively low (up to 30 %) all indicators are equally dependent of both α_s and α_i. However, with the increase of the PrEP protection against HIV the influence of the reduction of infectiousness decreases significantly. The prevalence indicator (P_I) projects the least effectiveness over 10 years of PrEP use. It predicts that more than 70 % and 55% PrEP efficacy is needed to achieve 20% reduction in HIV prevalence with uni-directional ($\alpha_i = 0$) and bi-directional ($\alpha_i = \alpha_s$) interventions, respectively. In comparison, 20% reduction in the expected HIV infections is possible with 65% effective uni-directional and 55% bi-directional PrEP while 20% reduction in HIV incidence is feasible even if less than 45% effective uni-directional and 30% bi-directional PrEP is used over 10 years. The order of predicted effectiveness by the prevalence and the fractional indicators is reversed over an evaluation period of 30 years. More than half of the parameter space results in more than 50% reduction in HIV prevalence (P_I), an unreachable threshold as a reduction in expected infections (F_I).

All indicators increase with coverage $k = k_1$ (Figure 2.7). The prevalence and fractional indicators are more sensitive to changes in the transmission rate (β) than is the annual incidence indicator, with increasing influence of β on the fraction of prevented infection for larger evaluation periods. The maximum PrEP effectiveness over 10 years of PrEP use is predicted for complete coverage (k) and modest level of the transmission rate (β) while over 30 years it is achieved for the lowest possible β.

2.3 Sensitivity Analysis

Finally, it is explored the sensitivity of the indicators to changes in each parameter. Using the algorithm presented in Blower and Dowlatabadi (1994) I calculate the Partial Rank Correlation Coefficients (PRCC) which evaluate the monotonicity of the model outcomes (indicators) in terms of the model parameters. Values of $PRCC$ closer to ± 1, imply stronger correlation between the output indicator and the input parameter while the sign of the coefficients determines if the outcomes grow or decrease with an increase of the input parameters.

I study separately the sensitivity of the indicators to the parameters (b_a, n, μ, and d) which are fitted using data from South Africa. In the analysis I choose 1000 random parameters combinations of those input parameters sampled uniformly from their corresponding ranges: $[0.0015, 0.0045]$ for b_a, $[32.9247, 98.7741]$ for n, $[0.0125, 0.0375]$ for μ, and $[0.0651, 0.1953]$ for d. Each range is chosen as $[0.5, 1.5]*$(baseline parameter value in Table 2.1). The rest of the parameters are fixed on their baseline values in Table 2.1. For each parameter set I simulate the models with "dual-protection" and no intervention and calculate PRCC matrix of all six indicators C_I, F_I, P_I, aI_I, $\hat{C}_I$, and $\hat{F}_I$ for the first 10 years (standard analysis) as well as for the 100 years (long-term analysis). Similarly, I investigate the indicators' sensitivity to the remaining epidemic and intervention parameters (α_s, α_i, P, $k_1(k = k_1)$, and θ), uniformly sampled from their corresponding ranges: $[0.25, 0.75]$ for α_s, $[0.25, 0.75]$ for α_i, $[0.083, 0.249]$ for P, $[0.1, 0.3]$ for k_1, and $[0.25, 0.75]$ for θ. Results are presented in Figure 2.8.

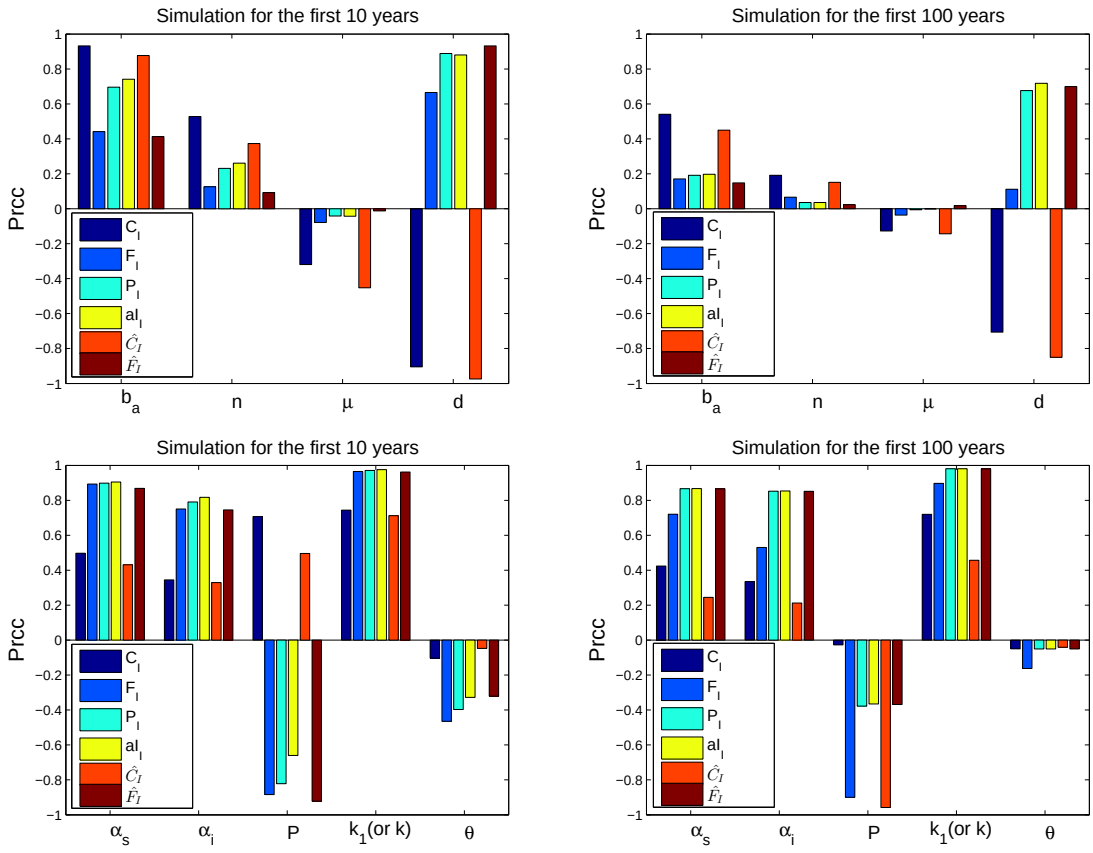


Figure 2.8: Partial rank correlation coefficients (PRCC) between model parameters and the quantitative indicators over 10 and 100 years.

Correlations for 10-year intervention suggest that from the fitted parameters the indicators are most sensitive to the factors (b_a and n) which determine the transmission rate β. However, their influence over time decreases. C_I and $\hat{C}_I$ are still positively correlated to the two factors while the rest of the indicators are negatively correlated to the two factors over 100 years, both dependencies are weak. The intervention outcomes are split into two groups with respect to their correlation with the HIV induced mortality (d): the cumulative indicators being negatively correlated while the rest being positively correlated with d. Similar discrepancy is observed with respect to the influence of the initial HIV prevalence P over 10 years but the corre-

lations are reversed (positive - for the the cumulative indicators and negative for the rest). Interestingly, in that case the difference between the indicators disappears in a long term. Note that although P appears in the initial conditions only, it continues to have strong influence on all the cumulative and reduction indicators for more than 10 years while its impact on the fraction of prevented infection gets even stronger over time.

Among the intervention parameters, PrEP coverage (k) and PrEP efficacies per act (α_s and α_i) express strong positive correlation with all the indicators in a short term. It remains significant in a long term for all outcomes. This confirms that PrEP coverage and protection level are critical to the intervention success regardless which qualitative metric is used. In contrast, the influence of the initial control on the PrEP use by HIV-positive individuals (θ) diminishes substantially in time.

The prevalence (P_I) and the annual incidence (aI_I) indicators have almost the same sensitivity to all parameters. Therefore they should have consistent projections when evaluating the impact of the intervention.

2.4 Discussion

Precise evaluation of the expected public-health impact of biomedical interventions for HIV preventions becomes increasingly important with more prevention options entering the pipeline toward licensure. The practice shows that even if the products are effective in reducing the individual risk of acquisition (individual efficacy) the benefits from general usage (population effectiveness) may be limited by variety of epidemic, behavioral and intervention factors.

It is demonstrated that the quantitative indicators have distinct dynamical profile shortly after the start of PrEP intervention which modifies substantially over time. As a result, when calculated over a fixed period of time these indicators may

project significantly different PrEP effectiveness and therefore influence the decision if particular products are potentially good enough for implementation. In general, new prevention methods need to prove their effectiveness in randomized clinical trial (RCT), i.e., to demonstrate that the observed efficacy is significantly larger than zero (positive 95% confidence interval), before applying for licensure. In reality, developers and public-health officials try to avoid PrEP products with low efficacy because the controlled environments of the clinical trials are difficult to be replicated at community level. Another concern is that the availability of PrEP may affect sexual behavior and encourage risky sex practices. Therefore, minimal efficacy thresholds of 20% or higher are often included in the design of RCTs and similar levels of effectiveness is expected when interventions are modeled at population level (Dimitrov *et al.* (2013); Grant *et al.* (2010)). Other studies imply that 50% biological efficacy is needed to guarantee significant public-health impact. The question remains what does 20% or 50% PrEP effectiveness mean? It has been shown the widely used evaluation metrics may disagree over practical intervals of time (Figure 2.5). A reduction in HIV incidence at pre-specified levels seems most realistic as an intervention goal but it is not easy to be estimated in the population. In contrast, a reduction in HIV prevalence is easier to be recorded but more difficult to be achieved in a short term. The reduction in the number of new HIV infections, which is the most popular public-health metric, projects strong PrEP effectiveness initially but grows slower than the other indicators over time.

Moreover, if used to compare the impact of PrEP interventions different indicators may give preference to different options (Figure 2.3). Public-health officials who consider PrEP to be integrated in HIV prevention programs should base their decision on a complex of quantitative metrics. Although is specifically focused on HIV prevention, the same theoretical approach may be extended to model other in-

fectious diseases, such as malaria, cholera, and tuberculosis, or evaluate the impact of interventions, such as male circumcision, vaccination or quarantine strategies.

Presented results, assuming perfect adherence and instantaneous uptake, are likely to give optimistic views of the potential impact of a PrEP intervention. Although, overall self-reported adherence in the concluded clinical trials is high it is unclear how consistently PrEP will be used in real settings. Perfect adherence and other simplifying assumptions allowed to support the observations on the indicators and their simulated dynamics with analytical expressions which were easier to be interpreted. I believe that more complex and realistic modeling setup will be more useful in projecting benefits due to PrEP use but it is unlikely to resolve the differences between the interventional outcomes reported in this section.

Chapter 3

DEMOGRAPHIC ENTRANCE RATES FOR MATHEMATICAL MODELING OF HIV INTERVENTIONS

The most often mathematically modeled prevention interventions for HIV are male circumcision, test and treat strategy as prevention, microbicides, and PrEP.

Three randomized controlled trials have shown that male circumcision reduces the risk of heterosexually acquired HIV infection in men by approximately 60%. Male circumcision for HIV prevention has been broadly modeled in heterosexual population (Alsallaq *et al.* (2013, 2009); Andersson *et al.* (2011); Cox *et al.* (2011); Dushoff *et al.* (2011); Hallett *et al.* (2008); Nagelkerke *et al.* (2007); Podder *et al.* (2007); Williams *et al.* (2006)), and also has been modeled in male homosexual population(Londish *et al.* (2010)).

Standard antiretroviral therapy (ART) consists of the combination of at least three antiretroviral (ARV) drugs to suppress the HIV virus and stop the progression of HIV disease. Besides individual-wise benefits, expanded access to ART can also reduce the HIV transmission at population level antiretroviral, so treatment is also considered as an HIV prevention strategy. Mathematical models have been used to study antiretroviral therapy (Alsallaq *et al.* (2013); Cremin *et al.* (2013); Law *et al.* (2001); Lima *et al.* (2008)), also to discuss treatment as prevention or 'test and treat' strategy (Andrews *et al.* (2012); Hallett *et al.* (2009); Sorensen *et al.* (2012)). Within these papers, Granich *et al.* (2009); Wagner and Blower (2012) proposed that universal test and treat of HIV may be used as a prevention method, and it may eliminate the HIV.

Microbicides are compounds that can be applied inside the vagina or rectum to

protect against sexually transmitted infections (STIs) including HIV. Many papers have used mathematical models to study this HIV prevention method (Boily *et al.* (2011); Breban *et al.* (2006); Cox *et al.* (2011); Dimitrov *et al.* (2010); Karmon *et al.* (2003); Wilson *et al.* (2008)).

PrEP is a new HIV prevention method in which people who do not have HIV infection take a pill daily to reduce their risk of becoming infected. Only people who are HIV-negative should use PrEP. An HIV test is required before starting PrEP and then every 3 months while taking PrEP. PrEP, although is a relatively new HIV prevention method, has also been extensively mathematically modeled (Abbas *et al.* (2007); Cremin *et al.* (2013); Desai *et al.* (2008); Dimitrov *et al.* (2012); Grant *et al.* (2010); Juusola *et al.* (2012); Nichols *et al.* (2013); Supervie *et al.* (2011, 2010); Zhao *et al.* (2013)).

Using related key words, I collected and screened papers from 'Web of Knowledge' database and 'PubMed' database. For each related paper, I collect the information about the population being modeled, recruitment mechanism, mechanisms of departures from the population, assumptions regarding migration. The results are summarized in Table B.1, B.2, and B.3.

As observed from this summary, models use different demographic assumptions on the population entrance rate and departure rate. Models use either constant entrance rate or linear entrance rate. Although logistic entrance rate has not been observed in the collected published papers, I will include it in the main study to be compared with constant and linear entrance rates. Population departure rates are assumed in these papers, while most of the papers do not include the migration in their studies.

The following study will discuss the different assumptions on the population entrance rate, on the study of PrEP intervention. Except that no evidence shows that PrEP reduces the infectiousness of HIV positive people, and I assume $\alpha_i = 0$, all the

modeling assumptions will be the same as in section 2.1.1. And for briefness, I do not present the details for model descriptions and parameterizations in this section, which are similar to what have been done in section 2.1. Indicators are also defined in the same way as in section 2.2.1.

3.1 Models without Intervention

First being studied is the impact of the assumptions about the recruitment rate on the population dynamics in absence of PrEP.

$$\begin{array}{rcl} \frac{dS}{dt} & = & f(N) - \beta\frac{SI}{N} - \mu S \triangleq P(S,I) \\ \frac{dI}{dt} & = & \beta\frac{SI}{N} - (\mu + d)I \triangleq Q(S,I). \end{array} \tag{3.1}$$

with $S(0) = (1-P)N(0)$ and $I(0) = PN(0)$, where P is the initial HIV prevalence in the initial population of size $N(0)$.

Here the total population N is divided into two major classes, susceptibles (S) and infected (I). Frequency-dependent transmission is assumed and the cumulative HIV acquisition risk per year β is calculated based on the HIV risk per act (β_a) with a HIV-positive partner and the average number of sex acts per year (n):

$$\beta = 1 - (1 - b_a)^n.$$

Individuals join the population (become sexually active) at rate $f(N)$. Recruitment formulations $f(N) = \Lambda$, $f(N) = rN$, and $f(N) = rN(1 - \frac{N}{K})$ will be explored, corresponding to constant, proportional to size and logistic entrance rate, respectively. All model parameters are described in Table 3.1.

3.1.1 Models with Constant Recruitment

In this section, it is assumed that $f(N) = \Lambda$ in model (3.1).

Table 3.1: Parameter description for models with different entrance rates

Par.	Description
Λ	Annual rate at which individuals become sexually active
r	Growth rate at which individuals become sexually active
K	Population carrying capacity
b_a	HIV acquisition risk per act
n	Number of sexual acts per year per individual
β	Cumulative HIV-acquisition risk(calculated from b_a and n)
$\frac{1}{\mu}$	Time (in years) to remain sexually active
d	HIV carrier's annual rate of progression to AIDS
k	Proportion of the new recruits that start using PrEP
α_s	Efficacy of PrEP in reducing susceptibility of PrEP users

Proposition 3.1.1.1. *With nonnegative initial conditions all solutions of model (3.1) are nonnegative and bounded with total population size* $N(t) \leq \max\{N(0), \frac{\Lambda}{\mu}\}$.

- *When the basic reproduction number* $R_0 = \frac{\beta}{\mu+d} < 1$, *model (3.1) has an unique disease-free equilibrium* $E_0 = (\frac{1}{\mu}\Lambda, 0)$ *which is (globally) stable.*
- *When* $R_0 > 1$, E_0 *is unstable and the model possesses an unique endemic equilibrium* $E^* = (\frac{1}{\beta-d}\Lambda, \frac{1}{(\beta-d)} \cdot (R_0 - 1)\Lambda)$ *which is (globally) stable.*

Proof. Proof can be found in Appendix C. Also see Hwang and Kuang (2003). □

3.1.2 Models with Linear Recruitment

In this section, it is assumed that $f(N) = rN$ in model (3.1).

Notice that

$$\lim_{(S,I)\to(0,0)} \frac{SI}{N} = \lim_{(S,I)\to(0,0)} \frac{SI}{S+I} = 0.$$

Thus it is defined that $P(0,0) = 0$ and $Q(0,0) = 0$. With this assumption, $E = (0,0)$ is a unique steady state of (3.1), and both $P(S,I)$ and $Q(S,I)$ are continuous on $\{(S,I)|S \geq 0, I \geq 0\}$.

Proposition 3.1.2.1. *With nonnegative initial conditions all solutions of model (3.1) are nonnegative.*

The unique steady state ($E = (0,0)$) implies population extinction. A solution of (3.1) either approaches E or the population size grows unbounded under endemic or infection-free conditions, depending on different parameter values:

- *When $r < \mu$, $(S(t), I(t)) \to (0,0)$.*
- *When $r > \mu$, and further $\beta > d$ and $\mu + d > r$,*
  - *when $\beta < \mu + d$, $(S(t), I(t)) \to (\infty, 0)$ unbounded infection free;*
  - *when $\mu + d < \beta < \frac{(\mu+d)d}{\mu+d-r}$, $(S(t), I(t)) \to (\infty, \infty)$ unbounded endemic;*
  - *when $\beta > \frac{(\mu+d)d}{\mu+d-r}$, $(S(t), I(t)) \to (0,0)$ extinction.*

See Appendix C for proof. Notice that boundedness of solutions is not observed.

HIV prevalence is estimated and reported periodically in the statistical data. Therefore to study the projected HIV prevalence by the model, the dynamics of the fractional form of model (3.1) in terms of $s = \frac{S}{N}$ and $i = \frac{I}{N} = 1 - s$ are also analyzed:

$$\begin{aligned}\frac{ds}{dt} &= [r-(\beta-d)s](1-s)\\ \frac{dN}{dt} &= [r-\mu-d(1-s)]N\\ \frac{dI}{dt} &= [\beta s-(\mu+d)]I.\end{aligned}$$

Notice that the first equation is independent of N which allows to study $s(t)$ directly. The following proposition summarizes the solution behavior for different parameter values and initial conditions.

Proposition 3.1.2.2. *With no infected individuals initially ($i(0)=0$,$s(0)=1$), the population remains disease free ($s(t)\equiv 1$) with population size given by $\frac{dN}{dt}=(r-\mu)N$.*

Any other initial conditions result in one of the following cases (assume $r>\mu$, $\beta>d$, and $\mu+d>r$):

- *when $\beta<\mu+d$, $(S(t),I(t))\to(\infty,0)$ with $(s(t),i(t),N(t))\to(1,0,\infty)$;*
- *when $\mu+d<\beta<r+d$, $(S(t),I(t))\to(\infty,\infty)$ with $(s(t),i(t),N(t))\to(1,0,\infty)$;*
- *when $r+d<\beta<\frac{(\mu+d)d}{\mu+d-r}$, $(S(t),I(t))\to(\infty,\infty)$ with $(s(t),i(t),N(t))\to(\frac{r}{\beta-d},1-\frac{r}{\beta-d},\infty)$;*
- *when $\beta>\frac{(\mu+d)d}{\mu+d-r}$, $(S(t),I(t))\to(0,0)$ with $(s(t),i(t),N(t))\to(\frac{r}{\beta-d},1-\frac{r}{\beta-d},0)$.*

See Appendix C for proof.

3.1.3 Models with Logistic Recruitment

In this section, it is assumed that $f(N)=rN(1-\frac{N}{K})$ in model (3.1).

Notice that

$$\lim_{(S,I)\to(0,0)}\frac{SI}{N}=\lim_{(S,I)\to(0,0)}\frac{SI}{S+I}=0.$$

It is defined that $P(0,0) = 0$ and $Q(0,0) = 0$. Under this assumption, $E_{00} = (0,0)$ is a steady state of (3.1) and the right-hand sides ($P(S,I)$ and $Q(S,I)$) of (3.1) are continuous in $\{(S,I)|S \geq 0, I \geq 0\}$.

Proposition 3.1.3.1. *The biologically relevant region* $\{(S(t),I(t))|S(t) \geq 0, I(t) \geq 0, S(t)+I(t) \leq K\}$ *is positively invariant with respect to (3.1).*

The model has three possible steady states: population extinction $E_{00} = (0,0)$, *disease-free* $E_{01} = (\frac{r-\mu}{r}K, 0)$ *and endemic* $E^* = \frac{r-\mu-d(1-\frac{1}{R_0})}{r}K(\frac{1}{R_0}, \frac{R_0-1}{R_0})$ *where* $R_0 = \frac{\beta}{\mu+d}$.

The following global stability results hold.

- *When* $r < \mu$, $(S(t),I(t)) \to E_{00}$ *the extinction steady state.*
- *When* $r > \mu$, *and further* $\beta > d$ *and* $\mu + d > r$,
  - *when* $\beta < \mu + d$, $(S(t),I(t)) \to E_{01}$ *the infection free steady state, while* E_{00} *is unstable and* E^* *does not exist;*
  - *when* $\mu + d < \beta < \frac{(\mu+d)d}{\mu+d-r}$, $(S(t),I(t)) \to E^*$ *the endemic steady state, while* E_{00} *and* E_{01} *are unstable;*
  - *when* $\beta > \frac{(\mu+d)d}{\mu+d-r}$, $(S(t),I(t)) \to E_{00}$ *the extinction steady state, while* E_{01} *and* E^* *are unstable.*

See Appendix C for proof.

The behavior of the model (3.1) can be further studied by investigating its fractional form in terms of $s = \frac{S}{N}$ and $i = \frac{I}{N} = 1 - s$:

$$\begin{aligned} \frac{ds}{dt} &= [r(1-\tfrac{N}{K}) + (d-\beta)s](1-s) \\ \frac{dN}{dt} &= [r(1-\tfrac{N}{K}) - \mu - d(1-s)]N. \end{aligned} \tag{3.2}$$

Proposition 3.1.3.2. *Model (3.2) is used to study the local stability of the extinction steady state E_{00}. The following results hold:*

- *E_{00} is unstable when $\beta < \frac{(\mu+d)d}{\mu+d-r}$;*
- *E_{00} is stable when $\beta > \frac{(\mu+d)d}{\mu+d-r}$.*

See Appendix C for proof.

Table 3.2 summarizes the long-term dynamic results when the model (3.1) is using different recruitment rates. Notice that $\beta = \tilde{\beta} \triangleq \frac{(\mu+d)d}{\mu+d-r}$ makes $E_4 = (\frac{1}{R_0}, \frac{r-\mu-d(1-\frac{1}{R_0})}{r}K)$ $= E_2 = (\frac{r}{\beta-d}, 0)$ for (3.2).

Table 3.2: Stability conditions for the models in absence of PrEP. Extinction steady state is globally stable when $r < \mu$, so in the table it is assumed that $r > \mu$ and further assumed that $\beta > d$ and $\mu + d > r$ for linear and logistic entrance rates.

Recruitment type	Parameter conditions	Outcomes	HIV prevalence
Constant	$\beta < \mu + d$	disease free state	0
	$\beta > \mu + d$	endemic state	$1 - \frac{\mu+d}{\beta}$
Linear	$\beta < \mu + d$	disease free state	0
	$\mu + d < \beta < \frac{(\mu+d)d}{\mu+d-r}$	endemic state	0 or $1 - \frac{r}{\beta-d}$
	$\beta > \frac{(\mu+d)d}{\mu+d-r}$	extinction	$1 - \frac{r}{\beta-d}$
Logistic	$\beta < \mu + d$	disease free state	0
	$\mu + d < \beta < \frac{(\mu+d)d}{\mu+d-r}$	endemic state	$1 - \frac{\mu+d}{\beta}$
	$\beta > \frac{(\mu+d)d}{\mu+d-r}$	extinction	$1 - \frac{r}{\beta-d}$

3.2 Models with PrEP Intervention

Next, it will be investigated the influence of the recruitment mechanisms on the projected impact of PrEP interventions. The model with intervention is formulated by the following system of differential equations:

$$\begin{aligned} \frac{dS^p}{dt} &= kf(N) - (1-\alpha_s)\beta\frac{S^pI}{N} - \mu S^p \\ \frac{dS}{dt} &= (1-k)f(N) - \beta\frac{SI}{N} - \mu S \\ \frac{dI}{dt} &= \beta\frac{SI}{N} + (1-\alpha_s)\beta\frac{S^pI}{N} - (\mu+d)I \end{aligned} \tag{3.3}$$

with initial conditions:

$$\begin{aligned} S^p(0) &= k(1-P)N(0) \\ S(0) &= (1-k)(1-P)N(0) \\ I(0) &= PN(0). \end{aligned}$$

Here P is the initial HIV prevalence, $N(0)$ is the initial population size, and k is the initial coverage of PrEP among susceptible individuals. All parameters are explained in Table 3.1.

In this model the population is divided into three major classes, according to their HIV and PrEP status: susceptible individuals who don't use PrEP (S); susceptible PrEP users (S^p) and infected individuals (I). A constant proportion k of the new recruits are assumed to start using PrEP. The same proportion of the susceptible individual are assumed to start on PrEP initially. Since PrEP provides imperfect protection against HIV some of the PrEP users become infected. The risk of drug-resistance emergence among infected PrEP users has been discussed in the HIV prevention community (Dimitrov *et al.* (2012); Supervie *et al.* (2011, 2010)) and wide-scale PrEP interventions will likely include periodic HIV screening of all prescribed users. Therefore, it is assumed that PrEP users stop

using the product after acquiring HIV and all infected individual accumulate in the compartment (I). The basic reproduction number of model (3.3) is given by $R_0 = (1-k)R_0(S) + kR_0(S^p) \triangleq (1-k)\frac{\beta}{\mu+d} + k\frac{(1-\alpha_s)\beta}{\mu+d} = \frac{(1-\alpha_s k)\beta}{\mu+d}$.

Different expressions ($f(N) = \Lambda$, $f(N) = rN$, and $f(N) = rN(1-\frac{N}{K})$) will be explored for the population recruitment.

