A MULTI-AGENT SIMULATION OF MSM HIV TRANSMISSION EXAMINING THE ROLE OF CONCURRENT RELATIONSHIPS

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI’I AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE AUGUST 2015

By
Robert Ross Puckett

Dissertation Committee:
Nancy Reed, Chairperson
Tim Brown
James B. Nation
Luz Marina Quiroga
Dennis Streveler
Copyright © 2015 by
Robert Ross Puckett
This dissertation is dedicated to all of the underrepresented MSM who are living with, coping with, dying from, and yet to be infected with HIV.

You are my friends and family. I wish I could have done more to help you.
ACKNOWLEDGMENTS

I would like to thank my adviser Dr. Nancy Reed for giving up so very much of her time in guiding my research and for her faith that I would see the process through to completion. Thank you to Dr. Streveler for your critical, incisive questions and encouragement to tackle difficult areas of research. I am very thankful to Dr. Quiroga for her detailed analysis of my work and for her support despite occasional international boundaries. To Dr. Nation, I am greatly thankful for your mathematical analysis of my dissertation. Last, but not least, I extend my heartfelt thanks to Dr. Tim Brown for his extensive support, expertise in the field of HIV modeling, and confidence in my abilities.
ABSTRACT

The Human Immunodeficiency Virus (HIV) is a complex virus with a high rate of death, several modes of transmission, and long asymptomatic incubation period. Over time the virus weakens the host’s immune system leading to Acquired Immune Deficiency Syndrome (AIDS). Transmission of HIV varies over the course of infection, from a high infectivity stage during the initial, acute infection phase, to a lower level during the asymptomatic phase, before rising again as AIDS develops. Once progressed to AIDS, a body becomes susceptible to a range of opportunistic infections and cancers. In developed countries, HIV disproportionately affects men who have sex with men (MSM), primarily due to the practice of unprotected anal receptive (UAR) sex.

Of particular interest for study, the role of concurrent relationships as a driving force of epidemics has become a hotly debated topic in recent years. Concurrent relationships occur when a person has multiple active sexual relationships with overlapping durations. Several factors may increase the impact of concurrency on transmission including higher infectivity during the primary HIV Infection (PHI) phase of HIV, greater frequency of sex, and an increased exposure to HIV from multiple partners [70, 59]. In 1998, Morris and Kretzschmar concluded from stochastic simulations that concurrent partnerships “dramatically increase the speed and pervasiveness of the epidemic spread” [101]. Numerous subsequent publications by various authors have argued for [115, 83, 125, 58] and against [104, 113, 146] the degree of impact for concurrency in HIV epidemics around the world. Thus, the role of concurrency is far from a settled subject.

Epidemic modeling helps researchers in developing hypotheses on the driving forces, testing intervention strategies, and predicting the course of the HIV epidemic. Although many models for HIV have been developed, HIV remains a difficult subject to model owing to complex transmission modes, limitations in data collection, and variability in human behavior. Instead of studying populations in aggregate, microsimulation techniques allow for the study of individuals over time. A multi-agent system (MAS) can be used to implement a microsimulation of HIV transmission among individual people existing within a complex network of sexual partners. Through simple rules governing agent interactions, complex behavior can emerge from the society of agents.

In our research, we have built a multi-agent simulation of HIV transmission dubbed “MASHIV” used to model HIV epidemics in a single-sex community of men who have sex with men (MSM). HIV progression and mortality is based upon a CD4 compartment model adapted from the Spectrum software [62] to the realm of daily timestep multi-agent systems. MASHIV supports both duration-based dating events and probabilistic dating events. We utilize normal hazard functions to impose normal distributions upon probabilistic events, allowing local event probability generation at the agent level without centralized control. Our simulator includes support for risk reduction strategies such as negotiated safety arrangements, HIV discovery-based risk reduction, assortative/disassortative partner mixing weights, and HIV-status specific serosorting. The simulator also
supports simple antiretroviral drugs (ARV) programs for pre-exposure prophylaxis (PrEP) and antiretroviral therapy (ART).

We applied MASHIV to a simplified representation of the MSM epidemic in Bangkok, Thailand to investigate whether concurrency may be a driving factor of the HIV epidemic of the Bangkok MSM community. Our test results suggest that, while concurrency does have a theoretical possibility of contributing significantly to HIV spread under certain conditions, concurrency does not currently represent a substantial contribution to the spread of HIV to MSM in Bangkok at this time. While we calculated a positive correlation between concurrency and the number of total infections, the overall change between 0% practicing concurrency and 100% concurrency yielded only a 2% change in HIV prevalence.
# CONTENTS

Acknowledgments ................................................................. iv

Abstract ............................................................................. v

List of Tables ........................................................................ xii

List of Figures ......................................................................... xiv

List of Abbreviations ............................................................. xvii

1 Introduction ......................................................................... 1

2 Related Work ....................................................................... 5

2.1 Network Models ............................................................. 5

2.2 Complex Agents ............................................................. 8

2.3 Injecting Drug Use .......................................................... 11

2.4 Proof of Concepts .......................................................... 12

3 Modeling HIV Spread ......................................................... 15

3.1 AIDS Pathology ............................................................. 16

3.2 Transmission ................................................................. 17

3.3 Testing ........................................................................... 18

3.4 Drug Therapies .............................................................. 19

3.4.1 Post-Exposure Prophylaxis ........................................ 21

3.4.2 Pre-exposure Prophylaxis .......................................... 21

3.4.3 Anti-Retroviral Therapy ............................................ 22

3.5 Risk Factors .................................................................. 23

3.5.1 Sexually Transmitted Infections ................................. 24
3.5.2 Concurrent Relationships ........................................... 24
3.5.3 Versatility .......................................................... 27
3.5.4 Condoms ............................................................ 29
3.5.5 Circumcision ......................................................... 29
3.5.6 Seroadaptation ....................................................... 30

4 MASHIV Overview ................................................................ 32

4.1 Parameter Set ........................................................... 33

4.2 HIV Model .................................................................. 36
  4.2.1 HIV Initialization .................................................... 36
  4.2.2 HIV Transmission .................................................. 37
  4.2.3 Disease Progression & Mortality ................................. 41
  4.2.4 Adaptation of CD4 Model ........................................ 46

4.3 Agent Model ............................................................. 52
  4.3.1 Partner Finding ....................................................... 54
  4.3.2 Dating Evaluation .................................................... 55
  4.3.3 Pair Formation & Dissolution ................................... 57
  4.3.4 Sexual Roles .......................................................... 57
  4.3.5 Sexual Frequency .................................................... 58
  4.3.6 Condom Use .......................................................... 60
  4.3.7 HIV Testing & Discovery ......................................... 61
  4.3.8 ART ................................................................. 61
  4.3.9 PrEP ................................................................. 62
  4.3.10 Mortality ............................................................. 62
7.3 Normal Hazard Computation Load ........................................ 143
7.4 Normal Hazard Standard Deviations ..................................... 146
7.5 Object-oriented Issues ...................................................... 146
7.6 Probability Calculations .................................................... 147
7.7 Fitting Prevalence with Agents ......................................... 147
7.8 PHI & Contact Rates ....................................................... 147

8 Future Work ........................................................................ 149
8.1 Agent Model Refinement .................................................... 149
8.2 Complex Agent Expansion ................................................ 149
8.3 PrEP in Thailand .............................................................. 150
8.4 Partnership Duration ......................................................... 150

9 Summary ............................................................................ 151

Bibliography ......................................................................... 152
LIST OF TABLES

3.1 HIV Transmission Estimates ................................................. 20
3.2 Factors Influencing ART Effectiveness .................................. 23
3.3 Relative Increase of Susceptibility and Infectiousness Due to Presence of STIs .... 24
3.4 Sydney MSM Per-Contact HIV Transmission Probability ................. 30

4.1 Operational Parameters ...................................................... 34
4.2 Population Parameters ...................................................... 35
4.3 Minimal Group Parameters ................................................ 36
4.4 HIV Parameters .............................................................. 40
4.5 Probabilities of HIV Transmission ......................................... 41
4.6 Distribution of New Infections By CD4 count ............................ 44
4.7 Average Number of Years in a CD4 Category ............................ 44
4.8 Mortality Without ART ..................................................... 45
4.9 Mortality With ART .......................................................... 45
4.10 Uniform Daily Probabilities ................................................. 47
4.11 Standard Deviation in Years for CD4 Categories ......................... 48
4.12 Hypothetical Mean Years Survival ....................................... 51
4.13 Agent State ................................................................. 90
4.14 HIV Instance ............................................................... 91
4.15 Relationship Object ....................................................... 91
4.16 Relationship Schema ...................................................... 92
LIST OF FIGURES

3.1 Possible Uses of Antiretroviral Drugs ............................................... 20
3.2 Simulating the Effects of PHI Transmission and Concurrency ................. 26
3.3 Simulating the Effects of PHI Transmission and Concurrency with Alternating Sex Roles ................................................................. 28