3.2.1 Intervention Models with Constant Recruitment

In this section, it is assumed that $f(N) = \Lambda$ in model (3.3).

Proposition 3.2.1.1. *With nonnegative initial conditions, solutions for (3.3) are nonnegative and bounded with* $N(t) = \max\{N(0), \frac{\Lambda}{\mu}\}$. *Now assume* $\beta > d$.

- *When* $R_0 < 1$ *the model (3.3) has an unique (disease-free) equilibrium* $E_0 = (\frac{k}{\mu}\Lambda, \frac{1-k}{\mu}\Lambda, 0)$ *which is stable. Global stability of* E_0 *has been proved given extra condition* $R_0 < \frac{\mu}{\mu+d}$.

- *When* $R_0 > 1$ *the disease-free equilibrium* E_0 *is unstable. The system has a single endemic equilibrium* $E^* = (S^{p*}, S^*, I^*)$ *which satisfies* $S^{p*} = \frac{\Lambda(\Lambda - dI^*)}{\mu} \cdot \frac{k}{\Lambda + ((1-\alpha_s)\beta - d)I^*}$ *and* $S^* = \frac{\Lambda(\Lambda - dI^*)}{\mu} \cdot \frac{1-k}{\Lambda + (\beta - d)I^*}$, *where* I^* *is the unique solution of* $F(I) \triangleq \Lambda - \frac{k\Lambda(\Lambda - dI)}{(1-\alpha_s)\beta I + \Lambda - dI} - \frac{(1-k)\Lambda(\Lambda - dI)}{\beta I + \Lambda - dI} - (\mu + d)I = 0$ *in the interval* $(0, \frac{\Lambda}{d})$. *The endemic equilibrium* E^* *is stable when* $\frac{\mu+d}{d}[(1-\alpha_s)\beta - d] + \alpha_s(1-k)d > 0$. *($E^*$ may be stable whenever exists, can be further studied in future.)*

Hyman and Li (2000, 2005a,b, 2006) have studied epidemic models with differential infectivity and differential susceptibility in several papers. In particular, they studied a general compartmental differential susceptibility SIR model with disease-induced mortality in Hyman and Li (2005a). So here to get some results on the

positive steady state(s), the same technique is applied as in section 3.2 of Hyman and Li (2005a) to the model with intervention and constant entrance rate. See proof for Proposition 3.2.1.1 in Appendix C.

3.2.2 Intervention Models with Linear Recruitment

In this section, it is assumed that $f(N) = rN$ in model (3.3).

Notice that

$$\lim_{(S^p,S,I)\to(0,0,0)} \frac{S^p I}{N} = \lim_{(S^p,S,I)\to(0,0,0)} \frac{S^p I}{S^p + S + I} = 0$$

and

$$\lim_{(S^p,S,I)\to(0,0,0)} \frac{SI}{N} = \lim_{(S^p,S,I)\to(0,0,0)} \frac{SI}{S^p + S + I} = 0.$$

It is defined that $P(0,0,0) = Q(0,0,0) = R(0,0,0) = 0$. Under this assumption, $E = (0,0,0)$ is a unique steady state of (3.3), and $P(S^p,S,I), Q(S^p,S,I)$ and $R(S^p,S,I)$ are continuous in $\{(S^p,S,I)|S^p \geq 0, S \geq 0, I \geq 0\}$.

Proposition 3.2.2.1. *All solutions of (3.3) with nonnegative initial conditions remain nonnegative.*

Periodic solutions for (3.3) do not exist when $\beta - d > 0$.

The unique steady state ($E = (0,0,0)$) implies population extinction. It is globally stable if $r < \mu$.

Assume $r > \mu$ and $\beta > d$,

- *when $(1-k)\beta + k(1-\alpha_s)\beta < d + r$, $E_1 = (k,0)$ is globally stable (see Proposition 3.2.2.2), with $\lim_{t\to\infty} N(t) = \infty$ and*
  - *$\lim_{t\to\infty} I(t) = 0$ if $R_0 < 1$,*
  - *$\lim_{t\to\infty} I(t) = \infty$ if $R_0 > 1$;*

- *when* $(1-k)\beta+k(1-\alpha_s)\beta > d+r$, $E_2 = (p^*, i^*)$ *is globally stable (see Proposition 3.2.2.2),*
  - $\lim_{t\to\infty} N(t) = 0$ *if* $i^* > \frac{r-\mu}{d}$, *and* $\lim_{t\to\infty} I(t) = 0$,
  - $\lim_{t\to\infty} N(t) = \infty$ *if* $i^* < \frac{r-\mu}{d}$, *and* $\lim_{t\to\infty} I(t) = \infty$.

See proof in Appendix C.

Similar to the scenarios without intervention, the dynamics of the fractional form of (3.1) will be studied in terms of $p = \frac{S^p}{N}$, $s = \frac{S}{N}$, then $i = \frac{I}{N} = 1 - p - s$:

$$\begin{aligned}
\frac{dp}{dt} &= kr - rp - [(1-\alpha_s)\beta - d]pi \triangleq X(p,i) && (3.4)\\
\frac{di}{dt} &= [\beta - (d+r) - \alpha_s\beta p - (\beta - d)i]i \triangleq Y(p,i) && (3.5)\\
\frac{dN}{dt} &= [r - \mu - di]N && (3.6)\\
\frac{dI}{dt} &= [\beta(1-p-i) + (1-\alpha_s)\beta p - (\mu+d)]I. && (3.7)
\end{aligned}$$

which allows to understand the expected changes in HIV prevalence due to PrEP use.

Notice that the first two equations are decoupled from the rest of the system which allows to study the reduced system (3.4) and (3.5).

Proposition 3.2.2.2. *The biologically relevant region* $\{(p(t), i(t)) | p(t) \geq 0, i(t) \geq 0, p(t) + i(t) \leq 1\}$ *is positively invariant with respect to (3.4) and (3.5).*

When $\beta - d > 0$, *periodic solutions do not exist for the transformed system.*

$E_1 = (k, 0)$ *for the reduced system is stable when* $(1-k)\beta+k(1-\alpha_s)\beta < d+r$, *and is globally stable if further assume* $\beta > d$. *A positive steady state* $E_2 = (p^*, i^*)$ *of the reduced system exists and is globally stable provided that* $(1-k)\beta+k(1-\alpha_s)\beta > d+r$.

See proof in Appendix C.

A summary on the global stability results confirmed by simulations is provided in Table 3.3.

3.2.3 Intervention Models with Logistic Recruitment

In this section, it is assumed that $f(N) = rN(1 - \frac{N}{K})$ in model (3.3).

Notice that

$$\lim_{(S^p,S,I)\to(0,0,0)} \frac{S^pI}{N} = \lim_{(S^p,S,I)\to(0,0,0)} \frac{S^pI}{S^p + S + I} = 0$$

and

$$\lim_{(S^p,S,I)\to(0,0,0)} \frac{SI}{N} = \lim_{(S^p,S,I)\to(0,0,0)} \frac{SI}{S^p + S + I} = 0.$$

It is defined that $P(0,0,0) = Q(0,0,0) = R(0,0,0) = 0$. Under this assumption, $E_{00} = (0,0,0)$ is a steady state of (3.3), and $P(S^p,S,I), Q(S^p,S,I)$ and $R(S^p,S,I)$ are continuous in $\{(S^p,S,I)|S^p \geq 0, S \geq 0, I \geq 0\}$.

Proposition 3.2.3.1. *The biologically relevant region* $\{(S^p(t),S(t),I(t))|S^p(t) \geq 0, S(t) \geq 0, I(t) \geq 0, S^p(t) + S(t) + I(t) \leq K\}$ *is positively invariant with respect to (3.3).*

The model has three possible steady states: population extinction $E_{00} = (0,0,0)$, *disease-free* $E_{01} = (k\frac{r-\mu}{r}K, (1-k)\frac{r-\mu}{r}K, 0)$ *and endemic* $E^* = (S^{p*}, S^*, I^*)$.

The extinction steady state is globally stable if $r < \mu$.

Assume $r > \mu$,

- *the infection free steady state* E_{01} *is stable when* $R_0 < 1$ *and unstable when* $R_0 > 1$;
- *the extinction steady state* E_{00} *is stable when* $(1-k)\beta + k(1-\alpha_s)\beta > d + r$ *and unstable when* $(1-k)\beta + k(1-\alpha_s)\beta < d + r$;
- *further assume* $\beta > \mu + d$, *then the positive steady state* E^* *exists when* $R_0 > 1$ *and* $i^* = \frac{I^*}{N^*} < \frac{r-\mu}{d}$ *and does not exist when* $R_0 < 1$ *or* $i^* = \frac{I^*}{N^*} > \frac{r-\mu}{d}$.

See proof in Appendix C.

The behavior of the model (3.3) is further studied by investigating its fractional form in terms of $p = \frac{S^p}{N}$, $s = \frac{S}{N}$, and $i = \frac{I}{N} = 1 - p - s$:

$$\begin{aligned} \frac{dp}{dt} &= kr(1 - \tfrac{N}{K}) - r(1 - \tfrac{N}{K})p - [(1 - \alpha_s)\beta - d]pi \\ \frac{di}{dt} &= [\beta - d - r(1 - \tfrac{N}{K}) - \alpha_s\beta p - (\beta - d)i]i \\ \frac{dN}{dt} &= [r(1 - \tfrac{N}{K}) - \mu - di]N \end{aligned} \tag{3.8}$$

Proposition 3.2.3.2. *Assume $r > \mu$.*

Equilibrium $E_1 = (k, 0, 0)$ for (3.8) is unstable.

Equilibrium $E_2 = (p^*, i^*, 0)$ ($p^* > 0, i^* > 0$) for (3.8) does not exist when $(1 - k)\beta + k(1 - \alpha_s)\beta < d + r$ and is stable otherwise.

Equilibrium $E_3 = (k, 0, \frac{r-\mu}{K})$ (infection free steady state) for (3.8) is stable when $R_0 < 1$ and unstable otherwise.

Further assume $\beta > \mu + d$, then a positive steady state $E_4 = (p^*, i^*, N^*)$ for (3.8) exits when $R_0 > 1$ and $i^* < \frac{r-\mu}{d}$ and does not exist otherwise.

See proof in Appendix C.

A summary of the asymptotic behavior of the system with PrEP confirmed by simulations is provided in Table 3.3.

Next, a threshold value $\tilde{\beta}$ (used in Table 3.3) will be determined for β. $\tilde{\beta}$ will be derived from substituting $(p^*_{\log}, i^*_{\log} = \frac{r-\mu}{d}, 0)$ into (3.8). First

$$\frac{r - \mu}{d} = i^* = \frac{\tilde{\beta} - \mu - d - \alpha_s\tilde{\beta}p^*}{\tilde{\beta}}$$

leads to $p^* = \frac{(1-i^*)\tilde{\beta} - \mu - d}{\alpha_s\tilde{\beta}}$. Then from (3.8),

$$\frac{dp^*}{dt} = 0$$

implies

$$kr(1 - \frac{N^*}{K}) - r(1 - \frac{N^*}{K})p^* - [(1 - \alpha_s)\tilde{\beta} - d]p^*i^* = 0$$

Table 3.3: Models with intervention, assume $r > \mu$ and $\beta > d$. Notice that $\tilde{\beta}$ is a solution to (3.9). i^*_{lin} is the i^* appears in Proposition 3.2.2.1 and 3.2.2.2. $i^*_{\log}$ is the i^* appears in Proposition 3.2.3.1 and 3.2.3.2.

Recruitment type	Parameter conditions	Outcomes	HIV prevalence
Constant	$\beta < \frac{\mu+d}{1-\alpha_s k}$	disease free state	0
	$\beta > \frac{\mu+d}{1-\alpha_s k}$	endemic state	$\frac{\mu I^*}{\Lambda - dI^*}$
Linear	$\beta < \frac{\mu+d}{1-\alpha_s k}$	disease free state	0
	$\frac{\mu+d}{1-\alpha_s k} < \beta < \max\{\frac{r+d}{1-\alpha_s k}, \tilde{\beta}\}$	endemic state	0 or $i^*_{\text{lin}} (< \frac{r-\mu}{d})$
	$\beta > \tilde{\beta}$	extinction	$i^*_{\text{lin}} (> \frac{r-\mu}{d})$
Logistic	$\beta < \frac{\mu+d}{1-\alpha_s k}$	disease free state	0
	$\frac{\mu+d}{1-\alpha_s k} < \beta < \max\{\frac{r+d}{1-\alpha_s k}, \tilde{\beta}\}$	endemic state	$i^*_{\log} (< \frac{r-\mu}{d})$
	$\beta > \tilde{\beta}$	extinction	$i^*_{\log} (> \frac{r-\mu}{d})$

, i.e.,

$$kr - rp^* - [(1-\alpha_s)\tilde{\beta} - d]p^* i^* = 0$$

with $p^* = \frac{(1-i^*)\tilde{\beta}-\mu-d}{\alpha_s \tilde{\beta}}$ and $i^* = \frac{r-\mu}{d}$. This eventually can be simplified to be

$$kr\alpha_s\tilde{\beta} - [(1-i^*)\tilde{\beta} - \mu - d][i^*(1-\alpha_s)\tilde{\beta} + \mu] = 0$$

where $i^* = \frac{r-\mu}{d}$ or

$$kr\alpha_s\tilde{\beta} - [(1 - \frac{r-\mu}{d})\tilde{\beta} - \mu - d][\frac{r-\mu}{d}(1-\alpha_s)\tilde{\beta} + \mu] = 0. \tag{3.9}$$

Some bifurcation results are presented in Figure 3.1. Note that the trajectories of the models with constant and logistic recruitment are bounded while the linear recruitment allows for unbounded solutions. As the HIV risk (β) increases, the model with constant recruitment switches from infection-free to endemic steady state while

the model with logistic recruitment equilibrium goes from infection-free to endemic equilibrium and further to population extinction. The trajectories of the linear model follow a similar pattern as the logistic recruitment model: from unbounded infection-free through unbounded endemic state to population extinction. However, the transition (bifurcation) points where the behavior changes occur are different for the three models. As a result the projected long-term prevalence with each of the three models differ for epidemic conditions with R_0 greater than one.

What in common for all the three models is when $R_0 < 1$, the status of the population approaches infection free; when $R_0 > 1$, the status of the population approaches coexistence or extinction.

These figures in Figure 3.2 illustrate the possible solutions that can not be expressed explicitly in the previous analysis work. Figure 3.2 and 3.1 both confirm the propositions that haven been proved and the summary proposed in Table 3.3.

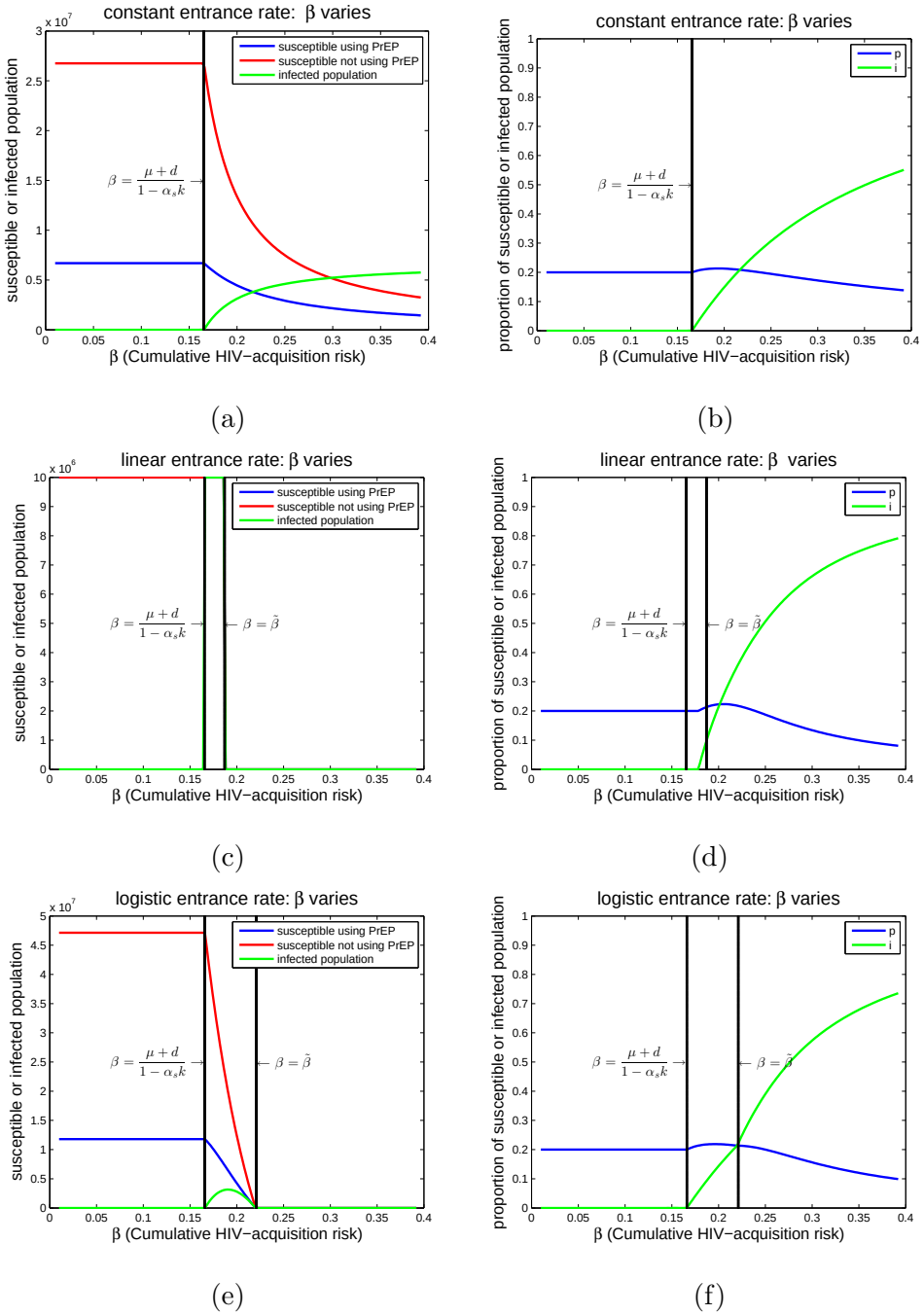


Figure 3.1: Bifurcation diagrams of models employing different recruitment functions: (a-b) constant; (c-d) linear; (e-f) logistic. Epidemic parameter values (except β) are chosen by fitting projected HIV populations to data from South Africa (see Table D.1). PrEP coverage ($k = 0.2$) and PrEP efficacy ($\alpha_s = 0.5$) are assumed. Initial conditions are adapted from year 2011.

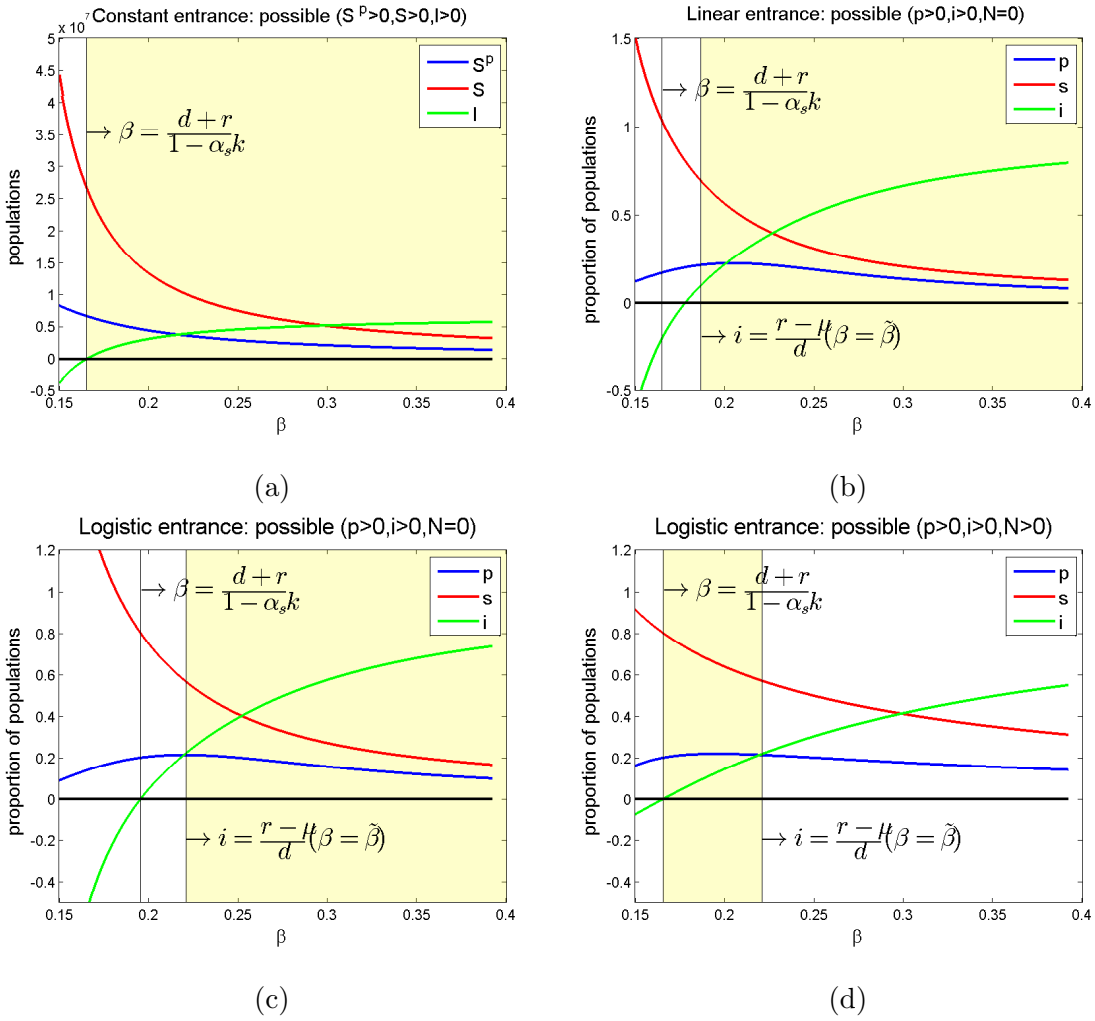


Figure 3.2: Parameter values(except β, see Table D.1) are chosen from data fitting, with $k = 0.2$ and $\alpha_s = 0.5$ are assumed. Initial conditions are adapted from year 2011. Figures are truncated for better view. Figure (a) corresponds to model (3.3) with constant entrance. Figure (b) corresponds to the transformed system (3.4) with linear entrance. Figures (c) and (d) both correspond to transformed system (3.8) with logistic entrance. (c) highlights possible steady states ($p^* > 0, i^* > 0, N^* = 0$), and (d) highlights possible steady state ($p^* > 0, i^* > 0, N^* > 0$).

3.3 Public-health Impact of PrEP Use

3.3.1 Parameterizations and Simulations

The models are simulated using different recruitment mechanisms (different $f(N)$), and keeping all the remaining parameters the same. It is assumed that the annual influx of people in the population is 1,000,000 initially, and then the parameter values are calculated corresponding to such recruitment for each mechanism. Therefore, $\Lambda = 10^6$ is used for the model with constant recruitment, $r = \frac{10^6}{N_0}$ for the model with linear recruitment and $r = \frac{10^6 K}{N_0(K-N_0)}$ with $K = 9 \times 10^7$ (Figure A.1) for the model with logistic recruitment, where $N_0 = 27172431$ is the initial population size from Table A.1.

The resulting population dynamics over 70 years under scenarios with and without PrEP are presented in Figure 3.3. Note that with identical initial recruitment and using the same values for all other parameters the population dynamics substantially diverge over the simulated period. In absence of PrEP, the population suffers the smallest decrease in size (34%) under the constant recruitment scenario because the disease-related mortality does not impact the influx of newly susceptible people (Figure 3.3 (a),(b),(c)). In comparison, the population loses 67% and almost 83% over 70 years under the logistic and linear recruitment scenarios. In addition to the population size the proportion of infected individuals is affected as well. Starting at 16.6% the models predict that the HIV prevalence will raise to 25.4% with constant, 32.9% with logistic and 41.9% with linear recruitment (Figure 3.3 (d)). The results do not change qualitatively if 20% of the population use PrEP. Naturally, for all recruitment methods the projected number of susceptibles is larger compared to the scenario without PrEP. However, the model with constant recruitment also projects smaller number of infected vs the scenario without PrEP while the model

with linear recruitment shows substantial increase in infected individuals due to the fact that healthier population size is preserved when PrEP is used. The relative order of projected population size by recruitment mechanism remains the same (Figure 3.3 (e),(f),(g)). The model with linear recruitment is most pessimistic with respect to HIV prevalence (18.8%) while the other two mechanisms predict decrease of HIV prevalence to 15.9% and 14.4%.

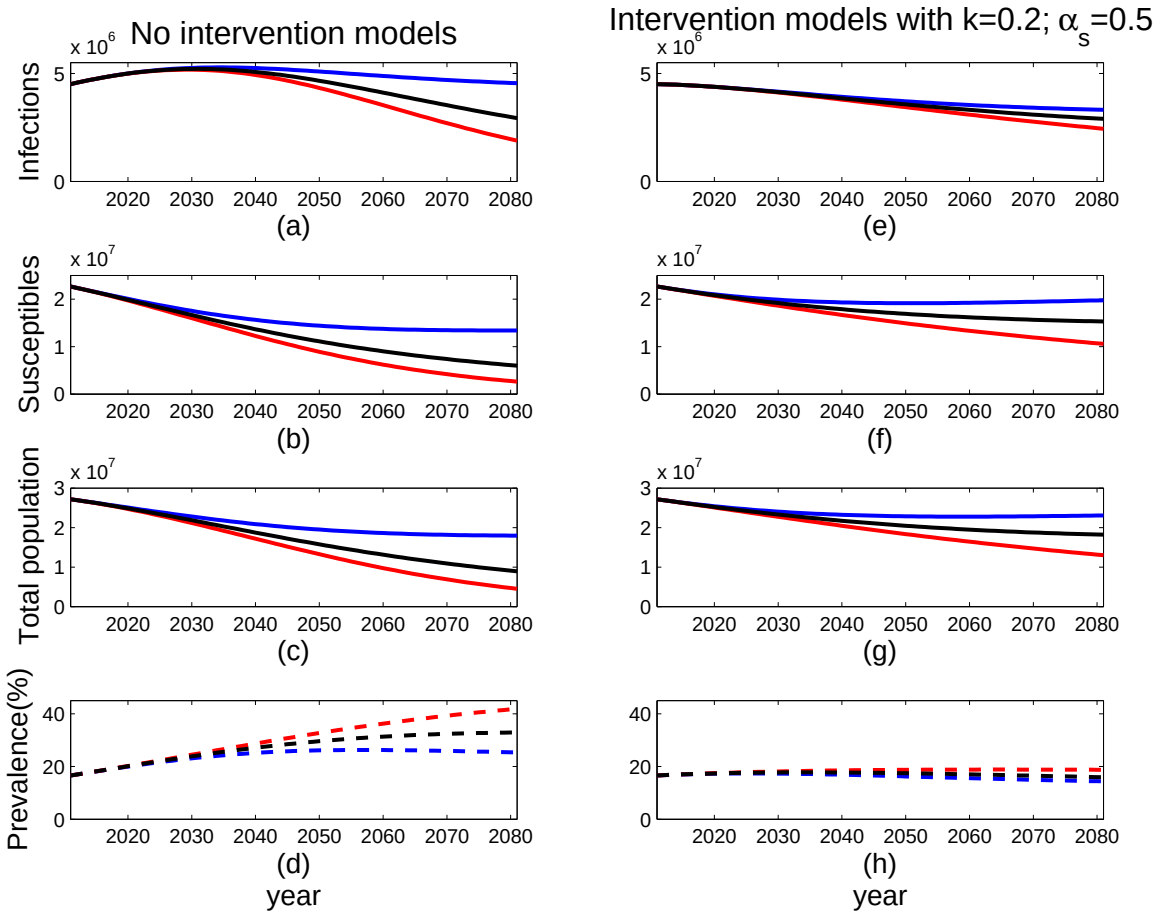


Figure 3.3: Population dynamics for models with different recruitment rates: constant (blue), linear (red), and logistic (black). Initial recruitment and parameter values unrelated to recruitment are kept the same across the models (see Table D.1).