4.1 The CD4 Compartment Model ............................................................. 43
4.2 Comparison of Uniform and Normal-Hazard Transition Patterns .............. 49
4.3 Comparison of Uniform and Normal-Hazard Transition Rates .................. 50
4.4 Algorithm Diagram ............................................................................ 53
4.5 Dating Module .................................................................................... 54
4.6 Set Editor ............................................................................................ 64
4.7 Parameter Editor .................................................................................. 65
4.8 Interactive Simulator ........................................................................... 67
4.9 Data Collection Selector ...................................................................... 68
4.10 Aggregate Editor ................................................................................ 69
4.11 Aggregate Simulations in Progress ....................................................... 72
4.12 Completed Aggregate Simulations Interface ......................................... 73
4.13 Aggregate Report Tool ....................................................................... 74
4.14 Aggregate Analysis Tool ................................................................... 75
4.15 Comparison of the Effect of Progression Rate on Mortality for Different Ages of Initial Infection .............................................................. 78
4.16 Spectrum Mortality Distribution Based Upon Infection Start Age ............ 79
5.12 Effect of PrEP on the Cumulative Number of Infections at 15 Years with 50% PrEP Coverage ................................................................. 112
5.13 Effect of PrEP on the Cumulative Number of Infections at 15 Years with 70% PrEP Coverage ................................................................. 113
5.14 Effect of Condom Use on the Cumulative Number of Infections at 15 Years ................................................................. 115
5.15 Effect of Discovery-based Risk Reduction on the Cumulative Number of Infections at 15 Years ................................................................. 117
5.16 Duration-based with Seropositive Serosorting ................................................................. 118
5.17 Duration-based with Seronegative Serosorting ................................................................. 118
5.18 Duration-based with Both Positive and Negative Serosorting ................................................................. 119
5.19 Probabilistic with Seropositive Serosorting ................................................................. 119
5.20 Probabilistic with Seronegative Serosorting ................................................................. 120
5.21 Probabilistic with Both Positive and Negative Serosorting ................................................................. 120
6.1 Analysis of Concurrency and Sexual Frequency ................................................................. 130
6.2 Analysis of Concurrency ................................................................. 131
6.3 Prevalence and Mortality for Concurrency Test ................................................................. 132
6.4 Prevalence and Mortality for Stochastic Variation Test ................................................................. 134
6.5 Analysis for Percent Initial PHI Test ................................................................. 135
6.6 Analysis of PHI Multiplier Test ................................................................. 136
6.7 Analysis of Sexual Frequency and Concurrent Relationships ................................................................. 138
6.8 Analysis of Sexual Frequency ................................................................. 139
7.1 Concurrency vs. Cumulative Infections ................................................................. 142
7.2 Comparison of Normal-Hazard and Weibull-Hazard Approaches ................................................................. 145
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIDS</td>
<td>acquired immune deficiency syndrome</td>
<td>4, 15, 17</td>
</tr>
<tr>
<td>ART</td>
<td>antiretroviral therapy</td>
<td>vi, 5, 17, 19, 22, 42, 61, 133, 151</td>
</tr>
<tr>
<td>ARV</td>
<td>antiretroviral drugs</td>
<td>vi, 19, 21, 22</td>
</tr>
<tr>
<td>AZT</td>
<td>azidothymidine</td>
<td>22</td>
</tr>
<tr>
<td>CDC</td>
<td>Centers for Disease Control and Prevention</td>
<td>15, 19, 21, 150</td>
</tr>
<tr>
<td>CDF</td>
<td>cumulative distribution function</td>
<td>47, 125, 126, 144</td>
</tr>
<tr>
<td>EPP</td>
<td>Estimates and Projections Package</td>
<td>41</td>
</tr>
<tr>
<td>FDA</td>
<td>Food and Drug Administration</td>
<td>18, 21</td>
</tr>
<tr>
<td>FTC</td>
<td>emtricitabine</td>
<td>21, 22</td>
</tr>
<tr>
<td>GIS</td>
<td>geographic information system</td>
<td>12–14</td>
</tr>
<tr>
<td>GRID</td>
<td>gay-related immune deficiency</td>
<td>15</td>
</tr>
<tr>
<td>GUI</td>
<td>graphical user interface</td>
<td>4, 33, 62</td>
</tr>
<tr>
<td>HAART</td>
<td>highly-active antiretroviral therapy</td>
<td>6, 10</td>
</tr>
<tr>
<td>HIV</td>
<td>human immunodeficiency virus</td>
<td>1, 4, 15, 16, 24</td>
</tr>
<tr>
<td>HTLV-III</td>
<td>human T-cell lymphotropic retrovirus type III</td>
<td>15</td>
</tr>
<tr>
<td>IA</td>
<td>immunological assay</td>
<td>18</td>
</tr>
<tr>
<td>IDU</td>
<td>injecting drug user</td>
<td>5, 11–13, 16–18, 21, 122, 150</td>
</tr>
<tr>
<td>KS</td>
<td>Kaposi’s sarcoma</td>
<td>15</td>
</tr>
<tr>
<td>MAS</td>
<td>multi-agent system</td>
<td>v, 3, 4, 12, 32, 142, 151</td>
</tr>
<tr>
<td>MSM</td>
<td>men who have sex with men</td>
<td>v, 1, 3, 4, 15, 16, 18, 21, 32, 122, 125–127, 141, 150, 151</td>
</tr>
<tr>
<td>NEP</td>
<td>needle exchange program</td>
<td>11, 12</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
<td>Page(s)</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>NIDU</td>
<td>non-injecting drug use</td>
<td>12</td>
</tr>
<tr>
<td>NNRTI</td>
<td>non-nucleoside reverse transcriptase inhibitors</td>
<td>22</td>
</tr>
<tr>
<td>NRTI</td>
<td>nucleoside reverse transcriptase inhibitors</td>
<td>22</td>
</tr>
<tr>
<td>PCP</td>
<td>pneumocystis carinii pneumonia</td>
<td>15</td>
</tr>
<tr>
<td>PDF</td>
<td>probability density function</td>
<td>47</td>
</tr>
<tr>
<td>PEP</td>
<td>post exposure prophylaxis</td>
<td>17, 19, 21</td>
</tr>
<tr>
<td>PHI</td>
<td>primary HIV Infection</td>
<td>v, 2, 10, 17, 25, 32, 37, 41, 42, 101, 127, 128, 133, 137, 141, 147</td>
</tr>
<tr>
<td>PI</td>
<td>protease inhibitors</td>
<td>22</td>
</tr>
<tr>
<td>PMTCT</td>
<td>prevention of mother-to-child transmission</td>
<td>19</td>
</tr>
<tr>
<td>PrEP</td>
<td>pre-exposure prophylaxis</td>
<td>vi, 17, 19, 21, 22, 37–39, 52, 61, 62, 109, 150, 151</td>
</tr>
<tr>
<td>STI</td>
<td>sexually transmitted infection</td>
<td>6, 7, 18, 21–24, 29</td>
</tr>
<tr>
<td>TDF</td>
<td>tenofovir disoproxil fumarate</td>
<td>21, 22</td>
</tr>
<tr>
<td>TG</td>
<td>transgender</td>
<td>150</td>
</tr>
<tr>
<td>UAI</td>
<td>unprotected anal intercourse</td>
<td>21, 31, 60, 151</td>
</tr>
<tr>
<td>UIAI</td>
<td>unprotected insertive anal intercourse</td>
<td>38, 40, 41</td>
</tr>
<tr>
<td>UNAIDS</td>
<td>the Joint United Nations Programme on HIV/AIDS</td>
<td>24, 41, 71</td>
</tr>
<tr>
<td>URAI</td>
<td>unprotected receptive anal intercourse</td>
<td>11, 16, 18, 38, 40, 41</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