3.3.2 *Influence of Recruitment on the PrEP Effectiveness*

The effectiveness of PrEP use is often evaluated by different quantitative indicators. It has been demonstrated in previous work that the choice of evaluation method may influence the conclusions of the modeling analyses (Zhao *et al.* (2013)). Here, study will focus on four indicators already used in modeling studies to quantify the impact of PrEP interventions (see Table 3.4). The fractional indicator (F_I) measures the intervention effectiveness based on the difference in expected infections between scenarios with and without PrEP. The prevalence (P_I) and incidence (aI_I) indicators measure the reduction in the projected HIV prevalence and incidence due to PrEP, respectively. The last evaluation method ($\hat{F}_I$) is based on the reduction of the number of infected individuals and correlates with the economic burden of the HIV epidemic on the public health system at community and state level since the money allocated for HIV treatment is proportional to the size of the infected population.

Table 3.4: Indicator description

Indicator	Name	Description
$F_I(T)$	Fractional indicator	Fraction of infections prevented over the period $[0, T]$ due to the usage of PrEP
$P_I(T)$	Prevalence indicator	Reduction in HIV-prevalence at time $t = T$ due to the usage of PrEP
$aI_I(T)$	Incidence indicator	Reduction in the annual HIV incidence at time $t = T$ due to the usage of PrEP
$\hat{F}_I(T)$	Fractional reduction indicator	Fraction of the projected number of infected at time $t = T$ reduced due to the usage of PrEP

To track the cumulative number of new infections over time, the following equation is added to model (3.1):

$$\frac{d(I_{New})}{dt} = \beta \frac{SI}{N},$$

and similarly to model (3.3):

$$\frac{d(I_{New})}{dt} = \beta \frac{SI}{N} + (1-\alpha_s)\beta \frac{S^p I}{N}.$$

[ ] is used to denote variables from the model without PrEP (3.1) and [ ]$_P$ for variables from the model with PrEP (3.3). Using these notations, the qualitative indicators are defined as follows:

$$\begin{aligned}
F_I(T) &= \frac{[I_{New}(T)]-[I^p_{New}(T)+I_{New}(T)]_P}{[I_{New}(T)]} = 1 - \frac{[I^p_{New}(T)+I_{New}(T)]_P}{[I_{New}(T)]} \\
P_I(T) &= 1 - \frac{[\frac{I^p(T)+I(T)}{S^p(T)+S(T)+I^p(T)+I(T)}]_P}{[\frac{I(T)}{S(T)+I(T)}]} \\
aI_I(T) &= 1 - \frac{[\frac{I^p_{New}(T+1)+I_{New}(T+1)-(I^p_{New}(T)+I_{New}(T))}{S^p(T)+S(T)}]_P}{[\frac{I_{New}(T+1)-I_{New}(T)}{S(T)}]} \\
\hat{F}_I(T) &= \frac{[I(T)]-[I^p(T)+I(T)]_P}{[I(T)]} = 1 - \frac{[I^p(T)+I(T)]_P}{[I(T)]}.
\end{aligned}$$

The impact of the recruitment on the projected PrEP effectiveness is investigated in Figure 3.4. The choice of recruitment mechanisms shows no substantial impact over the initial period of 20-30 years but leads up to 21% difference in predicted reduction in HIV prevalence and incidence after 70 years (Figure 3.4 (a),(b)). The model using linear recruitment is most optimistic predicting 55% reduction in HIV prevalence and 61% in HIV incidence, respectively. Conversely, the model with constant recruitment projects largest fraction of infection prevented (Figure 3.4 (c)). Interestingly, the same indicator projects negative overall PrEP impact of the models with linear and logistic recruitment in a long term. It is a result of the critical decline in population size under the scenario without PrEP which limits the number of HIV infections in a long term.

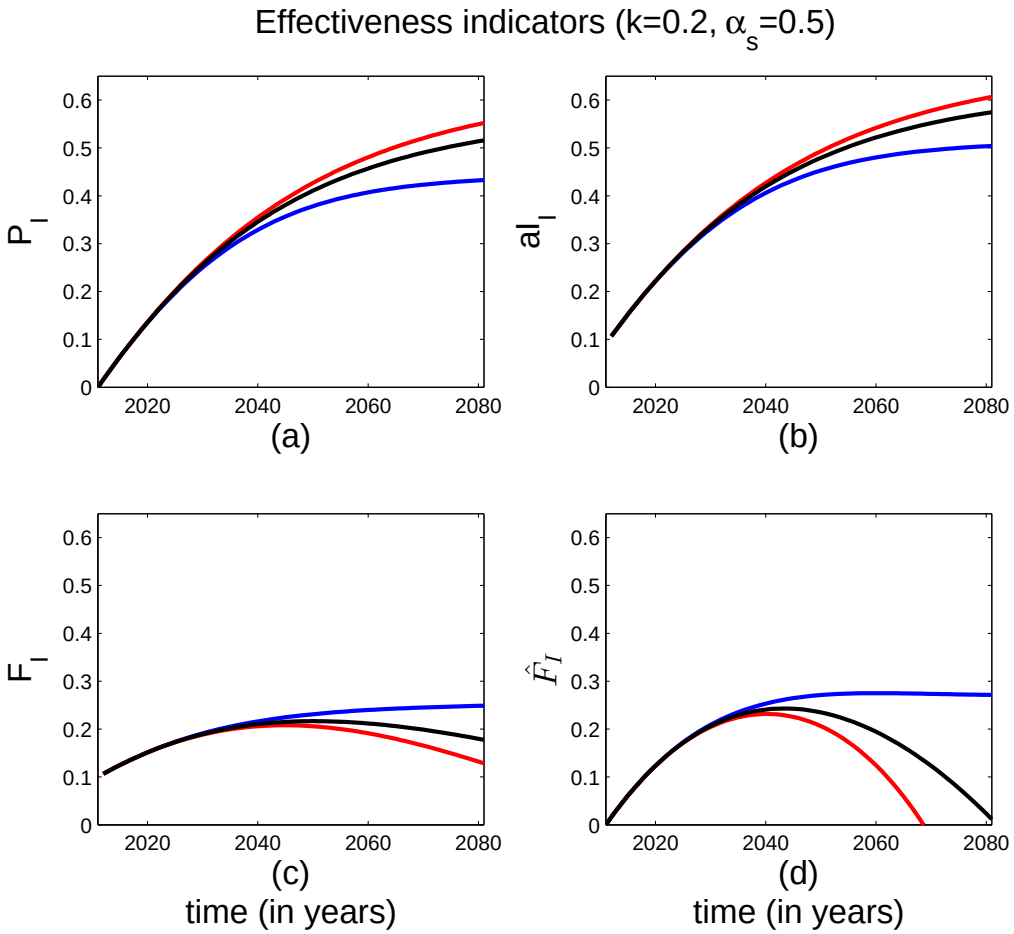


Figure 3.4: Dynamics of PrEP effectiveness for models with different recruitment rates: constant (blue), linear (red), and logistic (black). Initial recruitment and parameter values unrelated to recruitment are kept the same across the models (see Table D.1). Some indicators take negative values for models with linear entrance rate and logistic entrance rate (not shown).

Finally, simulation of the HIV epidemics has been done by fitting the models with different recruitment to 10-year HIV data representative for South Africa. Parameters values (see Table D.1) were determined to minimize the L^1 norm of the difference between projected HIV population and data in absence of PrEP following the approach

proposed in a previous study (Zhao *et al.* (2013)). The relative short duration of the fitted period did not allow for significant difference in the "best fit" parameters across models. As a result the predictions of HIV dynamics and PrEP effectiveness with that "best fit" parameter sets (see Figure D.1 and Figure D.2 in the Appendix), were qualitatively similar to the simulations with fixed parameter sets (Figure 3.3 and 3.4).

3.4 Discussion

Mathematical models are employed to estimate the expected effectiveness of different interventions for HIV prevention under various epidemic settings. In this paper it has been demonstrated that the assumptions regarding population recruitment have a strong influence on the future course of the HIV epidemic and as a result impacts the projected success of the planned interventions. Models are compared equipped with three distinct recruitment mechanisms (constant, linear and logistic) and studied their behavior. The analysis shows that the three models posses qualitatively different dynamic characteristics. The model with constant recruitment always stabilizes in population size and supports two asymptotic states, corresponding to disease-free and endemic equilibrium respectively. In comparison, linear and logistic recruitment support three asymptotic states including disease free equilibrium, endemic equilibrium and population extinction under the pressure of HIV. The parameter conditions (bifurcation points) where the transitions between asymptotic states occur are the same for these two models but different from those for the model with constant recruitment. On the other hand, constant and logistic model share an endemic fractional equilibrium which is different from the linear model, i.e., they project different HIV prevalence over long-term.

As a result the simulations of the HIV epidemics with the three models over 70

years show large discrepancies in population size and epidemic distribution under identical initial conditions and forces of infection. The projected HIV prevalence varies from 25% when constant recruitment is assumed and 42% when linear recruitment is assumed. In addition, significant difference in the reduction in HIV prevalence and incidence (almost 20%) is predicted when 50% effective PrEP is used by 20% of the population. Over the entire simulated period, linear recruitment provides the most optimistic estimates of the PrEP effectiveness in terms of prevalence reduction while constant recruitment predicts larger fraction of infections prevented.

It can be argued that regardless of the differences in the dynamic behavior all three models agree in their effectiveness projections over 20-30 years which is the usual period over which the intervention are evaluated. However, often the models are run for extended periods in order to simulate "mature" epidemics and the intervention is introduced afterwards. The key message of this analysis is that the way recruitment is incorporated in the models impacts the HIV epidemic and may have significant effect on the projected effectiveness of different HIV intervention in a short and long term. Demographic data, including statistics on births and age-specific mortality, should be used to inform the modeling mechanisms before HIV prevention is considered.

Chapter 4

BAYESIAN PARAMETER ESTIMATION

Previously, I tried to fit models to the data by minimizing the L^1 norm of the difference between projected HIV population and data (Figure D.1). This approach starts with a single set of initial guess values for the parameters, and ends up with another single set of approximated parameter values that make the model fit the data best. It works by finding the set of parameter values that minimize the sum of the absolute differences between the data and the model predictions. The issue is that this method only provides point estimates and does not tell how strongly the data supports these particular estimates. Distributions provide much more information than the point estimates. In reality, information on distributions of certain parameters can often be found in literature (see Table 4.1).

The Bayesian estimation of an unknown parameter Θ using some newly acquired data can be worked out based on Bayes' formula. First what is needed is the likelihood function $p(D|\Theta = \theta)$, which specifies the conditional distribution of the data (D) on each possible set of parameter values ($\Theta = \theta$). Then I choose some prior distribution $p(\theta)$ which indicates how strongly $\Theta = \theta$ is believed before accessing the new data. Finally using the information provided by the new data, Bayes' formula $p(\theta|D) = p(\theta)\dfrac{p(D|\theta)}{p(D)}$ provides the posterior distribution given the new data. The belief in how strong $\Theta = \theta$ now has been modified from $p(\theta)$ to $p(\theta|D)$.

Therefore, in this chapter, a second approach I will try is the Bayesian parameter estimation on fitting ODE model to the related HIV data. Now one can start with a set of prior distributions for the parameters as initial guess, and try to obtain a set of posterior distributions for the parameters which make the model fit the data

best. Prior distributions are distributions of the parameters assigned before data is observed. Prior distributions can either be derived from existing literature or be assumed. Posterior distributions are obtained by Bayesian inference given the observed data.

ODEs are not based on a probability model, to apply Bayesian inference directly on ODEs, likelihoods are therefore generally defined in a nonlinear regression context, such as assuming that the data are normally distributed around the deterministic solution(Toni *et al.* (2009); Vyshemirsky and Girolami (2008)).

In this chapter, I will try Bayesian method on estimating the parameters for model (4.1) from data on South Africa HIV population of year 2001-2011 (Table A.1).

4.1 Model

$$\begin{aligned} \frac{dS}{dt} &= \Lambda - \beta\frac{SI}{N} - \mu S \\ \frac{dI}{dt} &= \beta\frac{SI}{N} - (\mu + d)I \end{aligned} \tag{4.1}$$

Model (4.1) is the HIV model without intervention. See Table 4.1 for more information on the parameters. Throughout this chapter, Λ derived from Table A.1 will be assumed to be constant entrance rate. β, calculated from b_a and n ($\beta = 1-(1-b_a)^n$), will be studied directly. I will start with prior distributions on μ, d, and β, then apply Bayesian method to get posterior distributions for these parameters.

4.2 Parameter Prior Distributions

Table 4.1 has showed information on mean and possible confidence intervals from literature for parameters b_a, n, μ, and d. Then I assume lognormal distributions for b_a, μ, d, and Poisson distribution (Boily *et al.* (2009)) for n. The lognormal distributions were chosen for b_a, μ, and d since these parameters must be non-negative. To fit

Table 4.1: Parameter description and values from literature

Par.	Description	Mean	95% CI	Reference
Λ	Annual rate at which individuals become sexually active	996344		derived from Table A.1
b_a	HIV acquisition risk per act	0.0038	0.0013-0.011	Boily *et al.* (2009)
n	Number of sexual acts per year per individual	120		Wawer *et al.* (2005), Boily *et al.* (2009)
β	Cumulative HIV-acquisition risk			derived from b_a and n
$\frac{1}{\mu}$	Time (in years) to remain sexually active	35		UNAIDS (2009)
d	HIV carrier's annual rate of progression to AIDS	1/10.9	1/11.3-1/10.6	Porter and Zaba (2004)

the confidence intervals best (see (b) and (c) in Figure 4.1), I found the log standard deviations for b_a and d as in Table 4.2. For μ, log standard deviation is assumed to be 0.15.

Recall that when I minimized the L^1 norm of the difference between HIV populations from ODE model and from data (Figure D.1), I obtained $\mu = 0.029793556$, $d = 0.119146121$, and $\beta = 0.196924711$. Notice that the value for d is not consistent with the prior distribution as in Figure 4.1(b). This reminds me to revisit the literature from which I obtained the prior distribution for d. 10.9 years was observed in industrialized countries, as mean survival time for those aged 24-35 years at seroconversion, with 95% CI 10.6-11.3 years (Babiker *et al.* (2000); Porter and Zaba (2004)).

Here, I expect the mean time to be shorter to develop AIDS in South Africa compared with industrialized countries. Also notice that age group 15-49 years is studied instead of 24-35 years, so a wider 95% confidence interval should be expected. Therefore I will increase the corresponding standard deviation.

To make a better prior for d, I modify the prior distribution for d by increasing σ_2 from 0.0161875 to 0.25. Then the prior distribution for d is modified as in Figure 4.1(f).

Table 4.2: Assumptions for parameters. $\hat{d}$ is modified d. $\beta = 1 - (1 - b_a)^n$.

Par.	Distribution	Log mean	Log standard deviation	Mean	Standard deviation	95% CI
Λ	constant					
b_a	lognormal	ln(0.0038)	0.542394	0.004402	0.0025746	0.0013126 -0.011001
n	Poisson		120			
β	$\beta = \beta(b_a, n)$					
μ	lognormal	ln(1/35)	0.15(assumed)	0.028895	0.0043587	
d	lognormal	ln(1/10.9)	0.0161875	0.091755	0.0014854	0.088888 -0.09469
$\hat{d}$	lognormal	ln(1/10.9)	0.25	0.094655	0.024038	

Now the probability density functions (PDF) for the prior distributions have been derived as following.

μ has lognormal PDF $p(\mu) = \dfrac{1}{\mu\sigma_1\sqrt{2\pi}} e^{-\frac{(\ln(\mu)-\mu_1)^2}{2\sigma_1^2}}$ with $\mu_1 = \ln(1/35)$ and $\sigma_1 = 0.15$.

d has the lognormal PDF $p(d) = \dfrac{1}{d\sigma_2\sqrt{2\pi}} e^{-\frac{(\ln(d)-\mu_2)^2}{2\sigma_2^2}}$ with $\mu_2 = \ln(1/10.9)$ and $\sigma_2 = 0.25$.

b_a has the lognormal PDF $p(b_a) = \dfrac{1}{b_a\sigma_4\sqrt{2\pi}} e^{-\frac{(\ln(b_a)-\mu_4)^2}{2\sigma_4^2}}$ with $\mu_4 = \ln(0.0038)$ and $\sigma_4 = 0.542394$.

n has the Poisson PDF $p(n) = \dfrac{\lambda^n}{n!}e^{-n}$ with $\lambda = 120$. For convenience, the Poisson distribution for n is approximated by a normal distribution with mean 120 and standard deviation $\sqrt{120}$. So it is approximated by the normal PDF $p(n) = \dfrac{1}{\sqrt{2\pi}\sigma_5} e^{-\frac{(n-\mu_5)^2}{2\sigma_5^2}}$, with $\mu_5 = 120$ and $\sigma_5 = \sqrt{120}$ (see Figure 4.1(d)).

Since b_a and n are independent, then the cumulative distributive function (CDF) for β becomes

$$F(\beta) = \int_0^1 \int_0^{\ln_{1-b_a}(1-\beta)} p(b_a)p(n)\,\mathrm{d}n\mathrm{d}b_a,$$

and the PDF for β is

$$p(\beta) = \frac{d}{d\beta}F(\beta) = \frac{d}{d\beta}\left[\int_0^1 \int_0^{\ln_{1-b_a}(1-\beta)} p(b_a)p(n)\,\mathrm{d}n\mathrm{d}b_a\right].$$

Then by Leibniz rule,

$$\begin{aligned} p(\beta) &= \int_0^1 p(b_a)p(n(b_a,\beta))\frac{-1}{(1-\beta)\ln(1-b_a)}\,\mathrm{d}b_a \\ &= \int_0^1 \frac{1}{b_a\sigma_4\sqrt{2\pi}} e^{-\frac{(\ln(b_a)-\mu_4)^2}{2{\sigma_4}^2}} \frac{1}{\sqrt{2\pi}\sigma_5} e^{-\frac{(\log_{1-b_a}(1-\beta)-\mu_5)^2}{2{\sigma_5}^2}} \frac{-1}{(1-\beta)\ln(1-b_a)}\,\mathrm{d}b_a, \end{aligned} \tag{4.2}$$

with $\mu_4 = \ln(0.0038)$, $\sigma_4 = 0.542394$, $\mu_5 = 120$, and $\sigma_5 = \sqrt{120}$.

This can be approximated by a lognormal distribution of PDF $p(\beta) \approx \dfrac{1}{\beta\sigma_3\sqrt{2\pi}} \cdot e^{-\frac{(\ln(\beta)-\mu_3)^2}{2\sigma_3^2}}$ with $\mu_3 = \ln(0.3704)$ and $\sigma_3 = 0.4422$ (see Figure 4.1(e)).

Then the prior distributions for μ, d, and β are illustrated in Figure 4.1 and summarized in Table 4.3.

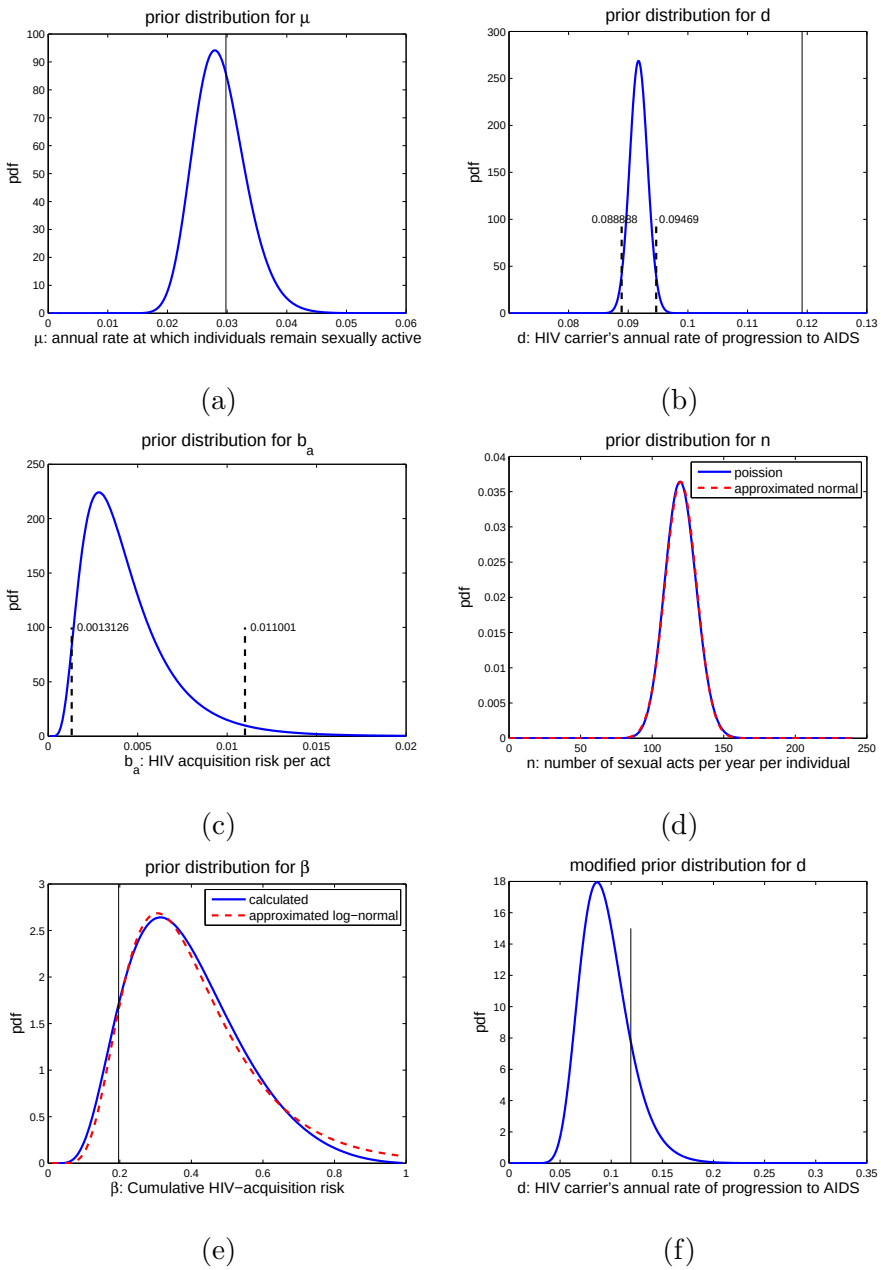


Figure 4.1: Parameter prior distributions. 95% confidence intervals(dashed black) are also included for b_a and d. Poisson distribution and approximated normal distribution have both been included for n. Calculated distribution (4.2) and approximated log-normal distribution have both been included for β. Solid black lines indicate values obtained from data fitting as in Figure D.1. (f) is a modification of (b) with a larger standard deviation.

Table 4.3: Parameter prior distributions

Par.	Distribution	Log mean	Log standard deviation	Mean	Standard deviation
μ	lognormal	$\mu_1 = \ln(1/35)$	$\sigma_1 = 0.15$	0.028895	0.0043587
d	lognormal	$\mu_2 = \ln(1/10.9)$	$\sigma_2 = 0.25$	0.094655	0.024038
β	lognormal	$\mu_3 = \ln(0.3704)$	$\sigma_3 = 0.4422$	0.40844	0.18981

Let z_i represent the i^{th} year HIV infected population since 2001 from data ($z_{0:T}$), and x_i represent the i^{th} year HIV infected population predicted by the model ($x_{0:T}$).

Because of the observation error (assumed to be normally distributed with standard deviation σ) in $z_{0:T}$, another two parameters are added into the study: the starting point of HIV population in the model x_0, and the related standard deviation σ. x_0 and σ are not independent since x_0 depends on both z_0 and σ.

Notice that only positive values are acceptable for x_0, so let $p(x_0|z_0, \sigma) = 0$ is assumed when $x_0 \leq 0$ and

$$p(x_0|z_0,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_0-z_0)^2}{2\sigma^2}} \Big/ \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-z_0)^2}{2\sigma^2}}\,\mathrm{d}x = e^{-\frac{(x_0-z_0)^2}{2\sigma^2}} \Big/ \int_0^\infty e^{-\frac{(x-z_0)^2}{2\sigma^2}}\,\mathrm{d}x \tag{4.3}$$

when $x_0 > 0$.

I did not find a prior distribution for σ from literature, therefore I choose a non-informative prior distribution which is invariant under reparameterization. Since the likelihood is normal with known mean and unknown standard deviation σ (see (4.4)), then I assume a Jeffreys prior for σ as

$$p(\sigma) \propto \frac{1}{\sigma} \text{ with } \sigma \in [1, 10^6]$$

or $p(\sigma) = \frac{C_\sigma}{\sigma}$ with $\sigma \in [1, 10^6]$ and $C_\sigma = \frac{1}{6\ln(10)}$. Since the observed HIV population as in Table A.1 shows $x_0 \in [3.8 \times 10^6, 4.6 \times 10^6]$, it is practical to assume that the possibility of σ lies outside $[1, 10^6]$ is negligible, with σ being the standard deviation when the observation is normally distributed.

Then prior for x_0 becomes

$$p(x_0|z_0) = \int_1^{10^6} p(\sigma)p(x_0|z_0,\sigma)\,\mathrm{d}\sigma = \int_1^{10^6} p(\sigma)\frac{1}{\int_0^\infty e^{-\frac{(x-z_0)^2}{2\sigma^2}}\,\mathrm{d}x}e^{-\frac{(x_0-z_0)^2}{2\sigma^2}}\,\mathrm{d}\sigma,$$

with $p(\sigma) = \frac{C_\sigma}{\sigma}$.