Worldwide, the year 2011 ended with 34 million people living with the human immunodeficiency virus (HIV), 2.5 million new infections, and 1.7 million deaths due to AIDS [169]. Since the recognition of HIV in 1981, over 36 million people have died of HIV related causes [181]. Men who have sex with men (MSM) are disproportionately affected by HIV due to a combination of a high transmission probability of unprotected anal sex, societal stigma [13, 110, 14], versatile sexual positioning [26], and personal behavioral factors [24, 35]. A survey of capitol cities measured HIV prevalence in the MSM populations to be on average 13 times higher than the prevalence of the general population of the country [169].

Unfortunately, detailed information on the status of the HIV epidemic in MSM communities is not universally available due to the stigma and criminalization of MSM behavior and HIV infection. For example, in 86 countries, MSM behavior is a criminal offense, and 128 countries fail to report any data on HIV among MSM [14]. One cannot expect full and honest disclosure of sexual behavior in surveys when one’s sexual behavior is illegal.

Epidemic modeling tools are often used to project the course of epidemics, highlight specific needs for further surveillance, and test hypotheses on intervention strategies. In the past decade, the contribution of concurrent sexual partners as a driving force in HIV epidemics has become a hotly contested topic. Concurrent sexual relationships are defined by UNAIDS as “overlapping sexual partnerships where sexual intercourse with one partner occurs between two acts of intercourse with another partner” [80]. Primarily, the concurrency debate centers around whether concurrent sexual relationships are a major factor driving HIV epidemics or whether the spread of HIV can be best explained via other contributing factors such as sexual frequency.

Early modeling has emphasized the role of the number of partners as a key indicator for the spread of HIV. For example, in 1989, May and Anderson noted that the incidence of AIDS in modeling depended critically on the rate of partner change [120]. Later efforts began examining the role of contact networks and network structure, evoking the possibility that concurrent relationships may play an important role in the spread of HIV. The theory behind the role of concurrent relationships, or concurrency as it is often called, is that a person with concurrent partnerships increases the risk of his partners to acquiring HIV from his sexual contacts with other people [124].

Since 1995, Morris and Kretzschmar have repeatedly demonstrated that overlapping sexual networks have a theoretical possibility of amplifying the spread of HIV [127, 126, 129]. Similarly, in 1992 Watts and May concluded that a sufficient number of concurrent partnerships can create a rapid initial spread of infection among the network partners [175]. In 1996, Morris and Kretzschmar declared that network structure played a critical role in the transmission of HIV, and in 1997, stated that interventions aimed at ensuring “one partner at a time” were as important as those promoting
fewer partners. The authors strongly encouraged the collection of data tailored at investigating the role of concurrency.

Lagarde et al. didn’t find any evidence that concurrent relationships were a determining factor in 5 sub-Saharan Africa cities and concluded that concurrent relationships were only one of many contributing factors [104]. Halperin and Epstein disagreed, noting that African men and women typically have more than one concurrent partnership that can overlap by months or years [83]. Such longer-term trusted relationships are noted to seldom involve consistent condom use. However, in a 2003 Botswana study of 546 people, Carter noted 23% of respondents reported having concurrent relationships in the past year with 76% condom use among the people with concurrent partners compared to 69% among those who did not have concurrent partners [36]. This increased condom use may be indicative of either regional variations in condom use or perhaps variations in the length of concurrent relationships.

However, Lurie and Rosenthal contend that there is wholly insufficient empirical evidence for sub-Saharan Africa that concurrency is associated with HIV prevalence, increases HIV epidemic size, increases HIV transmission, increases the persistence of HIV, or that concurrency has a large effect [114]. The authors critique other researchers as conflating theoretical possibility of the effect of concurrency with actual effect in the real world. Additionally, the authors emphasize the need for a common definition of concurrency, better study design, and the administering of longitudinal studies to further investigate concurrency. Mah and Halperin maintain that there is sufficient evidence to move forward given the social and cultural underpinnings of concurrency in the region coupled with the higher transmission probability of HIV during the early PHI phase [115].

Contrary to the earlier reports of high concurrency level, in 2013, Sawers reported that improved questionnaires yielded a mean concurrency of 3.4% for sub-Saharan Africa. This level, Sawers maintains, is insufficient to support the claim that concurrency can explain the exceptionally high HIV prevalence in the region [145]. In response, Morris et al. point to evidence that the DHS surveys, which Sawers bases part of his conclusions upon, “systematically underestimate levels of multiple and concurrent partnerships” [128].

Needless to say, the role of concurrency as a driving factor of HIV epidemics appears to be far from resolved. In developing intervention strategies to combat the spread of HIV it is necessary to know what the driving forces of the epidemic are and how best to counteract them. Given the often limited resources available to enact such intervention strategies, it is important that interventions are tailored towards producing the greatest impact for the least amount of resources. Thus, policy makers need to know to what degree they should be concerned about concurrency. Is it a critical issue that must be addressed, or is money better spent elsewhere such as toward reducing overall sexual frequency?