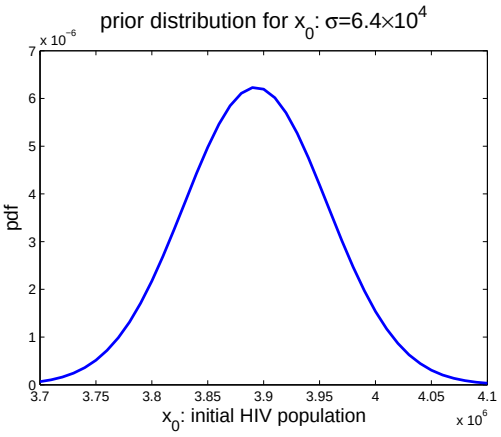
Then

$$p(x_0) \approx \int_1^{10^6} \frac{C_\sigma}{\sigma}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_0-z_0)^2}{2\sigma^2}}\,\mathrm{d}\sigma = \int_1^{10^6} \frac{C_\sigma}{\sqrt{2\pi}\sigma^2}e^{-\frac{(x_0-z_0)^2}{2\sigma^2}}\,\mathrm{d}\sigma$$
$$\xlongequal{v=\frac{x_0-z_0}{\sqrt{2}\sigma}} \frac{C_\sigma}{\sqrt{\pi}(x_0-z_0)}\int_{\frac{x_0-z_0}{\sqrt{2}}10^{-6}}^{\frac{x_0-z_0}{\sqrt{2}}} e^{-v^2}\,\mathrm{d}v(\text{when } x_0 \neq z_0),$$

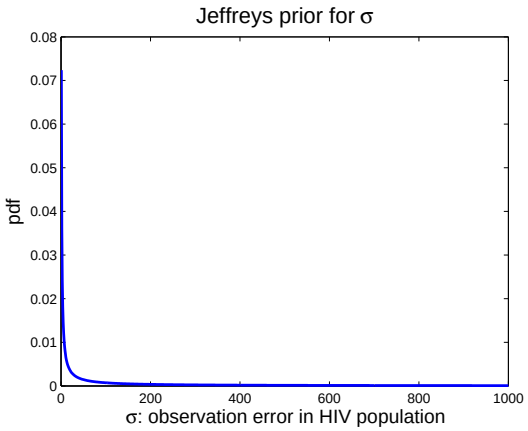
and

$$p(x_0 = z_0) \xlongequal{u=\frac{1}{\sigma}} \int_{10^{-6}}^1 \frac{C_\sigma}{\sqrt{2\pi}}\,\mathrm{d}u = \frac{C_\sigma(1-10^{-6})}{\sqrt{2\pi}}.$$

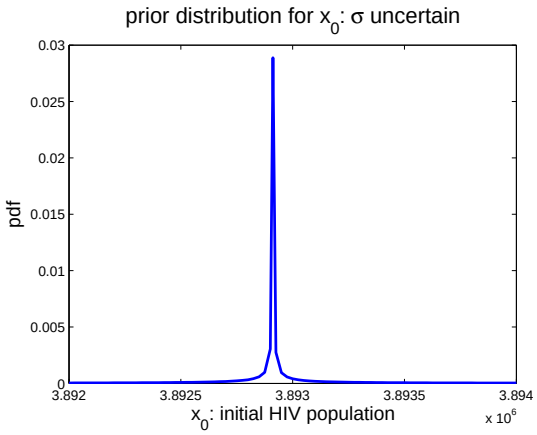
Thus, all the prior distributions needed for Bayesian inference have been derived (see Figure 4.1 and 4.2).



(a)



(b)



(c)

Figure 4.2: Figure (a) is the prior distribution for x_0 when $\sigma = 6.4 \times 10^4$ is fixed. Figure (c) is the prior distribution for x_0 when σ has a Jeffreys prior as in figure (b).

4.3 Bayesian Estimation

Now that parameter prior distributions have been derived, Bayesian method is ready to be applied.

Let $\theta = (\theta_1, \theta_2, \theta_3, \sigma, x_0) = (\mu, d, \beta, \sigma, x_0)$ represent 5 parameters to be studied and denote $D = z_{0:T}$.

Bayesian inference computes the posterior distributions according to the Bayes'

formula

$$p(\theta|D) = p(\theta)\frac{p(D|\theta)}{p(D)},$$

where

- $p(\theta)$ represents the prior density, the probability of θ before D is observed;
- $p(\theta|D)$ represents the posterior probability, the probability of θ after D is observed;
- $p(D|\theta)$ is the likelihood (or sampling probability for D), the probability of observing D given θ;
- $p(D)$ is the marginal likelihood, all possible hypotheses being considered:

$$p(D) = \int_\theta p(\theta)p(D|\theta)\,\mathrm{d}\theta.$$

Observation error is assumed to be normally distributed with standard deviation σ. Then the total likelihood function is given by

$$p(D|\theta) = \prod_{i=1}^{T} N(z_i|x_i;\theta) = \prod_{i=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(z_i-x_i)^2}{2\sigma^2}}, \tag{4.4}$$

where a Jeffreys prior distribution for σ is assumed and $x_i = x_i(\theta)$ for $i \geq 1$.

4.3.1 Analytic Solutions

From Bayes' formula $p(\theta|D) = p(\theta)\frac{p(D|\theta)}{p(D)}$, one can first determine the shape for $p(\theta|D)$ by calculating $p(\theta|D) \propto p(\theta)p(D|\theta) \triangleq g(\theta)$. Notice that $p(D)$ is a constant: $p(D) = \int_\theta p(\theta)p(D|\theta)\,\mathrm{d}\theta$.

4.3.1.1 σ uncertain

In this case, it is assumed that σ is a distribution with a Jeffreys prior.

The following parameters were assigned independent prior distributions: (μ, d, β, σ). But $x_0 = x_0(\sigma)$, and $x_i = x_i(\mu, d, \beta, \sigma)$ for $i \geq 1$. So using (4.3) with notation $\Delta \triangleq \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-z_0)^2}{2\sigma^2}} \, \mathrm{d}x$, the joint probability becomes

$$\begin{aligned} p(\theta) &= p(\mu)p(d)p(\beta)p(\sigma)p(x_0|z_0,\sigma) = p(\mu)p(d)p(\beta)p(\sigma)\frac{1}{\Delta}e^{-\frac{(x_0-z_0)^2}{2\sigma^2}} \\ &= \left(\prod_{j=1}^{3} \frac{1}{\sqrt{2\pi}\sigma_j\theta_j} e^{-\frac{(\ln\theta_j-\mu_j)^2}{2\sigma_j^2}}\right) \frac{C_\sigma}{\sigma}\frac{1}{\Delta}e^{-\frac{(x_0-z_0)^2}{2\sigma^2}}. \end{aligned} \tag{4.5}$$

Then using (4.5) and (4.4),

$$\begin{aligned} & g(\theta) \triangleq p(\theta)p(D|\theta) \\ &= \left(\prod_{j=1}^{3} \frac{1}{\sqrt{2\pi}\sigma_j\theta_j} e^{-\frac{(\ln\theta_j-\mu_j)^2}{2\sigma_j^2}}\right) \frac{C_\sigma}{\sigma}\frac{1}{\Delta}e^{-\frac{(x_0-z_0)^2}{2\sigma^2}} \prod_{i=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(z_i-x_i)^2}{2\sigma^2}} \\ &\propto \left(\prod_{j=1}^{3} \frac{1}{\theta_j} e^{-\frac{(\ln\theta_j-\mu_j)^2}{2\sigma_j^2}}\right) \frac{1}{\sigma^{T+1}}\frac{1}{\Delta} \prod_{i=0}^{T} e^{\frac{-(z_i-x_i(\theta))^2}{2\sigma^2}} \\ &\triangleq \tilde{g}(\theta). \end{aligned}$$

For product of small numbers, I will calculate the natural log of the product first:

$$\begin{aligned} & \ln(\tilde{g}(\theta)) \\ &= -\sum_{j=1}^{3} \ln\theta_j - \sum_{j=1}^{3} \frac{(\ln\theta_j-\mu_j)^2}{2\sigma_j^2} - (T+1)\ln(\sigma) - \ln(\Delta) - \frac{1}{2\sigma^2}\sum_{i=0}^{T}(x_i(\theta)-z_i)^2 \\ &\triangleq G(\theta). \end{aligned} \tag{4.6}$$

Thus,

$$p(\theta|D) \propto p(\theta)p(D|\theta) \triangleq g(\theta) \propto \tilde{g}(\theta) = e^{G(\theta)}.$$

Now the marginal distributions can be calculated. Remember that $(\theta_1, \theta_2, \theta_3) = (\mu, d, \beta)$. With

$$H = \int_0^\infty \int_1^{10^6} \int_0^1 \int_0^1 \int_0^1 e^{G(\theta)} \, \mathrm{d}\theta_1 \mathrm{d}\theta_2 \mathrm{d}\theta_3 \mathrm{d}\sigma \mathrm{d}x_0,$$

the following marginal distributions as posterior probability are obtained:

$$p(\theta_1)$$
$$= \left(\int_0^{\infty} \int_1^{10^6} \int_0^1 \int_0^1 e^{G(\theta)} \, \mathrm{d}\theta_2 \mathrm{d}\theta_3 \mathrm{d}\sigma \mathrm{d}x_0 \right) / H \propto \int_0^{\infty} \int_1^{10^6} \int_0^1 \int_0^1 e^{G(\theta)} \, \mathrm{d}\theta_2 \mathrm{d}\theta_3 \mathrm{d}\sigma \mathrm{d}x_0;$$

$$p(\theta_2)$$
$$= \left(\int_0^{\infty} \int_1^{10^6} \int_0^1 \int_0^1 e^{G(\theta)} \, \mathrm{d}\theta_1 \mathrm{d}\theta_3 \mathrm{d}\sigma \mathrm{d}x_0 \right) / H \propto \int_0^{\infty} \int_1^{10^6} \int_0^1 \int_0^1 e^{G(\theta)} \, \mathrm{d}\theta_1 \mathrm{d}\theta_3 \mathrm{d}\sigma \mathrm{d}x_0;$$

$$p(\theta_3)$$
$$= \left(\int_0^{\infty} \int_1^{10^6} \int_0^1 \int_0^1 e^{G(\theta)} \, \mathrm{d}\theta_1 \mathrm{d}\theta_2 \mathrm{d}\sigma \mathrm{d}x_0 \right) / H \propto \int_0^{\infty} \int_1^{10^6} \int_0^1 \int_0^1 e^{G(\theta)} \, \mathrm{d}\theta_1 \mathrm{d}\theta_2 \mathrm{d}\sigma \mathrm{d}x_0;$$

$$p(\sigma)$$
$$= \left(\int_0^{\infty} \int_0^1 \int_0^1 \int_0^1 e^{G(\theta)} \, \mathrm{d}\theta_1 \mathrm{d}\theta_2 \mathrm{d}\theta_3 \mathrm{d}x_0 \right) / H \propto \int_0^{\infty} \int_0^1 \int_0^1 \int_0^1 e^{G(\theta)} \, \mathrm{d}\theta_1 \mathrm{d}\theta_2 \mathrm{d}\theta_3 \mathrm{d}x_0;$$

$$p(x_0)$$
$$= \left(\int_1^{10^6} \int_0^1 \int_0^1 \int_0^1 e^{G(\theta)} \, \mathrm{d}\theta_1 \mathrm{d}\theta_2 \mathrm{d}\theta_3 \mathrm{d}\sigma \right) / H \propto \int_1^{10^6} \int_0^1 \int_0^1 \int_0^1 e^{G(\theta)} \, \mathrm{d}\theta_1 \mathrm{d}\theta_2 \mathrm{d}\theta_3 \mathrm{d}\sigma.$$

4.3.1.2 σ fixed

Then I also tried to set σ to be a constant, assuming information of σ can be found in literature. I tried $\sigma = 6.4 \times 10^4$, since this is approximately the mean of the posterior distribution for σ (see Figure 4.3 and 4.5). Then in this case, only 4 parameters need to be studied: $\theta = (\theta_1, \theta_2, \theta_3, x_0) = (\mu, d, \beta, x_0)$.

The following parameters are independent of each other: $(\mu, d, \beta, x_0(\sigma))$. So using (4.3) with notation $\Delta \triangleq \int_0^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-z_0)^2}{2\sigma^2}} \, \mathrm{d}x$, the prior probability becomes

$$\begin{aligned} p(\theta) &= p(\mu)p(d)p(\beta)p(x_0) = p(\mu)p(d)p(\beta)\frac{1}{\Delta} e^{-\frac{(x_0-z_0)^2}{2\sigma^2}} \\ &= \left(\prod_{j=1}^{3} \frac{1}{\sqrt{2\pi}\sigma_j\theta_j} e^{-\frac{(\ln\theta_j - \mu_j)^2}{2\sigma_j^2}} \right) \frac{1}{\Delta} e^{-\frac{(x_0-z_0)^2}{2\sigma^2}}. \end{aligned} \tag{4.7}$$

Then using (4.7) and (4.4),

$$g(\theta) = \left(\prod_{j=1}^{3} \frac{1}{\sqrt{2\pi}\sigma_j\theta_j} e^{-\frac{(\ln\theta_j-\mu_j)^2}{2\sigma_j^2}}\right) \frac{1}{\Delta} e^{-\frac{(x_0-z_0)^2}{2\sigma^2}} \prod_{i=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(z_i-x_i(\theta))^2}{2\sigma^2}}$$
$$\propto \left(\prod_{j=1}^{3} \frac{1}{\theta_j} e^{-\frac{(\ln\theta_j-\mu_j)^2}{2\sigma_j^2}}\right) \prod_{i=0}^{T} e^{\frac{-(z_i-x_i(\theta))^2}{2\sigma^2}} \triangleq \tilde{g}(\theta).$$

Then (4.6) becomes

$$\ln(\tilde{g}(\theta)) = -\sum_{j=1}^{3} \ln\theta_j - \sum_{j=1}^{3} \frac{(\ln\theta_j-\mu_j)^2}{2\sigma_j^2} - \frac{1}{2\sigma^2}\sum_{i=0}^{T}(x_i(\theta)-z_i)^2 \triangleq G(\theta). \tag{4.8}$$

Then with

$$H = \int_0^\infty \int_0^1 \int_0^1 \int_0^1 e^{G(\theta)}\,\mathrm{d}\theta_1\mathrm{d}\theta_2\mathrm{d}\theta_3\mathrm{d}x_0,$$

the following marginal distributions as posterior probability are obtained:

$$p(\theta_1) = \left(\int_0^\infty \int_0^1 \int_0^1 e^{G(\theta)}\,\mathrm{d}\theta_2\mathrm{d}\theta_3\mathrm{d}x_0\right)/H \propto \int_0^\infty \int_0^1 \int_0^1 e^{G(\theta)}\,\mathrm{d}\theta_2\mathrm{d}\theta_3\mathrm{d}x_0;$$
$$p(\theta_2) = \left(\int_0^\infty \int_0^1 \int_0^1 e^{G(\theta)}\,\mathrm{d}\theta_1\mathrm{d}\theta_3\mathrm{d}x_0\right)/H \propto \int_0^\infty \int_0^1 \int_0^1 e^{G(\theta)}\,\mathrm{d}\theta_1\mathrm{d}\theta_3\mathrm{d}x_0;$$
$$p(\theta_3) = \left(\int_0^\infty \int_0^1 \int_0^1 e^{G(\theta)}\,\mathrm{d}\theta_1\mathrm{d}\theta_2\mathrm{d}x_0\right)/H \propto \int_0^\infty \int_0^1 \int_0^1 e^{G(\theta)}\,\mathrm{d}\theta_1\mathrm{d}\theta_2\mathrm{d}x_0;$$
$$p(x_0) = \left(\int_0^1 \int_0^1 \int_0^1 e^{G(\theta)}\,\mathrm{d}\theta_1\mathrm{d}\theta_2\mathrm{d}\theta_3\right)/H \propto \int_0^1 \int_0^1 \int_0^1 e^{G(\theta)}\,\mathrm{d}\theta_1\mathrm{d}\theta_2\mathrm{d}\theta_3.$$

4.3.2 MCMC Method

Evaluating the posterior distribution often requires the integration of high dimensional functions (see previous section). I did not get good results from numerically calculating these triple integrals and quadruple integrals using Matlab. Therefore I will try the Markov Chain Monte Carlo (MCMC) method using the Metropolis-Hastings algorithm.

The original Monte Carlo method uses random number generation to compute integrals. To generate a Markov chain, one uses the previous sample value to generate

the next sample value, and the transition probability is just a function of the previous sample value. As the chain continues, the sample values get spread out over the possible state space. The idea of discrete-time Markov chain can be generalized to a continuous-time Markov process.

The MCMC method initially proposed by Metropolis requires the proposal distribution (or the transition probability) to be symmetric. Later this method was modified by Hastings to work for asymmetric proposal distributions as well (Gregory (2005)). In the simulations, to better make sure the convergence of the chains, one should use multiple chains each starting from different initial values.

4.3.2.1 Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm consists of two stages: the first stage is to generate a candidate from the proposal distribution, and the second stage is the accept-reject step. The proof of that the stationary distribution of the Markov chain generated by the M-H algorithm is the target posterior distribution can be found in many material (such as in (Gregory (2005))). Now let's present the algorithm with details.

Metropolis-Hastings algorithm(MCMC method):

1. Initialize parameters θ.

2. Repeat

   (a) Propose new values for θ as θ^* by sampling from the proposal density $Q(\theta^*, x^*_{1:T}|\theta, x_{1:T})$, and calculate corresponding $x^*_{1:T}$ from ODE.

   (b) With probability

$$\min\left(\frac{\Pr(\theta^*, x^*_{1:T}|z_{1:T})}{\Pr(\theta, x_{1:T}|z_{1:T})}\frac{Q(\theta, x_{1:T}|\theta^*, x^*_{1:T})}{Q(\theta^*, x^*_{1:T}|\theta, x_{1:T})}, 1\right),$$

   set $\theta = \theta^*$; otherwise set $\theta = \theta$.

Derived from the Bayes' formula : $\Pr(\theta, x_{1:T}|z_{1:T}) = \dfrac{\Pr(z_{1:T}, x_{1:T}|\theta)\Pr(\theta)}{\Pr(z_{1:T})}$ and $\Pr(\theta^*, x^*_{1:T}|z_{1:T}) = \dfrac{\Pr(z_{1:T}, x^*_{1:T}|\theta^*)\Pr(\theta^*)}{\Pr(z_{1:T})}$, where $\Pr(z_{1:T}, x_{1:T}|\theta)$ and $\Pr(z_{1:T}, x^*_{1:T}|\theta^*)$ are the total likelihoods. So the acceptance probability can be rewritten as:

$$\begin{aligned}&\min\left(\frac{\Pr(\theta^*, x^*_{1:T}|z_{1:T})}{\Pr(\theta, x_{1:T}|z_{1:T})}\frac{Q(\theta, x_{1:T}|\theta^*, x^*_{1:T})}{Q(\theta^*, x^*_{1:T}|\theta, x_{1:T})}, 1\right)\\ &= \min\left(\frac{\Pr(z_{1:T}, x^*_{1:T}|\theta)\Pr(\theta^*)}{\Pr(z_{1:T}, x_{1:T}|\theta)\Pr(\theta)}\frac{Q(\theta, x_{1:T}|\theta^*, x^*_{1:T})}{Q(\theta^*, x^*_{1:T}|\theta, x_{1:T})}, 1\right),\end{aligned}$$

as the marginal probability of the data $\Pr(z_{1:T})$ cancels out during the calculation.

4.3.2.2 σ uncertain

In this case, it is assumed that σ is a distribution with a Jeffreys prior, then $\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5) = (\mu, d, \beta, \sigma, x_0(\sigma))$. Then $\theta$is initialized as $\theta_0 = (0.01, 0.01, 0.01, 10^3,$ normrnd$(z_0, 10^3))$ with a normrnd$(z_0, 10^3) > 0$. Then the stochastic kernel $Q(\theta^*_{1:5}|\theta_{1:5})$ is chosen such that $\theta^*_{1:4} = e^{\ln(\theta_{1:4})+\varepsilon}$ (generates only positive values) and $\theta^*_5 = z_0 + N(0, \theta^{*2}_4)$ (accepts only positive values), where ε has the multivariate normal distribution $N(0, \Sigma^2)$. Here $\Sigma = (\tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{\sigma}_3, \tilde{\sigma}_4)$, which is assumed to be $\Sigma = (0.1, 0.1, 0.1, 10)$ in the simulations. Notice that I choose a much larger value $\tilde{\sigma}_4 = 10$ for σ.

Now calculate $\dfrac{Q(\theta, x_{1:T}|\theta^*, x^*_{1:T})}{Q(\theta^*, x^*_{1:T}|\theta, x_{1:T})}$ in details first.

$\theta^*_{1:4} = e^{\ln(\theta_{1:4})+\epsilon}$ and $x^*_0 \sim N(z_0; \sigma^*)$ (accepts only positive values) imply that

$$Q(\theta^*|\theta) = \left(\prod_{j=1}^{4}\frac{1}{\sqrt{2\pi\tilde{\sigma}^2_j}}e^{\frac{-(\ln(\theta^*_j)-\ln(\theta_j))^2}{2\tilde{\sigma}^2_j}}\right)e^{\frac{-(x^*_0-z_0)^2}{2\sigma^{*2}}}/\int_0^\infty e^{-\frac{(x-z_0)^2}{2\sigma^{*2}}}\,\mathrm{d}x.$$

Similarly,

$$Q(\theta|\theta^*) = \left(\prod_{j=1}^{4}\frac{1}{\sqrt{2\pi\tilde{\sigma}^2_j}}e^{\frac{-(\ln(\theta_j)-\ln(\theta^*_j))^2}{2\tilde{\sigma}^2_j}}\right)e^{\frac{-(x_0-z_0)^2}{2\sigma^2}}/\int_0^\infty e^{-\frac{(x-z_0)^2}{2\sigma^2}}\,\mathrm{d}x.$$

Then

$$\frac{Q(\theta|\theta^*)}{Q(\theta^*|\theta)} = \frac{\int_0^\infty e^{-\frac{(x-z_0)^2}{2\sigma^{*2}}}\,\mathrm{d}x}{\int_0^\infty e^{-\frac{(x-z_0)^2}{2\sigma^2}}\,\mathrm{d}x}e^{\frac{(x^*_0-z_0)^2}{2\sigma^{*2}}-\frac{(x_0-z_0)^2}{2\sigma^2}}.$$

Recall the likelihood function expressed in equation (4.4) and prior probability expressed in equation (4.5), then

$$
\begin{aligned}
&\frac{\Pr(z_{1:T}, x^*_{1:T}|\theta^*)\Pr(\theta^*)Q(\theta|\theta^*)}{\Pr(z_{1:T}, x_{1:T}|\theta)\Pr(\theta)Q(\theta^*|\theta)} \\
&= \exp\left(\sum_{j=1}^{3}\frac{(\ln(\theta_j)-\mu_j)^2-(\ln(\theta^*_j)-\mu_j)^2}{2\sigma_j^2}+\sum_{i=0}^{T}\left(\frac{(x_i-z_i)^2}{2\sigma^2}-\frac{(x^*_i-z_i)^2}{2\sigma^{*2}}\right)\right) \\
&\quad\cdot\left(\prod_{j=1}^{3}\frac{\theta_j}{\theta^*_j}\right)\left(\frac{\sigma}{\sigma^*}\right)^{T+1}\frac{\int_0^\infty e^{-\frac{(x-z_0)^2}{2\sigma^2}}\,\mathrm{d}x}{\int_0^\infty e^{-\frac{(x-z_0)^2}{2\sigma^{*2}}}\,\mathrm{d}x}\cdot\exp\left(\frac{(x^*_0-z_0)^2}{2\sigma^{*2}}-\frac{(x_0-z_0)^2}{2\sigma^2}\right)\frac{\int_0^\infty e^{-\frac{(x-z_0)^2}{2\sigma^{*2}}}\,\mathrm{d}x}{\int_0^\infty e^{-\frac{(x-z_0)^2}{2\sigma^2}}\,\mathrm{d}x} \\
&= \left(\prod_{j=1}^{3}\frac{\theta_j}{\theta^*_j}\right)\left(\frac{\sigma}{\sigma^*}\right)^{T+1} \\
&\quad\cdot\exp\left(\sum_{j=1}^{3}\frac{(\ln(\theta_j)-\mu_j)^2-(\ln(\theta^*_j)-\mu_j)^2}{2\sigma_j^2}+\sum_{i=1}^{T}\left(\frac{(x_i-z_i)^2}{2\sigma^2}-\frac{(x^*_i-z_i)^2}{2\sigma^{*2}}\right)\right) \\
&\triangleq e^F,
\end{aligned}
$$

with

$$
\begin{aligned}
F = &\left(\sum_{j=1}^{3}(\ln(\theta_j)-\ln(\theta^*_j))\right)+(T+1)(\ln(\sigma)-\ln(\sigma^*)) \\
&+\sum_{j=1}^{3}\frac{(\ln(\theta_j)-\mu_j)^2-(\ln(\theta^*_j)-\mu_j)^2}{2\sigma_j^2}+\sum_{i=1}^{T}\left(\frac{(x_i-z_i)^2}{2\sigma^2}-\frac{(x^*_i-z_i)^2}{2\sigma^{*2}}\right).
\end{aligned}
$$

Thus, the acceptance probability in each iteration (2(b)) becomes $\min(e^F, 1)$.

Notice that the expression for F is closely related to expression (4.6).

In the simulation, I assume that during each step (2(a)), a new sample is obtained by perturbing just one component of θ with equal probability. Thus, during each step, the probability of (only) perturbing θ_i is $\frac{1}{5}$ for $i = 1, 2, 3, 4, 5$. Therefore the stochastic kernel $Q(\theta^*_{1:5}|\theta_{1:5})$ is actually chosen such that $\theta^*_i = e^{\ln(\theta_i)+N(0,\tilde{\sigma}_i^2)}$ with probability $\frac{1}{5}$ for each $1 \le i \le 4$ and $\theta^*_5 = z_0 + N(0, \theta^{*2}_4)$ with probability $\frac{1}{5}$.

3×10^6 iterations have been performed in each simulation. The simulation results with the burn-in period (first 10^3 iterations) omitted is presented in Figure 4.3.

4.3.2.3 σ fixed

In this case, $\sigma = 6.4 \times 10^4$ is fixed, then $\theta = (\theta_1, \theta_2, \theta_3, \theta_4) = (\mu, d, \beta, x_0)$. It is initialized as $\theta_0 = (0.01, 0.01, 0.01, \text{normrnd}(z_0, 6.4\times 10^4))$ with a $\text{normrnd}(z_0, 6.4\times 10^4) > 0$. Then the stochastic kernel $Q(\theta^*_{1:4}|\theta_{1:4})$ is chosen such that $\theta^*_{1:3} = e^{\ln(\theta_{1:3})+\varepsilon}$ (generates only positive values) and $\theta^*_4 = z_0 + N(0, (6.4 \times 10^4)^2)$ (accepts only positive values), where ε has the multivariate normal distribution $N(0, \Sigma^2)$. Here $\Sigma = (\tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{\sigma}_3)$, which is assumed to be $\Sigma = (0.1, 0.1, 0.1)$ in the simulations.

In this case, $\theta^*_{1:3} = e^{\ln(\theta_{1:3})+\epsilon}$ and $x^*_0 \sim N(z_0; \sigma)$ (accepts only positive values) imply that

$$Q(\theta^*|\theta) = \left(\prod_{j=1}^{3} \frac{1}{\sqrt{2\pi\tilde{\sigma}_j^2}} e^{\frac{-(\ln(\theta_j^*)-\ln(\theta_j))^2}{2\tilde{\sigma}_j^2}} \right) e^{\frac{-(x_0^*-z_0)^2}{2\sigma^2}} / \int_0^\infty e^{-\frac{(x-z_0)^2}{2\sigma^2}} \, \mathrm{d}x.$$

Similarly,

$$Q(\theta|\theta^*) = \left(\prod_{j=1}^{3} \frac{1}{\sqrt{2\pi\tilde{\sigma}_j^2}} e^{\frac{-(\ln(\theta_j)-\ln(\theta_j^*))^2}{2\tilde{\sigma}_j^2}} \right) e^{\frac{-(x_0-z_0)^2}{2\sigma^2}} / \int_0^\infty e^{-\frac{(x-z_0)^2}{2\sigma^2}} \, \mathrm{d}x.$$

Then

$$\frac{Q(\theta|\theta^*)}{Q(\theta^*|\theta)} = e^{\frac{(x_0^*-z_0)^2-(x_0-z_0)^2}{2\sigma^{*2}}}.$$

Recall the likelihood function expressed in equation (4.4) and prior probability expressed in equation (4.7), then

$$\begin{aligned}
&\frac{\Pr(z_{1:T}, x^*_{1:T}|\theta^*) \Pr(\theta^*) Q(\theta|\theta^*)}{\Pr(z_{1:T}, x_{1:T}|\theta) \Pr(\theta) Q(\theta^*|\theta)} \\
&= \left(\prod_{j=1}^{3} \frac{\theta_j}{\theta_j^*} \right) \exp\left(\sum_{j=1}^{3} \frac{(\ln(\theta_j) - \mu_j)^2 - (\ln(\theta_j^*) - \mu_j)^2}{2\sigma_j^2} + \sum_{i=0}^{T} \frac{(x_i - z_i)^2 - (x_i^* - z_i)^2}{2\sigma^2} \right) \\
&\quad \cdot \exp\left(\frac{(x_0^* - z_0)^2 - (x_0 - z_0)^2}{2\sigma^{*2}} \right) \\
&\triangleq e^F,
\end{aligned}$$

with

$$F = \left(\sum_{j=1}^{3} (\ln(\theta_j) - \ln(\theta_j^*)) \right) + \sum_{j=1}^{3} \frac{(\ln(\theta_j) - \mu_j)^2 - (\ln(\theta_j^*) - \mu_j)^2}{2\sigma_j^2} + \sum_{i=1}^{T} \frac{(x_i - z_i)^2 - (x_i^* - z_i)^2}{2\sigma^2}.$$

Thus, the acceptance probability in each iteration (2(b)) becomes $\min(e^F, 1)$.

Notice that the expression for F is closely related to expression (4.8).

This time, during each iteration, the probability of (only) perturbing θ_i is $\frac{1}{4}$ for $i = 1, 2, 3, 4$. Therefore the stochastic kernel $Q(\theta_{1:4}^*|\theta_{1:4})$ is actually chosen such that $\theta_i^* = e^{\ln(\theta_i)+N(0,\tilde{\sigma}_i^2)}$ with probability $\frac{1}{4}$ for each $1 \leq i \leq 3$ and $\theta_4^* = z_0 + N(0, (6.4 \times 10^4)^2)$ with probability $\frac{1}{4}$.

3×10^6 iterations have been performed in each simulation. When $\sigma = 6.4 \times 10^4$ is fixed, the simulation results without burn in period (first 10^3 iterations) is presented in Figure 4.4. See the corresponding posterior distribution results also in Figure 4.4.

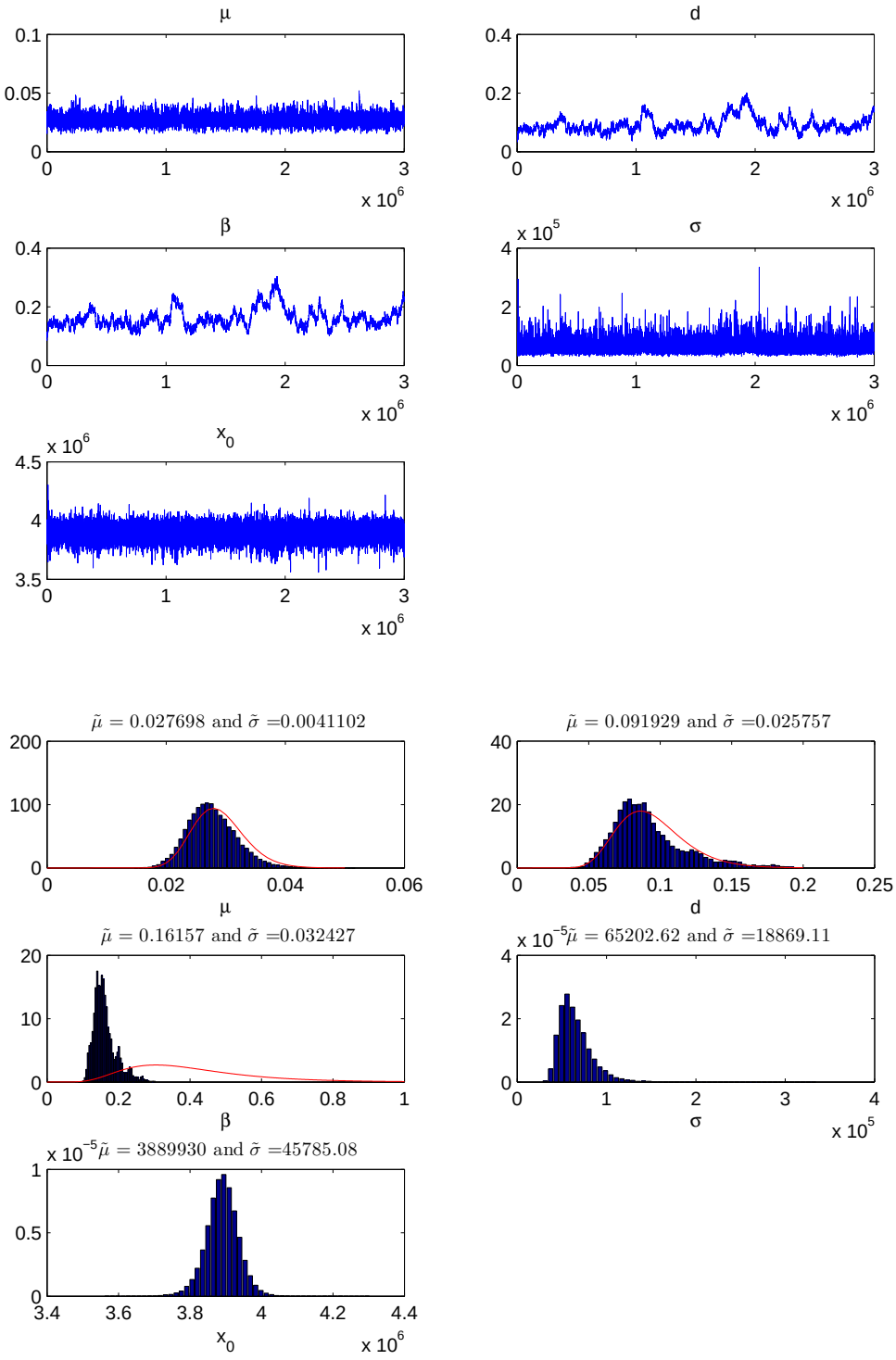


Figure 4.3: σ random. All 3×10^6 iterations. Burn in period dropped. Red curves indicate the prior distributions. See Figure 4.2 for prior distributions for σ and x_0.

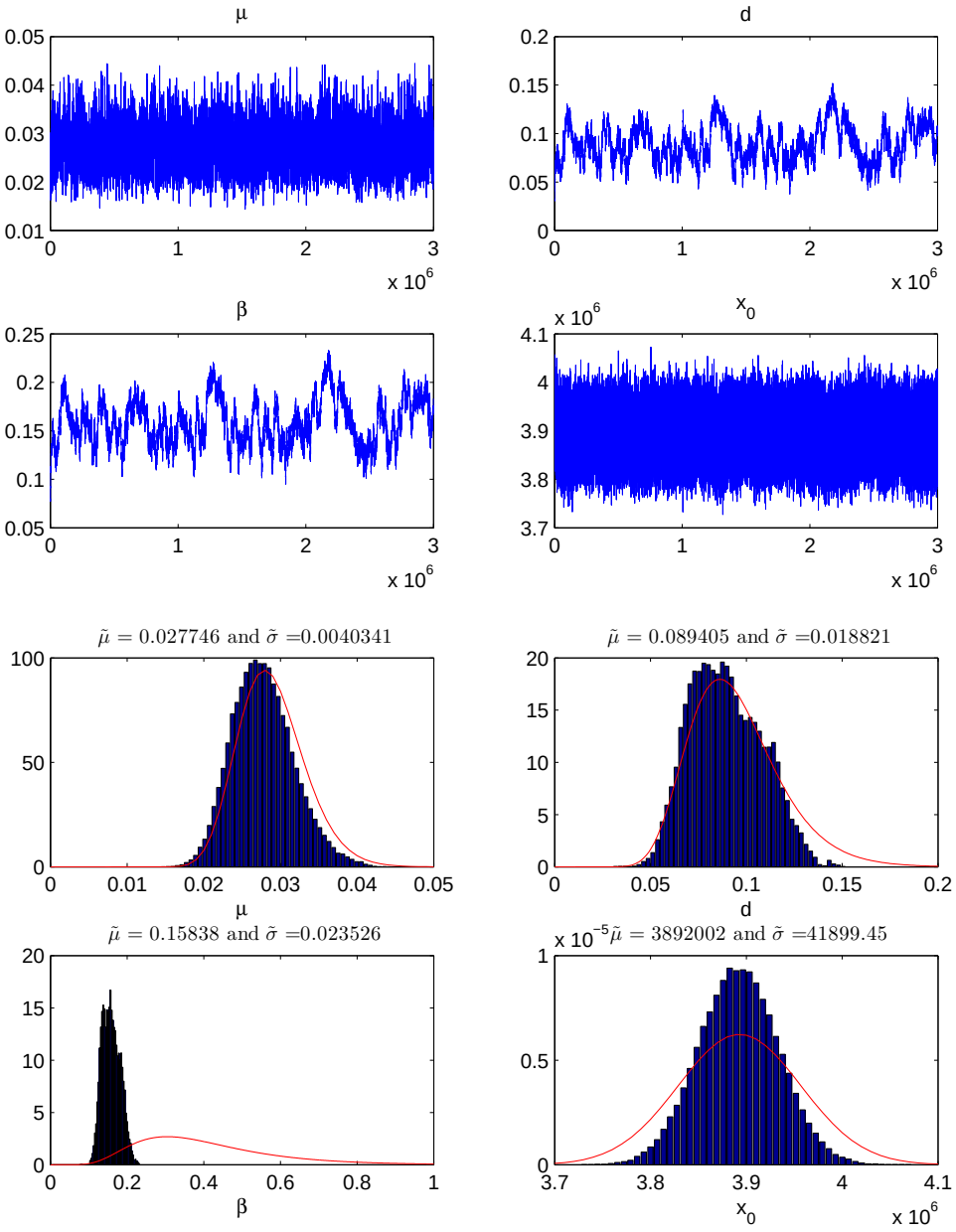


Figure 4.4: $\sigma = 6.4 \times 10^4$ fixed. All 3×10^6 iterations. Burn in period dropped. Red curves indicate the prior distributions.

4.3.2.4 Improved MCMC

From previous simulation results (Figure 4.3 and 4.4), it shows that the chains for d and β are not mixing as well as other parameters. Also, the chains for β and d are highly correlated to each other (correlation coefficients are 0.9822 and 0.9683 for Figure 4.3 and 4.4 respectively). I think the strong correlation can be interpreted in the following way. Once β (cumulative HIV-acquisition risk) increases, more susceptible population become infected. Then in order to balance the HIV population, d (annual rate of progression to AIDS) needs to be increased as well, so that more infections will become sexually inactive as they develop AIDS. Therefore to get better results, or better mixing chains for d and β, I will treat β and d as a group in the following way.

- When σ has a Jeffreys prior:
  - the probability to perturb μ is $\frac{1}{5}$;
  - the probability to perturb both d and β is $\frac{2}{5}$;
  - the probability to perturb σ is $\frac{1}{5}$;
  - the probability to perturb x_0 is $\frac{1}{5}$.

  Thus, for a probability of $\frac{2}{5}$, a new sample θ^* is proposed by proposing a new value d^* for d and a new value β^* for β independently at the same time. If θ^* is accepted, then both d^* and β^* are accepted. It is called that d and β are grouped in this case. See corresponding results in Figure 4.5.

- When $\sigma = 6.4 \times 10^4$ is fixed:
  - the probability to perturb μ is $\frac{1}{5}$;
  - the probability to perturb both d and β is $\frac{2}{5}$;

– the probability to perturb x_0 is $\frac{2}{5}$.

See the corresponding results in Figure 4.6.

The Markov chains in Figure 4.5 and 4.6 show much better mixing in d and β, after they have been grouped. Therefore the corresponding posterior distributions are better than those before d and β are grouped (Figure 4.3 and 4.4).

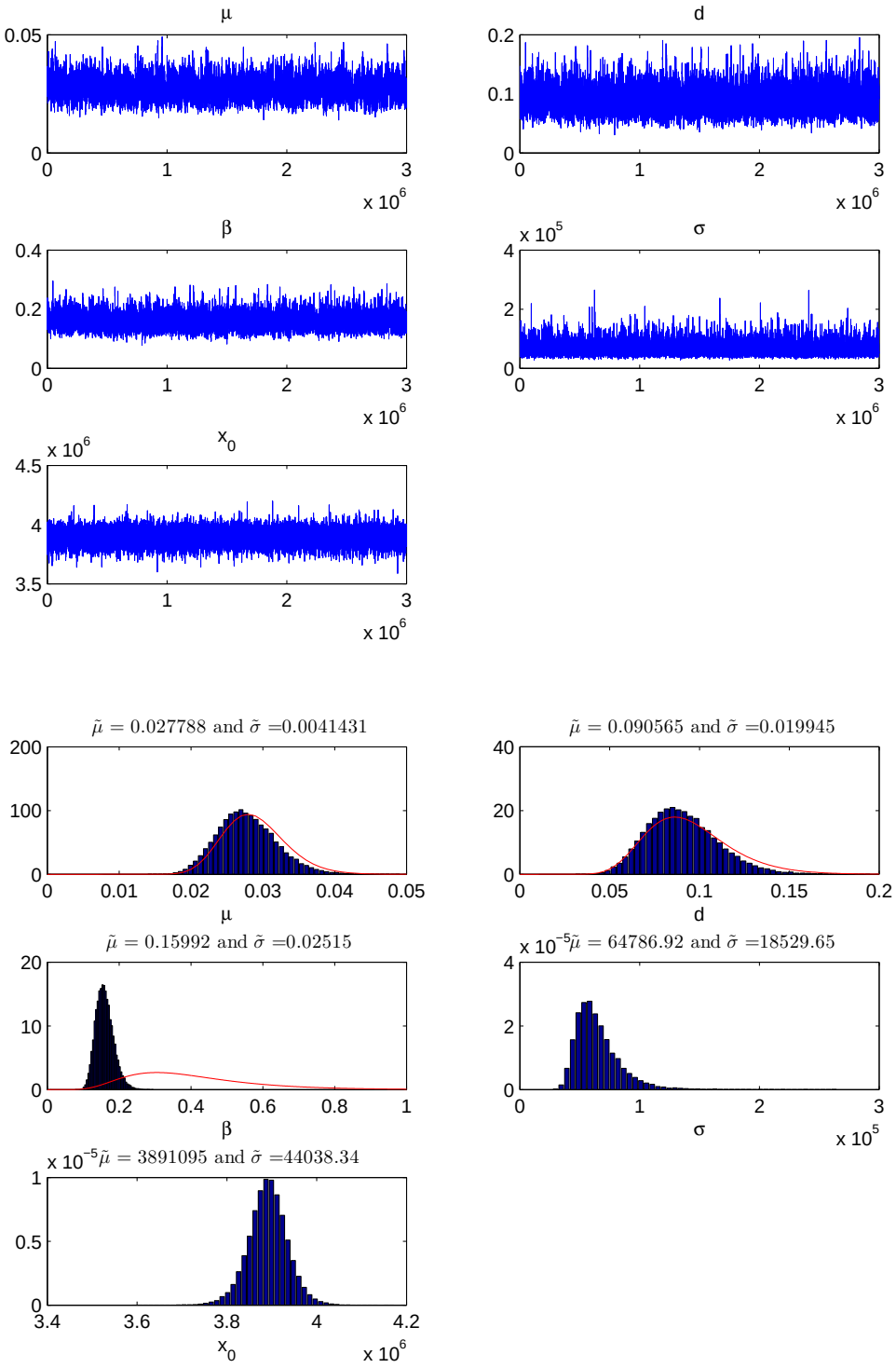


Figure 4.5: σ is random. All 3×10^6 iterations. Burn in period dropped. Red curves indicate the prior distributions. β and d grouped.

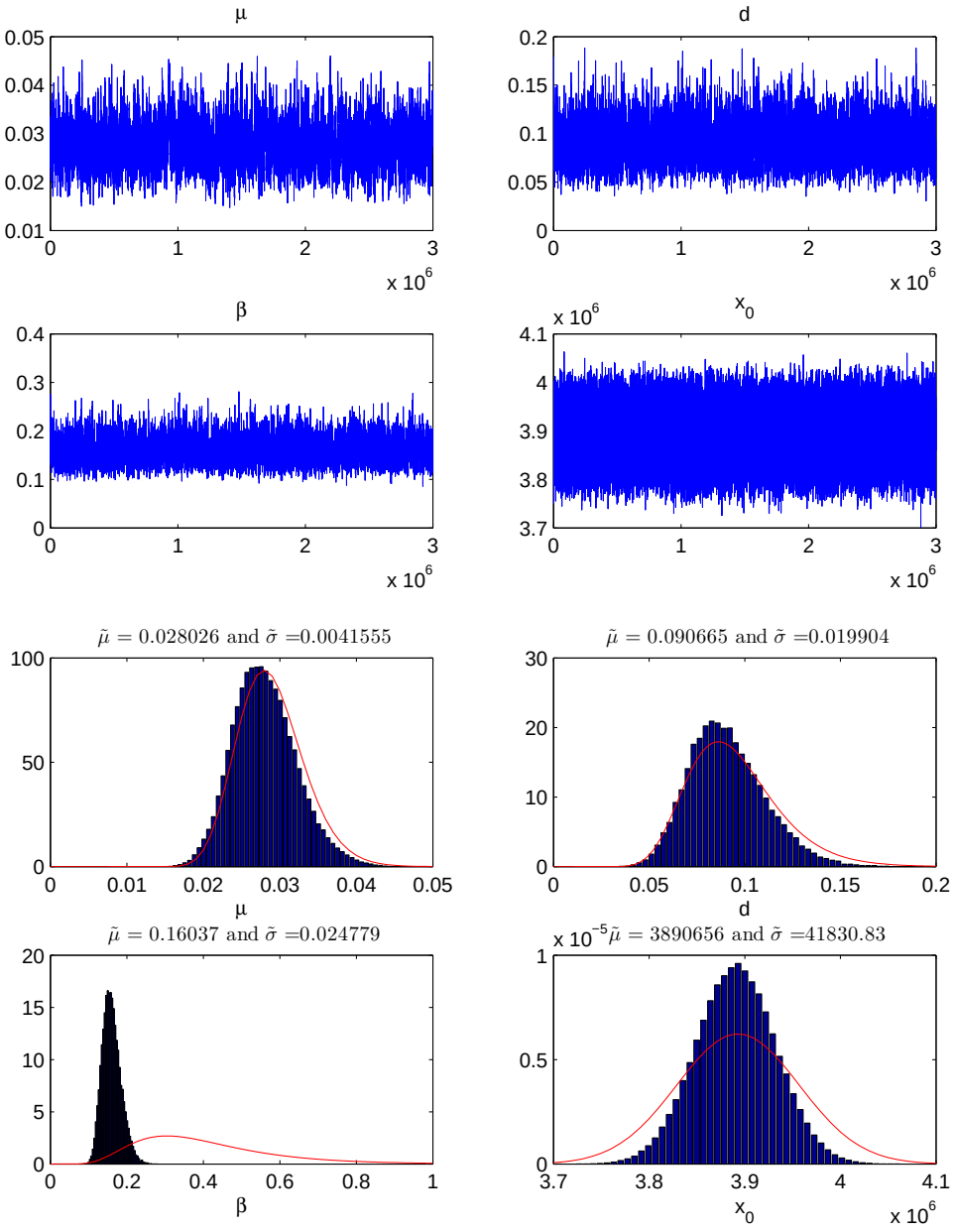


Figure 4.6: $\sigma = 6.4 \times 10^4$ fixed. All 3×10^6 iterations. Burn in period dropped. Red curves indicate the prior distributions. β and d grouped.

4.4 Parameter Posterior Distributions

From data fitting as in Figure D.1, $\mu = 0.02979$, $d = 0.1191$, and $\beta = 0.1969$ are obtained as a set of baseline parameter values. Now from Bayesian parameter estimation (Figure 4.5 and 4.6), the posterior distributions are obtained for these parameters. The mode and mean of these distributions are presented in Table 4.4. The mean values are quite different from the parameter values obtained from data fitting as in Figure D.1, especially a relative difference about $\frac{|0.197-0.16|}{|0.16|} \approx 23\%$ for β is observed.

Table 4.4: Parameter values comparison

	Figure D.1	σ as distribution		σ fixed	
		mean	mode	mean	mode
μ	0.02979	0.02779	0.027	0.02803	0.0276
d	0.1191	0.09057	0.085	0.09066	0.0827
β	0.1969	0.1599	0.154	0.1604	0.15
x_0	3892911.84	3.8911×10^6	3.89×10^6	3.8907×10^6	3.89×10^6

If the mode and mean are very different, the posterior PDF is too asymmetric to be adequately summarized by a single estimate (Gregory (2005)). Here, the mode and mean are quite close, but I still present the simulation results from two independent Markov chains.

For each simulation, two independent Markov chains are generated. The summary of mean, standard deviation, 95% credibility interval, and acceptance rate for each parameter is presented in Table 4.5-4.8.

After d and β are grouped (Table 4.7 and 4.8), the posterior distributions from

two independent Markov chains are very close to each other, and the convergence of chains is believed. On the other hand side, the posterior distributions from two independent Markov chains are not that close when d and β are not grouped (Table 4.5 and 4.6).

After combining d and β as a group during perturbation, although the acceptance has been decreased from 12.13% and 8.31% (Table 4.5) to 6.51% (Table 4.7), I have increased the probability to perturb d or β from $\frac{1}{5}$ to $\frac{2}{5}$ which is a compensation. Therefore a better mixing is obtained for the Markov chains of d and β (Figure 4.5).

Table 4.5: Posterior distributions of 2 independent chains, corresponding to Figure 4.3. Total acceptance rate is 18.08% for MCMC 1 and 18.09% for MCMC 2.

MCMC 1	μ	d	β	σ	x_0
Mean	0.0277	0.09193	0.1616	65202.6	3.8899×10^6
Standard Deviation	4.11×10^{-3}	25.76×10^{-3}	32.43×10^{-3}	18.87×10^3	45.79×10^3
95% Credibility Interval	[0.02018 0.0358]	[0.05527 0.15548]	[0.1146 0.2407]	[37450.7 104879.9]	$[3.7909\times10^6$ $3.969\times10^6]$
Acceptance Rate	31.78%	12.13%	8.31%	2.92%	35.27%
MCMC 2	μ	d	β	σ	x_0
Mean	0.02786	0.08886	0.1579	64900.5	3.8915×10^6
Standard Deviation	4.064×10^{-3}	16.8×10^{-3}	20.97×10^{-3}	18.53×10^3	44.18×10^3
95% Credibility Interval	[0.02066 0.03616]	[0.05983 0.1216]	[0.1209 0.1989]	[41664.3 105424.4]	$[3.8058\times10^6$ $3.9671\times10^6]$
Acceptance Rate	31.56%	12.14%	8.33%	2.94%	35.47%

Table 4.6: Posterior distributions of 2 independent chains, corresponding to Figure 4.4. Total acceptance rate is 21.95% for MCMC 1 and 21.73% for MCMC 2.

MCMC 1	μ	d	β	x_0
Mean	0.02775	0.0894	0.1584	3.892×10^{6}
Standard Deviation	4.034×10^{-3}	18.82×10^{-3}	23.53×10^{-3}	41.9×10^{3}
95% Credibility Interval	[0.02072	[0.05844	[0.1193	$[3.8068\times10^{6}$
	0.03574]	0.12386]	0.2033]	$3.9655\times10^{6}]$
Acceptance Rate	31.83%	12.15%	12.15%	35.55%
MCMC 2	μ	d	β	x_0
Mean	0.02782	0.09601	0.1668	3.8893×10^{6}
Standard Deviation	4.139×10^{-3}	23.59×10^{-3}	29.47×10^{-3}	41.99×10^{3}
95% Credibility Interval	[0.02057	[0.05479	[0.1136	$[3.8051\times10^{6}$
	0.03577]	0.15023]	0.234]	$3.9630\times10^{6}]$
Acceptance Rate	31.8%	11.56%	11.56%	35.48%

Table 4.7: Posterior distributions of 2 independent chains, corresponding to Figure 4.5. Total acceptance rate is 16.58% for MCMC 1 and 16.6% for MCMC 2.