In 2010, Morris et al. concluded that while, at the outset, concurrency measures for Uganda, Thailand, and the U.S. in the early 1990s demonstrated similar degrees of cumulative annual
concurrency, variation in the duration and coital risk exposure of these concurrent relationships produced substantial variations in HIV transmission [124]. HIV prevalence estimates for the MSM population in Bangkok, Thailand have risen from 17.3% in 2003 to 30.8% in 2007 demonstrating a concerning surge in infections [78]. In 2008, Li et al. reported on a survey of 456 male bisexuals and 1,125 MSM-only individuals divided among Phuket, Bangkok, and Chiang Mai [111]. 66.7% of the bisexual respondents reported 6 or more partners in the past 3 months compared to 27.4% of the MSM-only respondents. Condom use among bisexual males with male partners was 77.6% compared to MSM-only respondents at 62.9%. These reported levels of concurrent relationships and the observed explosive HIV growth in Bangkok raise the question of whether the relationship between the two is mere correlation or causation. Additionally, if concurrent relationships are contributing to the spread of HIV in Bangkok, to what degree is concurrency contributing?

To address that point, we have developed a multi-agent system (MAS) called MASHIV to explore the role of concurrency as a driving factor in the spread of HIV epidemics in MSM communities. A MAS is a computer system comprised of computational goal-oriented agents with local views and decentralized control. The agents exist and interact with each other in an environment according to a set of predefined actions and rules to accomplish their goals. Agents develop their own internal state unique to their personal history of interactions and observations.

In recent years, MAS have gained popularity in modeling human behavior. While an increasing number of researchers have applied MAS to a variety of aspects of the HIV epidemic, few have applied this technology to the study of concurrent behavior. Fewer still have applied MAS to the unique aspects of the MSM population as their primary focus [187, 121, 82, 178, 112]. Additionally, many of these MAS simulations rely on highly simplified models of HIV mortality and dating behavior, utilizing fixed-duration HIV progressions and static rates of relationship acquisition.

Concurrency has been previously examined via more traditional approaches such as differential equation models and contact network models. However, as concurrency relies on simultaneous sexual relationships between individuals, differential equation models that treat populations in aggregate fail to adequately represent individual-level population dynamics. While the addition of behavioral parameters to such models increases their flexibility, these models still lack an individual-level personal history of interactions. Network models of HIV tend to focus primarily on the impact of sexual and social network evolution and structure. While these models do allow for easier observation of concurrent behavior and the history of individual interactions, most network models lack goal-oriented behavior of the individuals.

However, MAS simulation allows for the modeling of interactions that are intractable or impossible to represent via analytical means due to the exponential proliferation of possible states in which the system could exist. Thus, we believe that the realm of simulating HIV with MAS offers promise towards addressing the debate of the role of concurrency in HIV epidemics. In this dissertation, we describe the development of a multi-agent system for examining the role of concurrency and
related transmission factors in HIV epidemics of MSM populations. Our simulator introduces the novel application of a CD4 compartment model to the realm of individual-based HIV simulation. While the original CD4 compartment model, developed by Stover et al. for Spectrum [158], applied annual HIV progression and mortality to aggregate populations, we utilize a normal distribution hazard function for estimating daily transition probabilities and an exponential hazard function for estimating the daily probability of mortality. A normal distribution hazard function produces a bell curve of compartment transitions centered about the mean whereas the exponential hazard function produces a constant daily probability of mortality. Our simulator is a capable platform for simulating a wide range of MSM HIV epidemics, allowing for both fixed duration and probabilistic distributed HIV progression and dating scenarios.

In Chapter 2 we review existing literature in the field of MAS modeling of HIV with particular emphasis given to researchers who study the unique attributes of MSM communities. In Chapter 3 we provide a brief overview of the history, etiology, pathology, and risk factors associated with HIV and acquired immune deficiency syndrome (AIDS). This chapter not only explains the complexities of HIV transmission to the reader, but describes our choice of HIV model parameters. Next, in Chapter 4 we provide a detailed description of the MASHIV simulator and its constituent parts of parameter set, HIV model, agent model, and graphical user interface (GUI). Following this specification, in Chapter 5 we analyze the effects and consequences of parameter values. In Chapter 6, we apply our model to a dataset we’ve developed for the MSM population in Bangkok, Thailand. In Chapter 7 we discuss the results of our simulations and the consequences of our design decisions. In Chapter 8 we outline future work remaining. Lastly, in Chapter 9 we summarize the findings of our research.
CHAPTER 2
RELATED WORK

Multi-agent simulations have been applied to numerous aspects of HIV Transmission. Research has tested the impact of ART, injecting drug user (IDU), intervention strategies, sexual mixing models, sexual/dating network models, and assumptions on sexual practices. While some have applied MAS to intracellular transmission dynamics [46, 68, 38, 105, 192], in this section we summarize the efforts to simulate populations at the individual (e.g. person) level.

As it is not relevant to our research, we do not cover models that treat populations in aggregate. Each person must be modeled independently, developing from a personal history of interactions with other people and observations of the environment.

2.1 Network Models

Network models focus on HIV transmission between individuals existing within social and sexual networks. These types of models tend to focus primarily on the evolution or impact of the network topology, giving little attention to adaptive behavior of the individuals. Individuals tend to follow static or probabilistic behavior with little variation. One may consider these individuals to be purely reactive agents, lacking goal-oriented behavior or a cognitive model. As the network is the primary focus, network models may impose constraints on individual behaviors which are not compatible with multi-agent design. As network models are not our primary focus, this section is not meant to be comprehensive as to all types of network models. In this section we summarize influential individual-level network models of HIV transmission.