MCMC 1	μ	d	β	σ	x_0
Mean	0.02779	0.09057	0.1599	64786.9	3.8911×10^6
Standard Deviation	4.143×10^{-3}	19.94×10^{-3}	25.15×10^{-3}	18.53×10^3	44.04×10^3
95%	[0.02062	[0.05545	[0.119	[39298.1	$[3.8018\times10^6$
Credibility Interval	0.03609]	0.13109]	0.2104]	105758]	3.9601×10^6]
Acceptance Rate	31.45%	6.51%	6.51%	2.94%	35.5%
MCMC 2	μ	d	β	σ	x_0
Mean	0.02781	0.09077	0.1602	64862.6	3.8912×10^6
Standard Deviation	4.194×10^{-3}	20.34×10^{-3}	25.52×10^{-3}	18.77×10^3	44.34×10^3
95%	[0.02061	[0.05783	[0.1155	[39039.5	$[3.8085\times10^6$
Credibility Interval	0.03589]	0.13135]	0.2117]	106115.7]	3.9709×10^6]
Acceptance Rate	31.52%	6.5%	6.5%	2.96%	35.5%

Table 4.8: Posterior distributions of 2 independent chains, corresponding to Figure 4.6. Total acceptance rate is 23.09% for MCMC 1 and 23.09% for MCMC 2.

MCMC 1	μ	d	β	x_0
Mean	0.02803	0.09066	0.1604	3.8907×10^6
Standard Deviation	4.155×10^{-3}	19.9×10^{-3}	24.78×10^{-3}	41.83×10^3
95% Credibility Interval	[0.02069	[0.05747	[0.119	$[3.8063\times10^6$
	0.03632]	0.13002]	0.2126]	$3.9654\times10^6]$
Acceptance Rate	31.56%	6.44%	6.44%	35.51%
MCMC 2	μ	d	β	x_0
Mean	0.02804	0.09134	0.1612	3.8904×10^6
Standard Deviation	4.17×10^{-3}	20.32×10^{-3}	25.39×10^{-3}	41.65×10^3
95% Credibility Interval	[0.02095	[0.058	[0.1166	$[3.8095\times10^6$
	0.03585]	0.13298]	0.2119]	$3.9651\times10^6]$
Acceptance Rate	31.49%	6.40%	6.40%	35.54%

It shows that the posterior distributions for both μ and d stays close to their corresponding prior distributions; while the posterior distribution for β differs a lot with its prior distribution, both in mean value and standard deviation. For β, the standard deviation (0.02515 from Table 4.7) of the posterior distribution is much smaller than the standard deviation (0.18981) of the prior distribution, providing a sharper estimation for β.

The little difference between prior and posterior distributions for μ and d implies that the data used is not very informative. Therefore the posterior distributions are mainly determined by the information provided by the prior distributions.

Notice that $\beta = 1 - (1 - b_a)^n$ is a composite parameter. Information on b_a and n

can both be found in literature (Boily *et al.* (2009),Wawer *et al.* (2005)). In this study, a posterior distribution for β (mean = 0.1599 and standard deviation = 0.02515 from Table 4.7) is also obtained, which can be a supplement to the previous HIV studies.

σ assumed with a Jeffreys prior, or assumed to be a constant, give similar posterior distributions for x_0 (as well as μ, d, and β), and standard deviation for the former is slightly larger. This is because, the posterior distribution of σ adds a diffusive effect on $x_0(\sigma)$ compared with when σ fixed as a constant.

From the simulation results, it is believed that $\sigma \in [39298, 105758]$ (95% CI from Table 4.7) gives a good approximation for the observation error in HIV population in South Africa. Correspondingly, it is believed that $x_0 \in [3.8018 \times 10^6, 3.9601 \times 10^6]$ (95% CI) gives a good approximation for the actual HIV population in 2001.

4.5 Model Validation

Because the consecutive values generated generated by a MCMC simulation are correlated, they provide less accurate estimates than independent samples. One measure of the accuracy of the estimate and how well the chain is mixing is the effective sample size (ESS).

The ESS is equal to

$$ESS = \frac{N}{1 + 2\sum_{k=1}^{\infty} \rho_k(\theta)},$$

where N is the number of posterior samples, ρ_k is the autocorrelation at lag k. The infinite sum is often truncated when $\rho_k < 0.05$ (ESS (2014),Kass *et al.* (1998)). I will use the sample autocorrelation function to estimate the ESS:

$$\rho_k(\theta_j) = \frac{\sum_{i=1}^{N-k} [\theta_j(i) - \tilde{\mu}] \cdot [\theta_j(i+k) - \tilde{\mu}]}{\sum_{i=1}^{N} [\theta_j(i) - \tilde{\mu}]^2}, \ k = 1 \cdots N - 1.$$

Here $\tilde{\mu}$ is the sample mean of θ_j, and $\sum_{i=1}^{N}[\theta_j(i) - \tilde{\mu}]^2$ is the sample variance of θ_j. The sum will be truncated when $\rho_k < 0.01$.

Table 4.9: Effective Sample Size, when d and β are not grouped.

number of samples	μ	d	β	σ	x_0
10^4	6	5	7	44	75
10^5	48	5	4	252	489
$10^6 - 10^3$	422	21	22	2610	4282
$3 \times 10^6 - 10^3$	1431	24	24	*2154*	*624*

Table 4.10: Effective Sample Size, when d and βd are grouped.

number of samples	μ	d	β	σ	x_0
10^4	19	10	10	41	59
10^5	39	131	123	302	388
$10^6 - 10^3$	404	935	877	3095	3154
$3 \times 10^6 - 10^3$	1298	2639	2326	7873	9081

Then for Figure 4.3 and 4.5, the corresponding results for ESS for each parameter are presented in Table 4.9 and 4.10 respectively. I have calculated the ESS given different number of samples. Originally, only 10^6 iterations were performed in each iteration. After dropping the first 10^3 iterations, the ESS of some parameters (μ, d and β in Table 4.10) can be less than 1000. Then I decided to run 3×10^6 iterations to get a better result. Then after dropping the first 10^3 iterations, the ESS for each parameter becomes 1298, 2639, 2326, 7873, and 9081, all greater than 1000 (ESS (2014)).

Also, notice that after d and β are grouped, the ESS for d and β have been

increased dramatically. This is consistent with the better mixing in Markov chains that have been observed. ESS given $3 \times 10^6 - 10^3$ samples is expected to greater than ESS given $10^6 - 10^3$ samples. However, for σ and x_0 in Table 4.9, this is not the case. I think this may be due to the fact that ESS can only be estimated but calculated exactly.

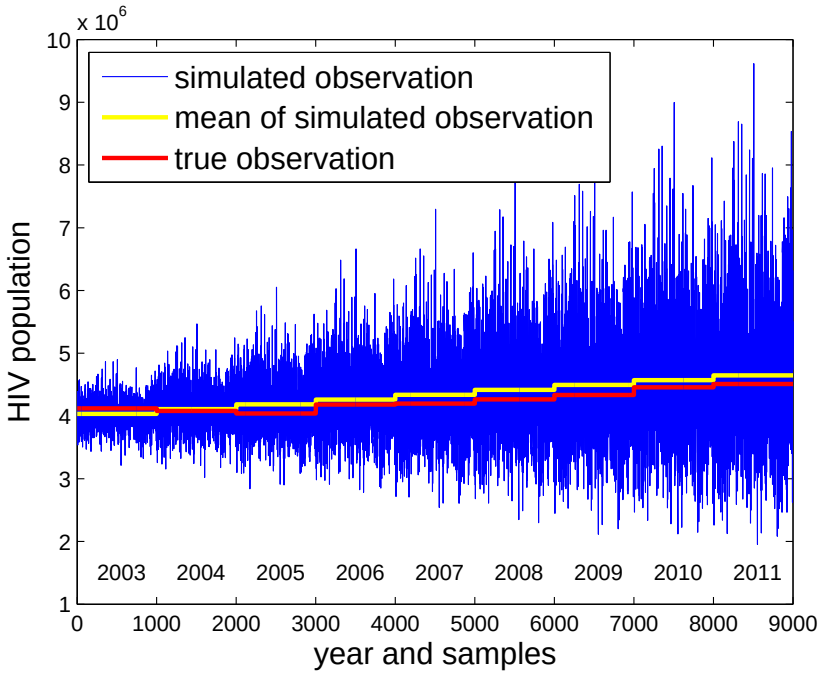


Figure 4.7: From year 2003 to 2011, 1000 samples of simulated HIV population observations are plotted for each year. The mean of the simulated observations is plotted in yellow for each year. The true observation is plotted in red for each year.

For model validation, 1000 independent samples are drawn from the posterior distributions, for all the parameters. Then for each sample of parameter values μ, d, β and x_0, the ODE (4.1) is solved and the solutions saved (HIV population from year 2001 to 2011). Finally for each year (2003 to 2011), simulated observations are generated with the ODE solutions as mean and sample values of σ as standard

deviation. These simulated observations for all 1000 samples are plotted in Figure 4.7, compared with the true observation from South Africa (Table A.1).

The mean of simulated observations from year 2003 to 2011 is

$$10^6 \times [4.0344, 4.1087, 4.1843, 4.2610, 4.3384, 4.4161, 4.4938, 4.5709, 4.6471],$$

while the true observations from year 2003 to 2011 is

$$10^6 \times [4.1198, 4.0816, 4.0408, 4.1815, 4.1985, 4.2632, 4.3349, 4.4582, 4.5106].$$

Not only the simulated observations contains the true observations from data, the mean of the simulates observations are very close to the true observations. This validates that the estimations are consistent with the true observations.

I think the Bayesian parameter estimation provides more reliable information for the parameters than the least square fitting (or similar approaches) does, as it provides distributions for each parameter, and it takes the observation error in HIV population into account. The posterior distributions give possible regions for parameter values such as 95% credibility intervals, also the the probability for a parameter to equal some specific value.

Chapter 5

CONCLUSION AND FUTURE WORK

In Chapter 2, the study was focused on six different effectiveness indicators for HIV interventions. Four of them are dimensionless (P_I, aI_I, F_I, $\hat{F}_I$) and two depend on the population size (C_I, $\hat{C}_I$). These indicators were defined for the HIV model with PrEP interventions. Then I studied their dynamics at the beginning of the interventions, and their asymptotic behaviors when the ODE system approaches the steady state. I mainly looked at how these indicators depend on the parameters. The parameters can be divided into two groups, the ones that are related to HIV infection (b_a, n, μ, d); the ones that are related to the epidemic (P), or effectiveness and implementation of PrEP (α_s, α_i, k_1, θ).

To intuitively show the dependence of the indicators on the parameters, by the end of this chapter, some results are presented from sensitivity analysis. I chose 10 years to illustrate the short term dynamics at the beginning of the intervention, and 100 years to illustrate the long time dynamics approaching the steady state. Despite all the discrepancies one can see, I also observed the similarity between indicators. Among the intervention parameters, PrEP coverage (k) and PrEP efficacies per act (α_s and α_i) express strong positive correlation with all the indicators in a short term. It remains significant in a long term for all outcomes. This confirms that PrEP coverage and protection level are critical to the intervention success regardless which qualitative metric is used. The prevalence (P_I) and the annual incidence (aI_I) indicators express almost the same sensitivity to all parameters. Therefore they should have consistent projections when evaluating the impact of the intervention.

Minimal efficacy thresholds of 20% or higher are often included in the design of RCTs and similar levels of effectiveness is expected when interventions are modeled at population level (Dimitrov *et al.* (2013); Grant *et al.* (2010)). Other studies imply that 50% biological efficacy is needed to guarantee significant public-health impact. It has been shown that the widely used evaluation metrics may disagree over practical intervals of time (Figure 2.5). A reduction in HIV incidence at pre-specified levels seems most realistic as an intervention goal but it is not easy to be estimated in the population. In contrast, a reduction in HIV prevalence is easier to be recorded but more difficult to be achieved in a short term. The reduction in the number of new HIV infections, which is the most popular public-health metric, projects strong PrEP effectiveness initially but grows slower than the other indicators over time. Public-health officials who consider PrEP to be integrated in HIV prevention programs should base their decision on a complex of quantitative metrics.

In Chapter 3, analytic study and bifurcation simulations have been done on models with 3 different entrance rates: constant recruitment, linear recruitment, and logistic recruitment. Analytic results haven been summarized for models without intervention in Section 3.1, Table 3.2. For models with interventions, similar summary based on both analytic study and bifurcation simulations is presented in Section 3.2, Figure 3.3.

Note that the trajectories of the models with constant and logistic recruitment are bounded while the linear recruitment allows for unbounded solutions. As the HIV risk (β) increases, the model with constant recruitment switches from infection-free to endemic steady state while the model with logistic recruitment equilibrium goes from infection-free to endemic equilibrium and further to population extinction. The trajectories of the linear model follow a similar pattern as the logistic recruitment model: from unbounded infection-free through unbounded endemic state to popu-

lation extinction. However, the transition (bifurcation) points where the behavior changes occur are different for the three models. As a result the projected long-term prevalence with each of the three models differ for epidemic conditions with R_0 greater than one. What in common for all the three models is when $R_0 < 1$, the status of the population approaches infection free; when $R_0 > 1$, the status of the population approaches coexistence or extinction.

I also looked at influence of different recruitment on the PrEP effectiveness. As a result the simulations of the HIV epidemics with the three models over 70 years show large discrepancies in population size and epidemic distribution under identical initial conditions and forces of infection. The projected HIV prevalence varies from 25% when constant recruitment is assumed and 42% when linear recruitment is assumed. In addition, significant difference in the reduction in HIV prevalence and incidence (almost 20%) is predicted when 50% effective PrEP is used by 20% of the population. Over the entire simulated period, linear recruitment provides the most optimistic estimates of the PrEP effectiveness in terms of prevalence reduction while constant recruitment predicts larger fraction of infections prevented.

It can be argued that regardless of the differences in the dynamic behavior all three models agree in their effectiveness projections over 20-30 years which is the usual period over which the intervention are evaluated. However, often the models are run for extended periods in order to simulate "mature" epidemics and the intervention is introduced afterwards. The key message of this analysis is that the way recruitment is incorporated in the models impacts the HIV epidemic and may have significant effect on the projected effectiveness of different HIV intervention in a short and long term. Demographic data, including statistics on births and age-specific mortality, should be used to inform the modeling mechanisms before HIV prevention is considered.

The models I studied are based on many simplified assumptions, presented in Section 2.1.1. There are many more practical models in literature considering more complicated models when age structure, population heterogeneity, drug resistance, and more are taken into account. However, to illustrate how different evaluation indicators and recruitment rates can affect the modeling results, it is better to start with the simplified model. The simplicity of the model makes these differences more significant. Modelers can refer to this and be aware of the possible different choices regarding the intervention evaluation methods and entrance rates. These different choices probably will also affect more complicated models at various levels.

In the previous chapters, the parameter values (β, μ, d) are either chosen from literature or obtained from fit model to data using an approach similar to least square fitting. In Chapter 4, to obtain more comprehensive information on parameter values, Bayesian inference is applied to fit the HIV model without intervention to South Africa HIV data (Table A.1). Bayesian parameter estimation eventually provide a distribution for each parameter. Evaluating the posterior distribution often requires the integration of high-dimensional functions, which is usually difficult to calculate numerically. However, the Markov Chain Monte Carlo (MCMC) method using the Metropolis-Hastings algorithm can be applied to approximate the posterior distribution, which is not difficult to implement. The convergence of the Markov chains is believed in this study. These chains provided good approximations for the posterior distributions of the key parameters (β, μ, d). The mean, standard deviation, and 95% confidence interval for the posterior distribution of each parameter can be easily summarized.

The little difference between prior and posterior distributions for μ and d implies that the data used is not very informative. Therefore the posterior distributions are mainly determined by the information provided by the prior distributions. Notice

that $\beta = 1 - (1 - b_a)^n$ is a composite parameter, which can not be measured directly. Information on b_a and n can both be found in literature (Boily *et al.* (2009),Wawer *et al.* (2005)). In this study, a posterior distribution for β (mean = 0.1599 and standard deviation = 0.02515 from Table 4.7) is obtained, which can be a supplement to the previous HIV studies.

Bayesian inference can also be applied in model selection (Gregory (2005); Raftery (1995); Toni *et al.* (2009)). If more data is available, the three models with different entrance rates may be differentiable.

"Mid-year population estimates, 2011", Statistics South Africa URL `http://www.statssa.gov.za/publications/P0302/P03022011.pdf` (2012).

"Effective sample size due to autocorrelation", `http://www.bayesian-inference.com/softwaredoc/ESS`, accessed: 2014-11-03 (2014).

Abbas, U. L., R. M. Anderson and J. W. Mellors, "Potential impact of antiretroviral chemoprophylaxis on hiv-1 transmission in resource-limited settings", PLoS One **2**, 9, e875 (2007).

Alsallaq, R. A., J. M. Baeten, C. L. Celum, J. P. Hughes, L. J. Abu-Raddad, R. V. Barnabas and T. B. Hallett, "Understanding the potential impact of a combination hiv prevention intervention in a hyper-endemic community", PloS one **8**, 1, e54575 (2013).

Alsallaq, R. A., B. Cash, H. A. Weiss, I. M. Longini Jr, S. B. Omer, M. J. Wawer, R. H. Gray and L. J. Abu-Raddad, "Quantitative assessment of the role of male circumcision in hiv epidemiology at the population level", Epidemics **1**, 3, 139–152 (2009).

Andersson, K. M., D. K. Owens and A. D. Paltiel, "Scaling up circumcision programs in southern africa: the potential impact of gender disparities and changes in condom use behaviors on heterosexual hiv transmission", AIDS and Behavior **15**, 5, 938–948 (2011).

Andrews, J. R., R. Wood, L.-G. Bekker, K. Middelkoop and R. P. Walensky, "Projecting the benefits of antiretroviral therapy for hiv prevention: the impact of population mobility and linkage to care", Journal of Infectious Diseases **206**, 4, 543–551 (2012).

Babiker, A., S. Darby, D. De Angelis, D. Kwart, K. Porter, V. Beral, J. Darbyshire, N. Day, N. Gill, R. Coutinho *et al.*, "Time from hiv-1 seroconversion to aids and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis", Lancet **355**, 9210, 1131–1137 (2000).

Baggaley, R. F., G. P. Garnett and N. M. Ferguson, "Modelling the impact of antiretroviral use in resource-poor settings", PLoS Medicine **3**, 4, e124 (2006).

Blower, S. and H. Dowlatabadi, "Sensitivity and uncertainty analysis of complex models of disease transmission: an hiv model, as an example", International Statistical Review/Revue Internationale de Statistique pp. 229–243 (1994).

Boily, M.-C., R. F. Baggaley, L. Wang, B. Masse, R. G. White, R. J. Hayes and M. Alary, "Heterosexual risk of hiv-1 infection per sexual act: systematic review and meta-analysis of observational studies", The Lancet infectious diseases **9**, 2, 118–129 (2009).

Boily, M.-C., D. Dimitrov, S. S. A. Karim and B. Mâsse, "The future role of rectal and vaginal microbicides to prevent hiv infection in heterosexual populations: implications for product development and prevention", Sexually transmitted infections **87**, 7, 646–653 (2011).

Breban, R., I. McGowan, C. M. Topaz, E. Schwartz, P. Anton and S. Blower, "Modeling the potential impact of rectal microbicides to reduce hiv transmission in bathhouses", Mathematical Biosciences and Engineering **3**, 3, 459–466 (2006).

Cox, A. P., A. M. Foss, L. A. Shafer, R. N. Nsubuga, P. Vickerman, R. J. Hayes, C. Watts and R. G. White, "Attaining realistic and substantial reductions in hiv incidence: model projections of combining microbicide and male circumcision interventions in rural uganda", Sexually transmitted infections **87**, 7, 635–639 (2011).

Cremin, I., R. Alsallaq, M. Dybul, P. Piot, G. Garnett and T. B. Hallett, "The new role of antiretrovirals in combination hiv prevention: a mathematical modelling analysis", Aids **27**, 3, 447–458 (2013).

Desai, K., S. L. Sansom, M. L. Ackers, S. R. Stewart, H. I. Hall, D. J. Hu, R. Sanders, C. R. Scotton, S. Soorapanth, M.-C. Boily, G. P. Garnett and P. D. McElroy, "Modeling the impact of hiv chemoprophylaxis strategies among men who have sex with men in the united states: Hiv infections prevented and cost-effectiveness", Aids **22**, 14, 1829–1839 (2008).

Dimitrov, D., M. Boily, B. Mâsse and E. Brown, "Impact of pill sharing on drug resistance due to a wide-scale oral prep intervention in generalized epidemics", J AIDS Clinic Res S **5**, 2 (2012).

Dimitrov, D. T., M.-C. Boily, R. F. Baggaley and B. Masse, "Modeling the gender-specific impact of vaginal microbicides on hiv transmission", Journal of theoretical biology **288**, 9–20 (2011).

Dimitrov, D. T., B. Masse and M.-C. Boily, "Who will benefit from a wide-scale introduction of vaginal microbicides in developing countries?", Statistical Communications in Infectious Diseases **2**, 1 (2010).

Dimitrov, D. T., B. R. Mâsse and M.-C. Boily, "Beating the placebo in hiv prevention efficacy trials: the role of the minimal efficacy bound", JAIDS Journal of Acquired Immune Deficiency Syndromes **62**, 1, 95–101 (2013).

Dushoff, J., A. Patocs and C.-F. Shi, "Modeling the population-level effects of male circumcision as an hiv-preventive measure: A gendered perspective", PloS one **6**, 12, e28608 (2011).

Edejer, T. T.-T., *Making choices in health: WHO guide to cost effectiveness analysis* (World Health Organization, 2003).

Granich, R. M., 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", The Lancet **373**, 9657, 48–57 (2009).

Grant, R. M., J. R. Lama, P. L. Anderson, V. McMahan, A. Y. Liu, L. Vargas, P. Goicochea, M. Casapía, J. V. Guanira-Carranza, M. E. Ramirez-Cardich *et al.*, "Preexposure chemoprophylaxis for hiv prevention in men who have sex with men", New England Journal of Medicine **363**, 27, 2587–2599 (2010).

Gregory, P., *Bayesian logical data analysis for the physical sciences*, vol. 10 (Cambridge University Press Cambridge, UK, 2005).

Hallett, T. B., K. Singh, J. A. Smith, R. G. White, L. J. Abu-Raddad and G. P. Garnett, "Understanding the impact of male circumcision interventions on the spread of hiv in southern africa", PLoS One **3**, 5, e2212 (2008).

Hallett, T., S. Dube, I. Cremin, B. Lopman, A. Mahomva, G. Ncube, O. Mugurungi, S. Gregson and G. Garnett, "The role of testing and counselling for hiv prevention and care in the era of scaling-up antiretroviral therapy", Epidemics **1**, 2, 77–82 (2009).

Hwang, T.-W. and Y. Kuang, "Deterministic extinction effect of parasites on host populations", Journal of mathematical biology **46**, 1, 17–30 (2003).

Hyman, J. M. and J. Li, "An intuitive formulation for the reproductive number for the spread of diseases in heterogeneous populations", Mathematical biosciences **167**, 1, 65–86 (2000).

Hyman, J. M. and J. Li, "Differential susceptibility epidemic models", Journal of mathematical biology **50**, 6, 626–644 (2005a).

Hyman, J. M. and J. Li, "The reproductive number for an hiv model with differential infectivity and staged progression", Linear algebra and its applications **398**, 101–116 (2005b).

Hyman, J. M. and J. Li, "Differential susceptibility and infectivity epidemic models", Mathematical Biosciences and Engineering **3**, 1, 89 (2006).

Juusola, J. L., M. L. Brandeau, D. K. Owens and E. Bendavid, "The cost-effectiveness of preexposure prophylaxis for hiv prevention in the united states in men who have sex with men", Annals of internal medicine **156**, 8, 541–550 (2012).

Kalichman, S. C., L. Simbayi, D. Cain and S. Jooste, "Heterosexual anal intercourse among community and clinical settings in cape town, south africa", Sexually transmitted infections **85**, 6, 411–415 (2009).

Karim, Q. A., S. S. A. Karim, J. A. Frohlich, A. C. Grobler, C. Baxter, L. E. Mansoor, A. B. Kharsany, S. Sibeko, K. P. Mlisana, Z. Omar *et al.*, "Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of hiv infection in women", science **329**, 5996, 1168–1174 (2010).

Karmon, E., M. Potts and W. M. Getz, "Microbicides and hiv: Help or hindrance?", Journal of Acquired Immune Deficiency Syndromes **34**, 1, 71–75 (2003).

Kass, R. E., B. P. Carlin, A. Gelman and R. M. Neal, "Markov chain monte carlo in practice: A roundtable discussion", The American Statistician **52**, 2, 93–100 (1998).

Law, M. G., G. Prestage, A. Grulich, P. Van de Ven and S. Kippax, "Modelling the effect of combination antiretroviral treatments on hiv incidence", Aids **15**, 10, 1287–1294 (2001).

Li, J., R. Gordon, J. Heffernan *et al.*, "Can we spend our way out of the aids epidemic? a world halting aids model", BMC Public Health **9**, Suppl 1, S15 (2009).

Lima, V. D., K. Johnston, R. S. Hogg, A. R. Levy, P. R. Harrigan, A. Anema and J. S. Montaner, "Expanded access to highly active antiretroviral therapy: a potentially powerful strategy to curb the growth of the hiv epidemic", Journal of Infectious Diseases **198**, 1, 59–67 (2008).

Londish, G. J., D. J. Templeton, D. G. Regan, J. M. Kaldor and J. M. Murray, "Minimal impact of circumcision on hiv acquisition in men who have sex with men", Sexual Health **7**, 4, 463–470 (2010).

Long, E. F., M. L. Brandeau and D. K. Owens, "Potential population health outcomes and expenditures of hiv vaccination strategies in the united states", Vaccine **27**, 39, 5402–5410 (2009).

Morgan, D., C. Mahe, B. Mayanja, J. M. Okongo, R. Lubega and J. A. Whitworth, "Hiv-1 infection in rural africa: is there a difference in median time to aids and survival compared with that in industrialized countries?", Aids **16**, 4, 597–603 (2002).

Nagelkerke, N. J., S. Moses, S. J. de Vlas and R. C. Bailey, "Modelling the public health impact of male circumcision for hiv prevention in high prevalence areas in africa", BMC infectious diseases **7**, 1, 16 (2007).

Nichols, B. E., C. A. Boucher, J. H. van Dijk, P. E. Thuma, J. L. Nouwen, R. Baltussen, J. van de Wijgert, P. M. Sloot and D. A. van de Vijver, "Cost-effectiveness of pre-exposure prophylaxis (prep) in preventing hiv-1 infections in rural zambia: A modeling study", PloS one **8**, 3, e59549 (2013).

Podder, C., O. Sharomi, A. Gumel and S. Moses, "To cut or not to cut: a modeling approach for assessing the role of male circumcision in hiv control", Bulletin of mathematical biology **69**, 8, 2447–2466 (2007).