Morris explored concurrency among heterosexual couples by testing several scenarios of partnership formation [127]. The simulation included an optional assortative pairing which used degree-based homophily. In assortative pairing, individuals prefer to pair with other individuals of similar attributes. In this case, the degree-based homophily may be perceived as the desire of individuals to pair with others of a similar level of promiscuity, the authors did not ground their implementation with a sociologically-informed basis. They concluded that while concurrent partnerships lowered the probability of an epidemic, when one did occur, the resulting epidemic developed faster and proved more pervasive. However, the simulation lacked direct modeling of sexual transmission and merely relied upon a highly simplified HIV model whereby HIV discordant couples exhibited a daily probability of transmission. The simulation did not include any HIV progression toward mortality, but this may be partially attributed toward the author’s three-year simulation window. While this is an individual-based simulation, the practice of fixing the number of concurrent relationships violates a principle tenet of multi-agent design whereby agents should act in an autonomous manner independent of centralized control.

In the following year, Kretzschmar and Morris continued their earlier efforts to examine the im-
pact of concurrency as a driving factor of heterosexual HIV spread [102]. Simulations demonstrated that disassortative sexual networks increase the odds of larger HIV epidemics. In disassortative pairing, individuals prefer to pair with people who have dissimilar characteristics. Subsequent analysis of a population with half of relationships set to be concurrent resulted in an HIV spread measured 10 times larger than a case with strict sequential monogamy [126]. Morris and Kretzschmar noted this effect can be attributed to concurrent relationships joining the population into a more easily reachable and connected social network.

Next, Kretzschmar and Dietz compare a differential equation model without pair formation to a model with explicit pair formation and partnership duration [101]. In the differential equation model, populations are considered in aggregate with separate groups for concordant and discordant HIV status pairings. The simulation demonstrated substantial contribution to the epidemic spread from the asymptomatic phase of HIV infection despite a low per-contact infectivity. Additionally, the short acute infection phase only appeared to contribute significantly when coupled with short durations of partnerships.

In 2000, Morris and Kretzschmar applied their microsimulation techniques to a population–Uganda [125]. Both long-term and short-term relationships are modeled. Relative to simulations of serial monogamy, the simulations assuming constant levels of concurrency resulted in a 26 percent increase in HIV prevalence after five years.

While the Morris and Kretzschmar models are useful tools for studying the theoretical impact of concurrency, their models involve significant simplification of HIV and partnership dynamics. For example, their models use a fixed daily HIV transmission rate for discordant heterosexual couples. Such a rate does not capture the difference in the probability of transmission based upon which partner is infected. For example, in the absence of circumcision, male to female transmission is approximately twice the probability of female to male transmission [134]. Circumcision, sexually transmitted infections (STIs), and other factors can significantly alter this balance in a population. Additionally, a fixed rate ignores the variance of infectivity due to the different stages of infection such as PHI and AIDS. A more realistic model would include the effects of HIV mortality and variable infectivity.

Building upon the earlier studies of Kretzschmar’s group [103], Xiridou et al. extended the above model allowing for AIDS mortality, concurrent casual partnerships, infection awareness influencing behavior, highly-active antiretroviral therapy (HAART) effects, and sexual roles [188]. Applied to homosexual men in Amsterdam, the simulation shows that steady, rather than casual, partnerships contribute more toward transmission. However, as the authors note, this may be attributed primarily to the greater degree of risky behavior assumed within steady relationships. Subsequently, Xiridou et al. concluded that primary HIV infections (PHI) contribute more heavily to the spread of HIV from casual partners than steady partners [187].

Along similar lines, Wei et al. presented a small-world network model of HIV transmission in
a community of MSM [178]. The authors adopted the casual and steady partner delineation set forth by Kretzschmar and Morris [102]. In Wei et al.’s simulation, agents will choose to engage in less risky behavior when the HIV prevalence in its local network, and perception of the global epidemic status, exceeds a certain threshold. Subsequent to a lowered infection rate, the agent will revert to its normal behavior. Agents will disassociate from HIV positive agents. While the agent design does exhibit some individual level behavior, it appears to be a predominantly reactive agent, lacking complex goal-oriented cognition.

Bracher et al. developed a simulation to investigate HIV prevalence in Malawi brides and grooms [29]. This age-based model allowed for an elaborately detailed specification of dating, marriage, divorce, and remarriage rates. However, for the purpose of this simulation, the authors assumed marriage only ended by death, and widows did not remarry. The rate of sex for an individual varies based upon age, risk propensity group, and marriage duration. Additionally, the model includes transmission of several other STIs which impact an individual’s infectiousness or susceptibility for HIV. The authors assume no condom use. From their simulations, the authors estimated HIV infection to be present in 13 to 20 percent of first-time marriages. The migration practices and premarital sexual activity of men resulted in their higher HIV prevalence compared to women who only married regionally. Overall, the role of extramarital sex in both men and women contributed significantly to the spread of HIV.

Later that year, Bracher et al. addressed their earlier limitation of marriage, extending the simulation to include divorce and remarriage[30]. The authors employ a novel idea of “strategic divorce” of which two cases are considered. In the first, divorce occurs upon observing symptoms of AIDS in a spouse. In the second, divorce occurs when a spouse’s infidelity total, which diminishes over time, reaches a certain threshold. Combined, these divorce methods resulted in a 20 percent reduction in lifetime risk of HIV, or about a 50 percent reduction if no remarriage is assumed. Without strategic divorce, but by introducing strict marital fidelity, HIV prevalence remained below 17 percent even when allowing for regular divorce and remarriage.

While the previous works of Bracher et al. assumed no condom use, in 2004 the authors addressed this omission, studying the impact of condom use on female lifetime risk of acquiring HIV [31]. Condom efficacy is reduced by breakage, slippage, degradation with age, or due to improper storage in warm climate. From this simulation, the authors derived a 42 percent lifetime risk of HIV infection for women when condoms are never used and a 8 percent risk for women when condoms are used consistently among non-married partners. This simulation does not cover the dynamics HIV testing nor does it allow for the mitigating effects of ART. Individuals exhibit limited goal-oriented behavior based upon assigned propensities toward specific actions. For example, an individual may marry, divorce, and remarry based upon their fixed inclination to do so, but this does not imply any goal-oriented decision making. An individual’s propensity for action is not affected by his personal history.
2.2 Complex Agents

There are numerous individual-level simulations of HIV transmission. However, few projects contain complex agents. That is, few have incorporated goal-oriented behavior or any form of cognitive model, relying solely on fixed reactive behavior. Many projects focused on knowledge diffusion models, based upon Axelrod [18]. Several others developed aspiration-based models.

Heuveline et al. developed a prototype complex MAS with aspiration-based courtship [86]. In a society where each person is assigned a quality level, an agent seeks partnerships with agents exceeding his aspiration level. Reacting to attempts at partnership formation, the aspiration level lowers due to rejection and raises from acceptance. The model incorporated agent migration between villages and a dynamic friendship network. A lack of publications after 2003 may indicate that this line of research may have been abandoned. However, their work would later influence other research endeavors [11, 166].

Sumodhee et al. presented a simulation where each agent has a specified propensity toward faithfulness, preferred number of partners, and frequency of condom use [160]. These parameters change over time in response to partner interactions and observing the environment. From a population divided into three risk groups, agents may form short-term and long-term partnerships. The agents may choose partners within their risk group (assortative pairing) or between risk groups (disassortative pairing). The model lacks variable infectivity, progression to AIDS, and all forms of mortality. The authors applied this model to the Taiwan population, adjusting risk group sizes to match the region’s HIV prevalence more accurately.

While Kretzschmar’s simulation [102] lacked any complex agent behavior, Rhee extended their model to simulate the social and sexual networks of Papua New Guinea for the purpose of examining the impact of idea diffusion on HIV spread [140]. Papua New Guinea presents an interesting test case as it contains extreme geographic isolation of ethnic groups. In the simulation, special advocate agents spread ideas of safe sex and HIV disclosure practices into the social network. An agent adjusts his personal awareness of safe sex practices according to his local social network of peers. The authors admit their assumption of equal weight between positive and negative ideas being spread may be naive in light of HIV stigma. Agents attaining a threshold of awareness become non-transmitters of HIV, practicing safe sex. Rhee concluded that the number of non-transmitters provided an indication as to the trajectory of the epidemic.

Alam, Meyer, and Ziervogel implemented a MAS to study the social impact of HIV spread in a South African community subjected to the stresses of food insecurity, climate variability, market fluctuations, and poor governance [10]. The simulation supports two dynamic small-world network layers—a social network, and a household network. Instead of supporting sexual transmission of HIV within the social network, the simulation employs a Gamma distribution to estimate incidence. Only preliminary results are presented.

Subsequently, Alam, Meyer, and Norling’s demonstrated that the initial network distribution
for their small-world networks provided no indication as to the final simulation outcome due to the dynamic nature of the network evolution [9]. Additional scenarios studied the interdependent effects of household dissolution, hunger, migration, and personal economy. This research project continued, reporting further on various scenarios [11] and developing network analysis tools to study their simulation networks [6, 3, 7, 4].

Finally correcting an omission in the group’s previous work, in 2008 Alam presents a sexual transmission network with per-coital act HIV transmission [2]. Alam introduces an aspiration-based dating mechanism whereby men choose female partners based upon log-normally distributed aspiration and quality levels. While the quality levels are static, the aspiration levels change over time in response to rejection and success in dating and courtship. Additionally, partner selection is influenced by an agent’s age, health, economic status, migration status, and social links. Females use similar criteria in deciding whether to reject suitors or begin a courtship period towards marriage. The actual length of courtship varies based upon the number of partners and certain cultural traditions. Each month, an agent’s risk of acquiring HIV is calculated based upon the agent’s preferred number of coital acts, the stage of HIV infection of partner, and the migration status of a partner.

Alam et al. continued the development of this simulator [8]. Concurrent partnerships are distributed according to a gender-specific log-normal distribution of maximal sexual partnerships. Sexual acts vary stochastically from 7 to 13 acts per month for an individual. To accommodate the variable infectivity of HIV, transmission rates proceed through three stages. The authors evaluate different reported per-contact transmission probabilities and the effect of exogenous incidence due to migration. Exogenous incidence is found to maintain epidemics that would normally have ceased. The authors conclude that their simulations emphasize the impact of sex workers and migrants on HIV transmission in Sub-Saharan Africa.

In 2010, Alam and Meyer use their simulator to compare their earlier aspiration-based mixing scheme to an endorsement-based strategy [5]. Static labels, such as kinship, are assigned between agents. Dynamic labels, such as trustworthiness, are applied and revoked over the course of the simulation. An agent assesses these positive and negative labels based upon his personal internal state. Agents build friendships based upon a similarity index. To form a relationship, both the male and female partners must find each other’s endorsement value exceeding an internal threshold value. Upon partnership formation, the lowest-ranked existing partner of an agent may be dropped to accommodate a personal limit on the number of relationships. Alam and Meyer concluded that their aspiration model exhibited greater volatility than the endorsement model, which demonstrated the typical epidemic plateauing effect.