Porter, K. and B. Zaba, "The empirical evidence for the impact of hiv on adult mortality in the developing world: data from serological studies", Aids **18**, S9–S17 (2004).

Pretorius, C., J. Stover, L. Bollinger, N. Bacaër and B. Williams, "Evaluating the cost-effectiveness of pre-exposure prophylaxis (prep) and its impact on hiv-1 transmission in south africa", PLoS One **5**, 11, e13646 (2010).

Raftery, A. E., "Bayesian model selection in social research", Sociological methodology **25**, 111–164 (1995).

Sorensen, S. W., S. L. Sansom, J. T. Brooks, G. Marks, E. M. Begier, K. Buchacz, E. A. DiNenno, J. H. Mermin and P. H. Kilmarx, "A mathematical model of comprehensive test-and-treat services and hiv incidence among men who have sex with men in the united states", PloS one **7**, 2, e29098 (2012).

Supervie, V., M. Barrett, J. S. Kahn, G. Musuka, T. L. Moeti, L. Busang and S. Blower, "Modeling dynamic interactions between pre-exposure prophylaxis interventions & treatment programs: predicting hiv transmission & resistance", Scientific reports **1** (2011).

Supervie, V., J. G. García-Lerma, W. Heneine and S. Blower, "Hiv, transmitted drug resistance, and the paradox of preexposure prophylaxis", Proceedings of the National Academy of Sciences **107**, 27, 12381–12386 (2010).

Toni, T., D. Welch, N. Strelkowa, A. Ipsen and M. P. Stumpf, "Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems", Journal of the Royal Society Interface **6**, 31, 187–202 (2009).

UNAIDS, *AIDS epidemic update: December 2009* (WHO Regional Office Europe, 2009).

UNAIDS, "Global report: Unaids report on the global aids epidemic 2013.", (2013).

Vickerman, P., F. Terris-Prestholt, S. Delany, L. Kumaranayake, H. Rees and C. Watts, "Are targeted hiv prevention activities cost-effective in high prevalence settings? results from a sexually transmitted infection treatment project for sex workers in johannesburg, south africa.", Sexually Transmitted Diseases **30**, Suppl 10, S122–S132 (2006).

Vissers, D. C., H. A. Voeten, N. J. Nagelkerke, J. D. F. Habbema and S. J. de Vlas, "The impact of pre-exposure prophylaxis (prep) on hiv epidemics in africa and india: a simulation study", PLoS One **3**, 5, e2077 (2008).

Vyshemirsky, V. and M. A. Girolami, "Bayesian ranking of biochemical system models", Bioinformatics **24**, 6, 833–839 (2008).

Wagner, B. G. and S. Blower, "Universal access to hiv treatment versus universal test and treat: transmission, drug resistance & treatment costs", PloS one **7**, 9, e41212 (2012).

Wawer, M. J., R. H. Gray, N. K. Sewankambo, D. Serwadda, X. Li, O. Laeyendecker, N. Kiwanuka, G. Kigozi, M. Kiddugavu, T. Lutalo *et al.*, "Rates of hiv-1 transmission per coital act, by stage of hiv-1 infection, in rakai, uganda", Journal of Infectious Diseases **191**, 9, 1403–1409 (2005).

Williams, B. G., 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**, 7, e262 (2006).

Wilson, D. P., P. M. Coplan, M. A. Wainberg and S. M. Blower, "The paradoxical effects of using antiretroviral-based microbicides to control hiv epidemics", Proceedings of the National Academy of Sciences **105**, 28, 9835–9840 (2008).

Zhao, Y., D. T. Dimitrov, H. Liu and Y. Kuang, "Mathematical insights in evaluating state dependent effectiveness of hiv prevention interventions", Bulletin of mathematical biology pp. 1–27 (2013).

APPENDIX A

DEMOGRAPHIC DATA FROM SOUTH AFRICA

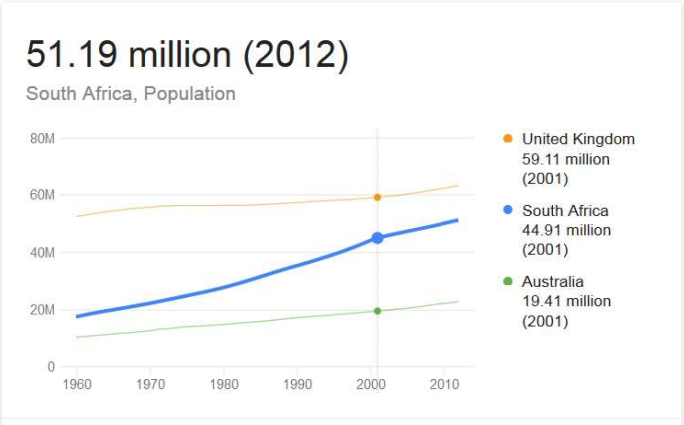


Figure A.1: Initial guess for capacity K is approximated by $2\times$(South Africa population at year 2001), because of the turning point observed at year 2001.

Table A.1: Age structured population in South Africa from year 2001 to 2011. Data adapted from Statistics South Africa Archive `http://www.statssa.gov.za/publications/statspastfuture.asp?PPN=P0302`, except age structured information could not be found for year 2002.

Age\Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
15-19	4981721		5263274	4923992	4898100	4938000	4975900	5152700	5214300	5226200	5175448
20-24	4294523		4392357	4679210	4621200	4653800	4675200	4783700	4920900	5018500	4900375
25-29	3934939		4100416	4291589	4211100	4271100	4335600	4367400	4423500	4518800	4598176
30-34	3340901		3422110	3696445	3762000	3841500	3863900	3913500	3888300	4035700	4040751
35-39	3071770		3216513	2851312	2780200	2842100	2972400	3147200	3282300	3465200	3600167
40-44	2619465		2794291	2538649	2483200	2428200	2400300	2389900	2443200	2524200	2612932
45-49	2087380		2241976	2213991	2187200	2215000	2222300	2240500	2260000	2230600	2244582
Total (15-49)	24330699	0	25430937	25195188	24943000	25189700	25445600	25994900	26432500	27019200	27172431
Prevalence (15-49)	0.16		0.162	0.162	0.162	0.166	0.165	0.164	0.164	0.165	0.166
Total HIV (15-49)	3892911.84	0	4119812	4081620	4040766	4181490	4198524	4263164	4334930	4458168	4510624
Total SUS (15-49)	20437787.2	0	21311125	21113568	20902234	21008210	21247076	21731736	22097570	22561032	22661807

APPENDIX B

DEMOGRAPHIC ENTRENCH RATES FROM LITERATURE

Table B.1: Population assumptions on HIV models from literature.

paper	population	recruitment	departure (all linear)	migration
Baggaley *et al.* (2006)	heterosexual(15-49 years)	constant	aging; death due to HIV; dropout due to side effects(linear);	No
Andrews *et al.* (2012)	heterosexual(15-49 years)	linear	background mortality; death due to HIV;	immigration(with specific HIV prevalence); emigration(with specific HIV prevalence);
Hallett *et al.* (2009)	heterosexual(15 years and older) and vertical	delayed linear	background mortality; death due to HIV(linear);	No
Karmon *et al.* (2003)	heterosexual(all ages)	constant	death due to HIV; emigration;	immigration as recruitment
Alsallaq *et al.* (2009)	heterosexual(sexually active)	linear	aging; death due to HIV;	No
Alsallaq *et al.* (2013)	heterosexual(sexually active)	linear	background mortality; death due to HIV;	No
Andersson *et al.* (2011)	heterosexual(17 years and older)	constant	background mortality; death due to HIV	No
Dushoff *et al.* (2011)	heterosexual(general)	linear	No	No
Granich *et al.* (2009)	heterosexual(15 years and older)	linear	background mortality; death due to HIV;	No
Hallett *et al.* (2008)	heterosexual(sexually active)	linear	aging; death due to HIV	No
Nagelkerke *et al.* (2007)	heterosexual(general)	linear	background mortality; death due to HIV;	No
Podder *et al.* (2007)	heterosexual(sexually active)	constant	background mortality; death due to HIV;	No

Table B.2: Continued from Table. B.1. MSM: men who have sex with men.

paper	population	recruitment	departure (all linear)	migration
Li *et al.* (2009)	general	constant	background mortality; death due to HIV;	immigration(linear); emigration(linear);
Sorensen *et al.* (2012)	MSM(sexually active)	constant	aging; background mortality; death due to HIV; drop out of care;	immigration as recruitment
Wagner and Blower (2012)	heterosexual(sexually active)	constant	background mortality; death due to HIV; drop out of treatment;	No
Nichols *et al.* (2013)	heterosexual	constant	background mortality; death due to HIV; drop out of treatment;	No
Abbas *et al.* (2007)	heterosexual(sexually active)	linear	aging; background mortality; death due to HIV;	No
Zhao *et al.* (2013)	heterosexual(sexually active)	constant	aging	No
Breban *et al.* (2006)	MSM(general)	constant	bathroom visit	No
Cox *et al.* (2011)	heterosexual(general)	constant	death, aging, emigration all in one	emigration(all in one, linear)
Cremin *et al.* (2013)	heterosexual(sexually active)	linear	death due to HIV; background mortality	No
Desai *et al.* (2008)	MSM(13 years and older)	constant	background mortality; death due to HIV;	No
Juusola *et al.* (2012)	MSM(13-64 years)	linear	aging; background mortality; death due to HIV;	No
Dimitrov *et al.* (2010)	heterosexual(sexually active)	constant	aging	No

Table B.3: Continued from Table. B.2. IDU: injection drug users.

paper	population	recruitment	departure (all linear)	migration
Dimitrov *et al.* (2011)	heterosexual(sexually active)	constant	aging; death due to HIV;	No
Dimitrov *et al.* (2012)	heterosexual(sexually active)	constant	aging	No
Supervie *et al.* (2010)	MSM(sexually active)	constant	No	No
Supervie *et al.* (2011)	heterosexual(sexually active)	constant	No	No
Vickerman *et al.* (2006)	heterosexual(sex workers)	constant	move away	No
Williams *et al.* (2006)	heterosexual(sexually active)	linear(fixed total)	death rate due to HIV	No
Wilson *et al.* (2008)	heterosexual(sexually active)	constant	background departure; death due to HIV;	No
Long *et al.* (2009)	heterosexual and homosexual(sexually active)	linear	mortality and maturation rate;	No
Londish *et al.* (2010)	MSM(sexually active)	constant	combined death and exit rates	emigration as exit rate
Lima *et al.* (2008)	MSM(and IDU)	constant	mortality rate (not) due to HIV	No
Law *et al.* (2001)	MSM(sexually active)	constant	aging;death rate due to HIV;	No

APPENDIX C

PROOFS

Proof for Proposition 3.1.1.1:

Proof. Positivity and boundedness of solutions can be easily proved. Then periodic solutions can be excluded by Dulacs Criteria: $\frac{\partial}{\partial S}\frac{P}{SI} + \frac{\partial}{\partial I}\frac{Q}{SI} = -\frac{\Lambda}{S^2 I} < 0$.

(3.1) has two possible steady states $E_0 = (\frac{1}{\mu}\Lambda, 0)$ and $E^* = (\frac{1}{\beta-d}\Lambda, \frac{1}{\beta-d}(R_0-1)\Lambda)$, with notation $R_0 = \frac{\beta}{\mu+d}$.

Eigenvalues $\lambda_1 = -\mu < 0$ and $\lambda_2 = (\mu+d)(R_0-1)$ for E_0 imply that E_0 is stable when $R_0 < 1$ and unstable when $R_0 > 1$.

Notice that E^* (positive steady state) exists if and only if $R_0 > 1$. Then corresponding eigenvalues $\lambda_1 + \lambda_2 = (\mu+d)(1 - R_0 - \frac{\mu}{\mu+d}) < 0$ and $\lambda_1 \times \lambda_2 = (\mu+d)[\beta(R_0-1)^2 + \mu R_0(R_0-1)] > 0$ imply that when exists E^* is stable.

Finally by Poincaré-Bendixson theorem, about global stability we have the following results:

- when $R_0 < 1$, E_0 is globally stable and E^* does not exist;
- when $R_0 > 1$, E^* is globally stable and E_0 is unstable.

□

Proof for Proposition 3.1.2.1:

Proof. The positivity of the solutions can be easily proved. Then $\frac{dN}{dt} = \frac{dS}{dt} + \frac{dI}{dt} = rN - \mu S - (\mu+d)I = (r-\mu)N - dI \leq (r-\mu)N$ implies that $N(t) \to 0$ extinction if $r < \mu$. Next we study the cases when $r > \mu$. Further if we assume $\beta > d$ and $\mu + d > r$, by Proposition 3.1.2.2, we have the following results:

- when $\beta < \mu + d$, $(S(t), I(t)) \to (\infty, 0)$ unbounded infection free.
- when $\mu + d < \beta < \frac{(\mu+d)d}{\mu+d-r}$, $(S(t), I(t)) \to (\infty, \infty)$ unbounded endemic.
- when $\beta > \frac{(\mu+d)d}{\mu+d-r}$, $(S(t), I(t)) \to (0, 0)$ extinction.

□

Proof for Proposition 3.1.2.2:

Proof. Assume that $r > \mu$, $\beta > d$ and $\mu + d > r$. If $s(0) \in [0, 1)$, then $\lim_{t\to\infty} s(t) = 1$ ($\lim_{t\to\infty} i(t) = 0$) if $\frac{r}{\beta-d} \geq 1$; $\lim_{t\to\infty} s(t) = \frac{r}{\beta-d}$ ($\lim_{t\to\infty} i(t) = 1 - \frac{r}{\beta-d}$) if $\frac{r}{\beta-d} < 1$. Further, we can study $N(t)$ from $\frac{dN}{dt}$, and $I(t)$ from $\frac{dI}{dt}$. In each case, for different parameter values, we have $\lim_{t\to\infty} N(t) = 0$ or $\lim_{t\to\infty} N(t) = \infty$; and $\lim_{t\to\infty} I(t) = 0$ or $\lim_{t\to\infty} I(t) = \infty$:

- when $\beta < \mu + d$, $(S(t), I(t)) \to (\infty, 0)$ with $(s(t), i(t), N(t)) \to (1, 0, \infty)$;
- when $\mu+d < \beta < r+d$, $(S(t), I(t)) \to (\infty, \infty)$ with $(s(t), i(t), N(t)) \to (1, 0, \infty)$;
- when $r + d < \beta < \frac{(\mu+d)d}{\mu+d-r}$, $(S(t), I(t)) \to (\infty, \infty)$ with $(s(t), i(t), N(t)) \to (\frac{r}{\beta-d}, 1 - \frac{r}{\beta-d}, \infty)$;

- when $\beta > \frac{(\mu+d)d}{\mu+d-r}$, $(S(t), I(t)) \to (0,0)$ with $(s(t), i(t), N(t)) \to (\frac{r}{\beta-d}, 1-\frac{r}{\beta-d}, 0)$.

□

Proof for Proposition 3.1.3.1:

Proof. Periodic solutions can be excluded by Dulac's Criteria: $\frac{\partial}{\partial S}(\frac{P}{SI}) + \frac{\partial}{\partial S}(\frac{Q}{SI}) = -\frac{r}{K}(\frac{1}{S}+\frac{1}{I}) - \frac{r}{S^2}(1-\frac{S+I}{K}) < 0$.
(3.1) has three possible steady states $E_{00} = (0,0)$, $E_{01} = (\frac{r-\mu}{r}K, 0)$ and $E^* = \frac{r-\mu-d(1-\frac{1}{R_0})}{r}K(\frac{1}{R_0}, \frac{R_0-1}{R_0})$, with notation $R_0 = \frac{\beta}{\mu+d}$.

Similar to Proposition 3.1.2.1, we can show that $N(t) \to 0$ if $r < \mu$, also E_{01} and E^* both do not exist. So next we will study the cases when $r > \mu$. Further we assume $\beta > d$ and $\mu + d > r$.

Eigenvalues $\lambda_1 = -(r-\mu) < 0$ and $\lambda_2 = \beta - (\mu+d)$ for E_{01} imply that when exists E_{01} is stable if $R_0 < 1$, and unstable if $R_0 > 1$.

Notice that E^* (positive steady state) exists if and only if $R_0 > 1$ and $r > \mu + d(1-\frac{1}{R_0})$ ($\Leftrightarrow R_0 r + d > \beta > \mu + d \Rightarrow R_0 r > \mu$). Eigenvalues $\lambda_1 + \lambda_2 = -[r-\mu-d(1-\frac{1}{R_0})] - (1-\frac{1}{R_0})(\beta-d) < 0$ and $\lambda_1 \cdot \lambda_2 = [r-\mu-d(1-\frac{1}{R_0})][\beta-(\mu+d)] > 0$ for E^* imply that E^* is stable when exists.

Thus,

- when $r < \mu$, E_{00} is (globally) stable, while E_{01} and E^* do not exist.
- When $r > \mu$(assume $\beta > d$ and $\mu + d > r$),
  - if $\beta < \mu + d (R_0 < 1)$, E_{01} is stable and E^* does not exist;
  - if $R_0 r + d > \beta > \mu + d (R_0 > 1)$, E^* exists and is stable and E_{01} is unstable;
  - if $\beta > R_0 r + d (R_0 > 1)$, E_{01} is unstable and E^* does not exist.

This can be simplified as: when $r > \mu$, $\beta > d$, and $\mu + d > r$,

- if $\beta < \mu + d$, E_{01} is stable and E^* does not exist;
- if $\mu + d < \beta < \frac{(\mu+d)d}{\mu+d-r}$, E^* exists and is stable and E_{01} is unstable;
- if $\beta > \frac{(\mu+d)d}{\mu+d-r}$, E_{01} is unstable and E^* does not exist.

Next we combine the local stability results for E_{00} from Proposition 3.1.3.2.

Finally by Poincaré-Bendixson theorem, about global stability we have the following conclusions(assume $r > \mu$, $\beta > d$, and $\mu + d > r$):

- when $\beta < \mu + d$, $(S(t), I(t)) \to E_{01} = (\frac{r-\mu}{r}K, 0)$ the infection free steady state, while E_{00} is unstable and E^* does not exist;
- when $\mu + d < \beta < \frac{(\mu+d)d}{\mu+d-r}$, $(S(t), I(t)) \to E^* = \frac{r-\mu-d(1-\frac{1}{R_0})}{r}K(\frac{1}{R_0}, \frac{R_0-1}{R_0})$ the endemic steady state, while E_{00} and E_{01} are unstable;

- when $\beta > \frac{(\mu+d)d}{\mu+d-r}$, $(S(t), I(t)) \to E_{00} = (0, 0)$ the extinction steady state, while E_{01} and E^* are unstable.

□

Proof for Proposition 3.1.3.2:

Proof. There are four possible steady states for (3.2): $E_1 = (1, 0)$, $E_2 = (\frac{r}{\beta-d}, 0)$, $E_3 = (1, \frac{r-\mu}{r}K)$, and $E_4 = (\frac{1}{R_0}, \frac{r-\mu-d(1-\frac{1}{R_0})}{r}K)$. Notice that E_3 and E_4 are equivalent with E_{01} and E^* respectively, while E_1 and E_2 are both corresponding to E_{00}. So we only study E_1 and E_2 here, and we assume $r > \mu$, $\beta > d$, and $\mu + d > r$.

Eigenvalues $\lambda_1 = \beta - d - r$ and $\lambda_2 = r - \mu > 0$ for E_1 imply that E_1 is unstable. Eigenvalues $\lambda_1 = d + r - \beta$ and $\lambda_2 = -\frac{\mu+d}{\beta-d}(\beta - d - R_0 r)$ for E_2 imply that E_2 is stable if $\beta > d + r$ and $\beta > d + R_0 r$. Thus, E_{00} is stable if $\beta > \max\{d + r, d + R_0 r\}$ with $(s(t), i(t), N(t)) \to (\frac{r}{\beta-d}, 1 - \frac{r}{\beta-d}, 0)$ and unstable if $\beta < \max\{d + r, d + R_0 r\}$.

Then,

- when $R_0 < 1 (\beta < \mu + d)$, E_{00} is unstable since $\beta < d + r$;
- when $R_0 > 1 (\beta > \mu + d)$, E_{00} is stable if $\beta > d + R_0 r$ and unstable if $\beta < d + R_0 r$.

This can be simplified to be:

- if $\beta < \frac{(\mu+d)d}{\mu+d-r}$, E_{00} is unstable;
- if $\beta > \frac{(\mu+d)d}{\mu+d-r}$, E_{00} is stable.

□

Proof for Proposition 3.2.1.1:

Proof. Positivity and boundedness of solutions can be easily proved. Then $\frac{dS^p}{dt} \leq k\Lambda - \mu S^p$ implies $\limsup_{t\to\infty} S^p \leq \frac{k\Lambda}{\mu}$ and $\frac{dS}{dt} \leq (1-k)\Lambda - \mu S$ implies $\limsup_{t\to\infty} S \leq \frac{(1-k)\Lambda}{\mu}$. Therefore $\frac{dN}{dt} = \frac{dS^p}{dt} + \frac{dS}{dt} + \frac{dI}{dt} = \Lambda - \mu N - dI \geq \Lambda - (\mu + d)N$ implies $\liminf_{t\to\infty} N \geq \frac{\Lambda}{\mu+d}$. Then in long term $\frac{S}{N} \leq \frac{(1-k)(\mu+d)}{\mu}$, $\frac{S^p}{N} \leq \frac{k(\mu+d)}{\mu}$, and so $\frac{dI}{dt} \leq I\left[\beta\frac{(1-k)(\mu+d)}{\mu} + (1-\alpha_s)\beta\frac{k(\mu+d)}{\mu} - (\mu+d)\right] = I\frac{(\mu+d)^2}{\mu}\left[\beta\frac{1-k}{\mu+d} + (1-\alpha_s)\beta\frac{k}{\mu+d} - \frac{\mu}{\mu+d}\right] = I\frac{(\mu+d)^2}{\mu}\left(R_0 - \frac{\mu}{\mu+d}\right)$. Now if $R_0 < \frac{\mu}{\mu+d}$, then because of the positivity of the solution, we know $\lim_{t\to\infty} I = 0$. Then combine with equations in (3.3), we know $\lim_{t\to\infty} S^p = \frac{k\Lambda}{\mu}$ and $\lim_{t\to\infty} S = \frac{(1-k)\Lambda}{\mu}$. Thus, global stability of infection free steady state $E_0 = (\frac{k\Lambda}{\mu}, \frac{(1-k)\Lambda}{\mu}, 0)$ under condition $R_0 < \frac{\mu}{\mu+d}$ is proved.

For local stability of E_0, we look the corresponding eigenvalues $\lambda_1 = -\mu < 0$, $\lambda_2 = -\mu < 0$ and $\lambda_3 = (\mu + d)(R_0 - 1)$. Therefore, E_0 is stable when $R_0 < 1$ and unstable when $R_0 > 1$.

Now assume that $\beta > d$. For $F(I)$ over $[0, \frac{\Lambda}{d}]$, the following results hold: $F(0) = 0$, $F(\frac{\Lambda}{d}) = -\mu < 0$, $F'(0) = (\mu+d)(R_0-1)$, and $F''(I) = -\frac{2k(1-\alpha_s)\beta[(1-\alpha_s)\beta - d]\Lambda^2}{[\Lambda - dI + (1-\alpha_s)\beta I]^3}$ $-\frac{2(1-k)\beta(\beta-d)\Lambda^2}{[\Lambda - dI + \beta I]^3} = \frac{2k(1-\alpha_s)\beta[d-(1-\alpha_s)\beta]\Lambda^2}{\{\Lambda - [d-(1-\alpha_s)\beta]I\}^3} - \frac{2(1-k)\beta(\beta-d)\Lambda^2}{[\Lambda + (\beta-d)I]^3}$. Further if $(1-\alpha_s)\beta \geq d$ then $F''(I) < 0$ over$[0, \frac{\Lambda}{d}]$; else if $(1-\alpha_s)\beta < d$ then $F''(I)$ increases over $[0, \frac{\Lambda}{d}]$. Thus, F is either concave down, or concave up, or changes from concave down to concave up over $[0, \frac{\Lambda}{d}]$. Now if $R_0 > 1$, then $F(0) = 0$, $F'(0) > 0$ and $F(\frac{\Lambda}{d}) < 0$ implies a unique solution of $F(I) = 0$ over $[0, \frac{\Lambda}{d}]$, because of the concavity of $F(I)$ over $[0, \frac{\Lambda}{d}]$. Else if $R_0 < 1$, then $F(0) = 0$, $F'(0) < 0$ and $F(\frac{\Lambda}{d}) < 0$ implies no solution of $F(I) = 0$ over $[0, \frac{\Lambda}{d}]$, because of the concavity of $F(I)$ over $[0, \frac{\Lambda}{d}]$. Thus, assume that $\beta > d$, then we have a unique E^* if $R_0 > 1$ and no E^* if $R_0 < 1$.

Notice that Because of the complexity of the expressions, we will not express the positive steady state explicitly. Now assume that $R_0 = \frac{(1-\alpha_s k)\beta}{\mu+d} > 1$. The Jacobian matrix for a positive steady state can be expressed as:
$J = \begin{bmatrix} -(1-\alpha_s)\beta i^*(1-p^*) - \mu & (1-\alpha_s)\beta p^* i^* & -(1-\alpha_s)\beta p^*(1-i^*) \\ \beta s^* i^* & -\beta i^*(1-s^*) - \mu & -\beta s^*(1-i^*) \\ [(1-\alpha_s)\beta - (\mu+d)]i^* & [\beta - (\mu+d)]i^* & -(\mu+d)i^* \end{bmatrix}$, with $p^* = \frac{S^{p*}}{N^*}$, $s^* = \frac{S^*}{N^*}$, and $i^* = \frac{I^*}{N^*}$. With $P = \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, J is similar to

$$H = P^{-1}JP = \begin{bmatrix} -\mu & 0 & d \\ \beta s^* i^* & -\mu - \beta i^* & -\beta s^* \\ [(1-\alpha_s)\beta - (\mu+d)]i^* & \alpha_s \beta i^* & -(1-\alpha_s)\beta i^* \end{bmatrix}.$$

Then J and H share the same eigenvalues. Let λ be an arbitrary eigenvalue, then we can eventually get the characteristic equation

$$\frac{1}{d}(\lambda+\mu)[\lambda + (1-\alpha_s)\beta i^*] + \frac{\alpha_s \beta^2 i^* s^*}{d} \cdot \frac{\lambda + \mu + d i^*}{\lambda + \mu + \beta i^*} + [(1-\alpha_s)\beta - (\mu+d)]i^* = 0.$$

Now plug $\lambda = u + iv$ in the characteristic equation, and extract the imaginary part we obtain:

$$\frac{v}{d}[2u + \mu + (1-\alpha_s)\beta i^* + \frac{\alpha_s \beta^2 (\beta - d) i^{*2} s^*}{(u + \mu + \beta i^*)^2 + v^2}] = 0.$$

Notice that $\beta - d > 0$ since $R_0 > 1$, then $v \neq 0$ implies $u < 0$. Otherwise $v = 0$ implies $f(\lambda) = f(u) \triangleq \frac{1}{d}(u+\mu)[u + (1-\alpha_s)\beta i^*] + \frac{\alpha_s \beta^2 i^* s^*}{d} \cdot \frac{u + \mu + d i^*}{u + \mu + \beta i^*} + [(1-\alpha_s)\beta - (\mu+d)]i^*$ and $f(u)$ increases over $[0, \infty)$ since $\beta > d$. Then

$$\begin{aligned} f(0) &= \frac{1}{d}\mu(1-\alpha_s)\beta i^* + \frac{\alpha_s\beta^2 i^* s^*}{d}\cdot\frac{\mu+di^*}{\mu+\beta i^*} + [(1-\alpha_s)\beta-(\mu+d)]i^* \\ &> \frac{1}{d}\mu(1-\alpha_s)\beta i^* + \frac{\alpha_s\beta^2 i^*\frac{(1-k)d}{\beta}}{d}\cdot\frac{d}{\beta} + [(1-\alpha_s)\beta-(\mu+d)]i^* \\ &= \frac{1}{d}\mu(1-\alpha_s)\beta i^* + \alpha_s(1-k)di^* + (1-\alpha_s)\beta i^* - (\mu+d)i^* \\ &= \{\frac{\mu+d}{d}[(1-\alpha_s)\beta-d] + \alpha_s(1-k)d\}i^*. \end{aligned}$$

Notice that $s^* = \frac{S^*}{N^*} = \frac{\Lambda(1-k)}{\Lambda+(\beta-d)I^*} > \frac{\Lambda(1-k)}{\Lambda+(\beta-d)\frac{\Lambda}{d}} = \frac{(1-k)d}{\beta}$. I hope to show that $f(0) > 0$ and so $f(u) > 0$ over $[0,\infty)$, and further $f(u) = 0$ can only have negative solutions. But it seems I can only show this given extra condition $\frac{\mu+d}{d}[(1-\alpha_s)\beta-d]+\alpha_s(1-k)d > 0$.

I believe in $f(0) > 0$ without this extra condition, which needs to be further studied in future. □

Proof for Proposition 3.2.2.1:

Proof. The positivity of the solutions can be easily proved. Then similar to Proposition 3.1.2.1, we can show that the unique steady state $E = (0,0,0)$ is globally stable if $r < \mu$. Next we will study the cases when $r > \mu$.

From Proposition 3.2.2.2, we know when $\beta > d$, there is no periodic solution for the reduced system (3.4) , and (3.5), and therefore $E_1 = (k,0)$ is globally stable when $(1-k)\beta + k(1-\alpha_s)\beta < d+r$, while $E_2 = (p^*, i^*)$ is globally stable when $(1-k)\beta + k(1-\alpha_s)\beta > d+r$. Then by (3.6), the total population $N(t)$ either approaches 0 when $r-\mu-d\lim_{t\to\infty} i(t) < 0$, or blows up when $r-\mu-d\lim_{t\to\infty} i(t) > 0$. Similarly by (3.7), the infected population $I(t)$ either approaches 0 when $\beta(1-\lim_{t\to\infty} p(t) - \lim_{t\to\infty} i(t)) + (1-\alpha_s)\beta\lim_{t\to\infty} p(t) - (\mu+d) < 0$, or blows up when $\beta(1-\lim_{t\to\infty} p(t) - \lim_{t\to\infty} i(t)) + (1-\alpha_s)\beta\lim_{t\to\infty} p(t) - (\mu+d) > 0$.

Thus, periodic solutions for (3.3) do not exist when $\beta > d$. □

Proof for Proposition 3.2.2.2:

Proof. Positivity of solutions for the transformed system can be easily checked. Further,

$$\begin{aligned} \frac{dp}{dt}+\frac{di}{dt} &= kr - rp - \beta pi + \alpha_s pi + dpi + (\beta-d)i - ri - \alpha_s\beta pi - (\beta-d)i^2 \\ &= kr - r(p+i) - \beta pi - (\beta-d)pi + (\beta-d)i - (\beta-d)i^2 \\ &= kr - r(p+i) - \beta pi + (\beta-d)i[1-(p+i)]. \end{aligned}$$

Then at $p+i=1$, $\frac{d(p+i)}{dt} = -r + kr - \beta pi = -r(1-k) - \beta pi < 0$ implies that $(p+i)(t) \le 1$ for $t > 0$ given that $(p+i)(0) \le 1$.

The periodic solutions for the reduced system (therefore for the transformed system) can be excluded by Dulac's Criteria: $\frac{\partial}{\partial p}(\frac{X}{pi}) + \frac{\partial}{\partial i}(\frac{Y}{pi}) = -\frac{kr}{p^2 i} - \frac{\beta - d}{p} < 0$, when given $\beta - d > 0$.

For the reduced system, there are possibly several steady states: $E_1 = (k, 0)$(always exists), $E_2 = (p^*, i^*)$(positive steady state(s), existence depends on parameter values).

For E_1, we have eigenvalues $\lambda_1 = -r < 0$ and $\lambda_2 = (1 - \alpha_s k)\beta - (d + r)$. Then if $(1 - \alpha_s k)\beta < (d + r)$, E_1 is locally stable; if $(1 - \alpha_s k)\beta > (d + r)$, E_1 is unstable.

For E_2, we have $Ap^{*2} + Bp^* + C = 0$ and $i^* = 1 - \frac{\alpha_s \beta p^* + r}{\beta - d}$, with $A = \alpha_s \beta[(1 - \alpha_s)\beta - d]$, $B = -\{(\beta - d - r)[(1 - \alpha_s)\beta - d] + (\beta - d)r\}$, and $C = (\beta - d)kr$. If further $0 < p^* < 1$ and $0 < i^* < 1$, then (p^*, i^*) exists as a positive steady state. (There may be more than one positive steady state.)

Notice that $p^* \in (0, \frac{\beta - (d+r)}{\alpha_s \beta})$ and $i^* \in (0, \frac{\beta - (d+r)}{\beta - d})$. So we assume that $\beta > d + r$. Denote $F(p) = Ap^2 + Bp + C$, then we have the following results for $F(p)$ over $(0, \frac{\beta - (d+r)}{\alpha_s \beta})$: $F(0) = (\beta - d)kr > 0$, and $F(\frac{\beta - (d+r)}{\alpha_s \beta}) = -\frac{(\beta - d)r}{\alpha_s \beta}[(1 - k\alpha_s)\beta - (d + r)]$. Now if $(1 - k\alpha_s)\beta > d + r$, then $F(0) > 0$ and $F(\frac{\beta - (d+r)}{\alpha_s \beta}) < 0$ implies a unique solution of $F(p) = 0$ over $(0, \frac{\beta - (d+r)}{\alpha_s \beta})$, because $F(p)$ is a parabolic function.

Now look at the case when $(1 - k\alpha_s)\beta < d + r (\Rightarrow (1 - \alpha_s)\beta - d < r)$. If $(1 - \alpha_s)\beta \leq d$, then $F(0) > 0$ and $F(\frac{\beta - (d+r)}{\alpha_s \beta}) > 0$ implies no solution of $F(p) = 0$ over $(0, \frac{\beta - (d+r)}{\alpha_s \beta})$, because $F(p)$ is linear or concave down. If $(1 - \alpha_s)\beta > d$, then $F(p)$ is concave up and attains its minimum at $\hat{p} = \frac{\beta - (d+r)}{2\alpha_s \beta} + \frac{(\beta - d)r}{2\alpha_s \beta[(1 - \alpha_s)\beta - d]} > \frac{\beta - (d+r)}{2\alpha_s \beta} + \frac{(\beta - d)r}{2\alpha_s \beta r} > \frac{\beta - (d+r)}{\alpha_s \beta}$. Therefore if $(1 - \alpha_s)\beta > d$, then $F(0) > 0$ and $F(\frac{\beta - (d+r)}{\alpha_s \beta}) > 0$ implies no solution of $F(p) = 0$ over $(0, \frac{\beta - (d+r)}{\alpha_s \beta})$, because $F(p)$ is decreasing over $(0, \frac{\beta - (d+r)}{\alpha_s \beta})$.

Together if $\beta \geq d + r$, a unique solution of $F(p) = 0$ exists over $(0, \frac{\beta - (d+r)}{\alpha_s \beta})$ when $(1 - k\alpha_s)\beta > d + r$ and no solution exists over $(0, \frac{\beta - (d+r)}{\alpha_s \beta})$ when $(1 - k\alpha_s)\beta < d + r$.

Next, we can show that E_2 when exists, is stable. Notice that the existence of E_2 requires $(1 - \alpha_s k)\beta > (d + r)$ (when E_1 is unstable). We have

$$\begin{aligned} & kr - rp^* - [(1 - \alpha_s)\beta - d]p^* i^* = 0 \\ & \Rightarrow -r - [(1 - \alpha_s)\beta - d]i^* = -\frac{kr}{p^*}, \quad -[(1 - \alpha_s)\beta - d]p^* = -\frac{kr - rp^*}{i^*}, \end{aligned}$$

and

$$\beta - (d + r) - \alpha_s \beta p^* - (\beta - d)i^* = 0.$$

And we have the following Jacobian matrix for E_2:

$$J(E_2) = \begin{vmatrix} -r - [(1 - \alpha_s)\beta - d]i^* & -[(1 - \alpha_s)\beta - d]p^* \\ -\alpha_s \beta i^* & -(\beta - d)i^* + \beta - (d + r) - \alpha_s \beta p^* - (\beta - d)i^* \end{vmatrix}.$$

Then

$$J(E_2) = \begin{vmatrix} -\frac{kr}{p^*} & -\frac{kr - rp^*}{i^*} \\ -\alpha_s \beta i^* & -(\beta - d)i^* \end{vmatrix}.$$

The corresponding eigenvalues satisfy

$$\lambda_1 + \lambda_2 = -\frac{kr}{p^*} - (\beta - d)i^* < 0,$$

and

$$\lambda_1 \cdot \lambda_2 = \frac{kr}{p^*}(\beta - d)i^* - \alpha_s\beta(kr - rp^*).$$

Further

$$\begin{aligned}
\lambda_1 \cdot \lambda_2 &= \frac{kr}{p^*}(\beta - d)(1 - \frac{\alpha_s\beta p^* + r}{\beta - d}) - \alpha_s\beta(kr - rp^*) \\
&= \frac{r}{p^*}[k(\beta - d - \alpha_s\beta p^* - r) - \alpha_s\beta(k - p^*)p^*] \\
&= \frac{r}{p^*}[\alpha_s\beta p^{*2} - 2\alpha_s\beta k p^* + k(\beta - d - r)].
\end{aligned}$$

We next conclude $\lambda_1 \cdot \lambda_2 > 0$ from $f(p^*) = \alpha_s\beta p^{*2} - 2\alpha_s\beta k p^* + k(\beta - d - r) > 0$. We know $f(p^*)$ attains the minimum value $f(k)$ at $p^* = k$. Now it is sufficient to show that $f(k) > 0$. Since $(1 - \alpha_s k)\beta > (d + r) \Rightarrow \beta - d - r > \alpha_s k\beta$, then

$$\begin{aligned}
f(k) &= \alpha_s\beta k^2 - 2\alpha_s\beta k^2 + k(\beta - d - r) \\
&= k(\beta - d - r) - \alpha_s\beta k^2 \\
&> k\alpha_s k\beta - \alpha_s\beta k^2 \\
&= 0.
\end{aligned}$$

Thus, we have proved that E_2 when exists (require $(1-\alpha_s k)\beta > (d+r)$), is stable. Further since $(1 - \alpha_s k)\beta > (d + r)$ implies $\beta > d$ (no periodic solutions) and E_1 is unstable, then E_2 when exists is globally stable.

If $(1 - \alpha_s k)\beta < (d + r)$, then E_1 is stable and E_2 does not exist. Further $\beta > d$ implies no periodic solutions, therefore we conclude that E_1 is globally stable provided that $(1 - \alpha_s k)\beta < (d + r)$ and $\beta > d$. □

Proof for Proposition 3.2.3.1:

Proof. The invariance of the biologically region can be easily proved.

Then similar to Proposition 3.1.2.1, we can show that the extinction steady state E_{00} is globally stable if $r < \mu$.

From now on, we assume $r > \mu$. About the eigenvalues for E_{01}, we have $\lambda_1 + \lambda_2 = -r < 0$, $\lambda_1 \times \lambda_2 = \mu(r - \mu) > 0$ and $\lambda_3 = (\mu + d)(R_0 - 1)$. Therefore E_{01} is stable when $R_0 < 1$ and unstable when $R_0 > 1$.

From Proposition 3.2.3.2, we know when $r > \mu$, then extinction steady state is stable if $(1 - k)\beta + k(1 - \alpha_s)\beta > d + r$ and unstable if $(1 - k)\beta + k(1 - \alpha_s)\beta < d + r$; further when $\beta > \mu + d$, the positive steady state exists if $R_0 > 1$ and $i^* = \frac{I^*}{N^*} < \frac{r-\mu}{d}$ and does not exist if $R_0 < 1$ or $i^* = \frac{I^*}{N^*} > \frac{r-\mu}{d}$. □

Proof for Proposition 3.2.3.2:

Proof. For (3.8), there are possibly several steady states: $E_1 = (k, 0, 0)$(always exists), $E_2 = (p^*, i^*, 0)$(existence depends on parameter values), $E_3 = (k, 0, \frac{r-\mu}{r}K)$, and $E_4 = (p^*, i^*, N^*)$(existence depends on parameter values). Notice that E_3 is equivalent with $E_{01} = (k\frac{r-\mu}{\mu}K, (1-k)\frac{r-\mu}{\mu}K, 0)$ for (3.3), and E_4 is equivalent with positive steady state E^* for (3.3) if exists, while E_1 and E_2 are both corresponding to $E_{00} = (0, 0, 0)$ for (3.3). We will assume $r > \mu$.

For E_1, we have eigenvalues $\lambda_1 = -r < 0$, $\lambda_2 = (1-\alpha_s k)\beta - (d+r)$, and $\lambda_3 = r - \mu > 0$. So E_1 is unstable.

For E_2, we have $Ap^{*2} + Bp^* + C = 0$ and $i^* = 1 - \frac{\alpha_s\beta p^* + r}{\beta - d}$, with $A = \alpha_s\beta[(1-\alpha_s)\beta - d]$, $B = -\{(\beta - d - r)[(1-\alpha_s)\beta - d] + (\beta - d)r\}$, and $C = (\beta - d)kr$. If further $0 \le p^* \le 1$ and $0 \le i^* \le 1$, then $(p^*, i^*, 0)$ exists as a steady state. This is similar to the case $E_2 = (p^*, i^*)$ in Proposition 3.2.2.2(with linear entrance rate).

For E_4, we have $Ap^{*2} + Bp^* + C = 0$, $i^* = \frac{\beta - \mu - d - \alpha_s\beta p^*}{\beta}$, $N^* = \frac{r - \mu - di^*}{r}K$, with $A = \alpha_s(1-\alpha_s)\beta$, $B = -[(1-\alpha_s)(\beta - \mu - d) + \mu + k\alpha_s d]$, and $C = kd\frac{\beta - \mu - d}{\beta} + k\mu$. Therefore, positive steady states may exist but too complicated to be expressed explicitly.

Notice that $i^* \in (0, \frac{r-\mu}{d})$ and $p^* \in \left(\frac{\beta - (\mu+d) - \frac{r-\mu}{d}}{\alpha_s\beta}, \frac{\beta - (\mu+d)}{\alpha_s\beta}\right)$. So we assume that $\beta > \mu + d$ and $r > \mu$ so that E_4 may exist as an endemic steady state. Denote $F(p) = Ap^2 + Bp + C$, then we have the following results for $F(p)$ over $(0, \frac{\beta - (\mu+d)}{\alpha_s\beta})$: $F(0) = \frac{k(\mu+d)(\beta - d)}{\beta} > 0$, and $F(\frac{\beta - (\mu+d)}{\alpha_s\beta}) = -\frac{\mu(\mu+d)}{\alpha_s\beta}(R_0 - 1)$. Now if $R_0 > 1$, then $F(0) > 0$ and $F(\frac{\beta - (\mu+d)}{\alpha_s\beta}) < 0$ implies a unique solution of $F(p) = 0$ over $(0, \frac{\beta - (\mu+d)}{\alpha_s\beta})$, because $F(p)$ is a parabolic function. Further if the solution $p^* > \frac{\beta - (\mu+d) - \frac{r-\mu}{d}}{\alpha_s\beta}$, then there exists a unique E_2. Thus, if $\beta > \mu + d$ and $r > \mu$, then there is a unique E_2 when $R_0 > 1$ and $p^* > \frac{\beta - (\mu+d) - \frac{r-\mu}{d}}{\alpha_s\beta} (\Leftrightarrow i^* < \frac{r-\mu}{d})$.

Now look at the case when $R_0 < 1 (\Rightarrow \beta < \frac{\mu+d}{1-k\alpha_s})$. $F(p)$ is concave up and attains its minimum at $\hat{p} = \frac{\beta - (\mu+d)}{2\alpha_s\beta} + \frac{\mu + k\alpha_s d}{2\alpha_s(1-\alpha_s)\beta} > \frac{\beta - (\mu+d)}{\alpha_s\beta}$, since

$$\begin{aligned}
& \frac{\beta - (\mu+d)}{2\alpha_s\beta} + \frac{\mu + k\alpha_s d}{2\alpha_s(1-\alpha_s)\beta} - \frac{\beta - (\mu+d)}{\alpha_s\beta} \\
= & \frac{\mu + k\alpha_s d - (1-\alpha_s)[\beta - (\mu+d)]}{2\alpha_s(1-\alpha_s)\beta} \\
> & \frac{\mu + k\alpha_s d - (1-\alpha_s)[\frac{\mu+d}{1-k\alpha_s} - (\mu+d)]}{2\alpha_s(1-\alpha_s)\beta} \\
= & \frac{(2 - \alpha_s - \frac{1-\alpha_s}{1-k\alpha_s})\mu + k\alpha_s(1 - \frac{1-\alpha_s}{1-k\alpha_s})d}{2\alpha_s(1-\alpha_s)\beta} > 0.
\end{aligned}$$

Therefore $F(\frac{\beta - (\mu+d)}{\alpha_s\beta}) > 0$ implies no solution of $F(p) = 0$ over $\left(\frac{\beta - (\mu+d) - \frac{r-\mu}{d}}{\alpha_s\beta}, \frac{\beta - (\mu+d)}{\alpha_s\beta}\right)$, because $F(p)$ is decreasing over $\left(\frac{\beta - (\mu+d) - \frac{r-\mu}{d}}{\alpha_s\beta}, \frac{\beta - (\mu+d)}{\alpha_s\beta}\right)$.

Together if $\beta > \mu + d$ and $r > \mu$, a unique solution of $F(p) = 0$ exists over $\left(\frac{\beta - (\mu+d) - \frac{r-\mu}{d}}{\alpha_s\beta}, \frac{\beta - (\mu+d)}{\alpha_s\beta}\right)$ when $R_0 > 1$ and $i^* < \frac{r-\mu}{d}$ and no solution exists over

$\left(\frac{\beta-(\mu+d)-\frac{r-\mu}{d}}{\alpha_s\beta}, \frac{\beta-(\mu+d)}{\alpha_s\beta}\right)$ when $R_0 < 1$ or $i^* > \frac{r-\mu}{d}$.

Because of the complexity of the expressions, we will not express the positive steady state explicitly. Now assume that $R_0 > 1$ and $i^* < \frac{r-\mu}{d}$. The Jacobian matrix for a positive steady state can be expressed as:
$J = \begin{bmatrix} -\mu-(1-\alpha_s)\beta i^* & -[(1-\alpha_s)\beta-d]p^* & \frac{r}{K}(-k+p^*) \\ -\alpha_s\beta i^* & -(\beta-d)i^* & \frac{r}{K}i^* \\ 0 & -dN^* & -\frac{r}{K}N^* \end{bmatrix}$. The study of the eigenvalues may be continued in future, for corresponding stability results.

□

APPENDIX D

ADDITIONAL SIMULATION RESULTS

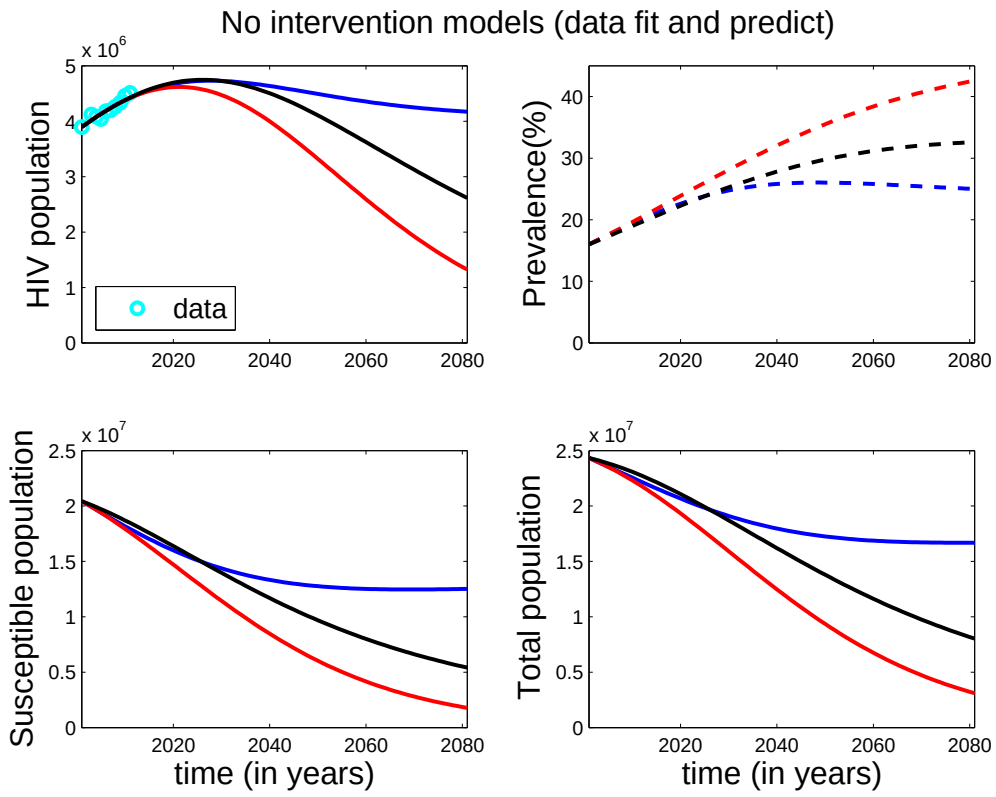


Figure D.1: Blue color corresponds to model with constant entrance rate. Red color corresponds to model with linear entrance rate. Black color corresponds to model with linear entrance rate.

Table D.1: Baseline parameter values generated from data fitting

	constant entrance	linear entrance	logistic entrance	initial guess
β	0.196924711	0.196935536	0.196924711	$1-(1-0.0038)^{80}$ ≈ 0.26256628106
μ	0.029793556	2.93E-02	0.02438153	$1/35$ $\approx 0.028571428571429$
d	0.119146121	0.11955454	0.125	$1/10 = 0.1$
Λ	996344			fixed value calculated from data
r		0.04095	0.0511875	fixed value or initial guess 0.04095
K			1.13E+08	9.00E+07 (Figure (A.1))
err	0.044056109	0.044755966	0.043080597	

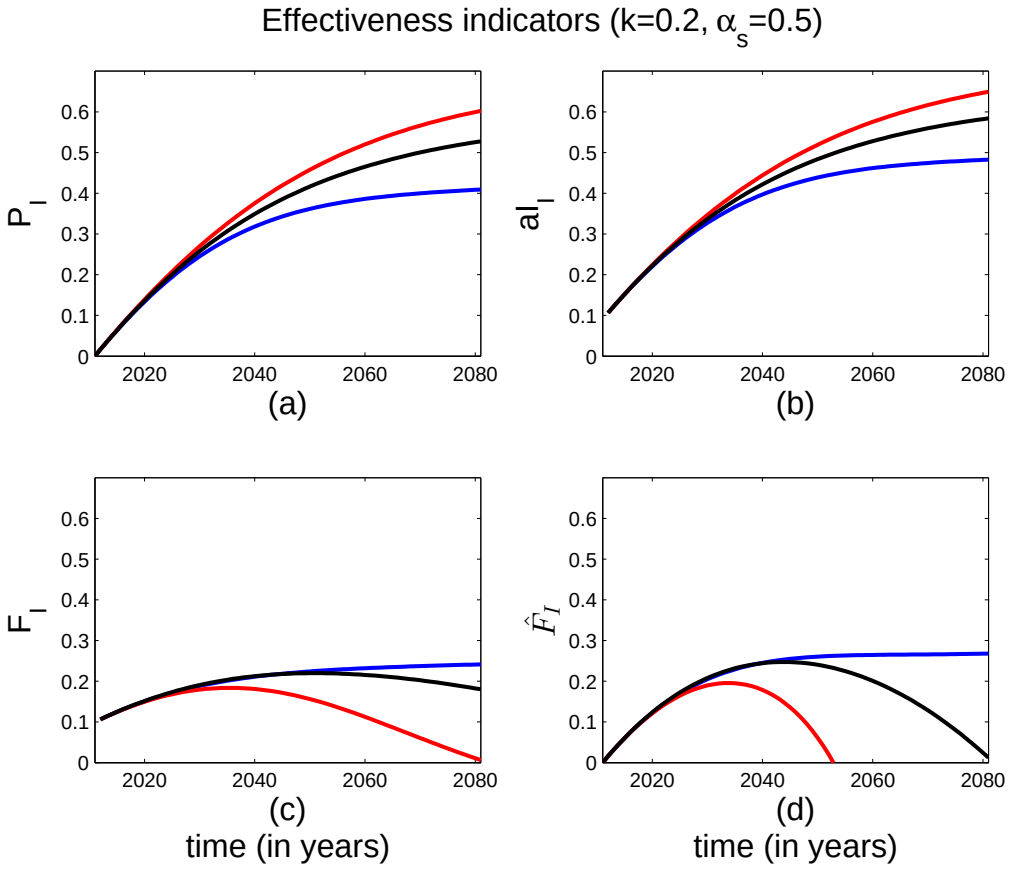


Figure D.2: For each simulation, models with different entrance rates use the corresponding parameter values from outfitting(see Table D.1). Blue color corresponds to model with constant entrance rate. Red color corresponds to model with linear entrance rate. Black color corresponds to model with linear entrance rate. Notice that for models with linear entrance rate and logistic entrance rate, some indicators are not well defined and can have negative values which are not shown in the figures.