Borrowing heavily from the courtship and aspiration relationship model of Alam et al. [8], Knittel et al. developed a multi-agent system for the study of concurrent partnerships [98]. The authors compared the output of this model to numerous empirical data sources of the United States. A sen-
Sensitivity analysis determined which parameter values increased or decreased the annual and lifetime number of partners for an agent. The authors admit the limitations of supporting only heterosexual transmission and the necessity to estimate parameter values for which there is insufficient data.

Mei et al. combine complex network and multi-agent simulation strategies to simulate the MSM population of Amsterdam [121]. The implementation incorporates steady partners, who have sex many times a year, and casual partners, who only have sex once a year. Only one steady partner is permitted, the presence of which reduces risky behavior with casual partners. For each year, an annual probability of transmission between each pair of agents is calculated based upon the HIV status of the partners, use of HAART, and type of relationships. We believe that the annual time step results in underestimating the spread of HIV due to the dilution of the impact of PHI. While PHI lasts only between a few weeks and a few months, transmission to sexual partners during this period increases substantially. Thus, while the author’s model only allows a primary infection to pass through one layer of the social network in a year, there is actually sufficient time for multiple rounds of PHI. The authors also assume random sexual role selection resulting in a highly sexually versatile population. This practice has been shown to potentially double HIV prevalence [69]. Lastly, the authors admittedly violate the local view tenet of agent design, allowing global statistics to directly influence individual agent behavior. The following year, Mei et al. illustrated how an increase of 30 percent in risky MSM behavior can counteract the mitigating influence of ART [122]. This result underscores the importance of reducing risky behavior.

A student of Sloot, Zarrabi joins the two normally disparate worlds of cellular and individual based modeling [193]. Zarrabi employs a filter-reduction method to approximate contact networks from HIV patient information. Contact and genetic networks are then coupled to build a HIV transmission network. In the MSM network studied, the author concluded the risk of being a super-spreader node correlated with higher viral loads and older age. Unfortunately, while the author did propose the development of a complex agent network, he did not actually build one. We believe that, although a multi-scale multi-agent system would be an interesting research challenge, the intrinsic level of detail required in such a simulation would likely render efficient computation intractable.

For the development of an AIDS prevention game, Klatt et al. developed a novel virtual agent game character based upon the Theory of Mind for the purpose of sexual negotiation [96]. Sexual negotiation involves offers, counter-offers, and a compliment-based persuasion tactic. Agents have a fixed level of attractiveness and a propensity towards a certain level of risk. During negotiation, an agent may offer a condom, offer unsafe sex, and compliment the other agent in hopes of reducing inhibition. An agent will work to maximize his goals of safety and arousal. An agent may form beliefs about a partner’s goals and may choose to maximize the partner’s perceived goal. Upon sex, arousal is increased proportionally to the attractiveness of the partner. An agent’s safety measure depends on whether or not safe sex occurred. The authors evaluate variations in goals to maximize
for each agent and the degree of sympathy toward each other’s goals. However, the authors did not attempt to justify their behavioral parameters with a literature basis. Given the abstract nature of the cognitive model presented, this is not surprising.

In 2012, Beyrer et al. reported on their development and use of a MSM multi-agent system of HIV transmission for illuminating the driving forces of MSM HIV epidemics [26]. The model, inspired by unpublished work of Goodreau, included sexual behaviors within steady and casual partnerships, circumcision prevalence, testing frequencies, and existing HIV treatment. The model was applied to urban USA and urban Peru with an initial population size of 5000 per country and an initial HIV prevalence of 15%. Beyrer et al. used the model to demonstrate the impact of role versatility coupled with high probability of transmission of unprotected receptive anal intercourse (URAI) in comparison to heterosexual sex.

### 2.3 Injecting Drug Use

Injecting drug users are disproportionately affected by HIV due to the high probability of transmission from needle sharing. With a per-contact transmission risk of 0.67 percent, needle sharing is comparable to the riskiest sexual practice, unprotected anal receptive, which has a 1 percent transmission probability [92]. Intravenous transmission accounts for 20 percent of all HIV-1 infections [48]. Frequency of injection is often quite high, roughly 3 times a day for heroin users, which elevates the risk of HIV transmission. Multi-agent simulations involving IDU transmission tend to treat the IDU population in isolation, avoiding complications from additional sexual transmission networks. However, IDU transmission among agents bears many similarities to sexual transmission networks. Thus, in this section we summarize IDU individual-based simulations.

Atkinson adapted the simulation technique of [108] to IDU shooting galleries, places people come to inject drugs and often share the provided needles [16]. Parameters include injecting seniority, injecting frequency, needle cleaning rate, and sharing frequency. Notably, instead of using a per-contact risk of transmission, a static number of HIV discordant contacts sufficient for infection was calculated from the per-contact risk. This practice represents a significant simplification of the HIV transmission process.

In 2003, Raboud et al. used a Monte Carlo simulation to study the impact of needle exchange programs (NEPs) on HIV spread within the IDU population of Vancouver, Canada [139]. A NEP, as the name suggests, is a place where people can exchange used needles for new ones. It is hoped that through needle exchanges, that there will be less shared use of contaminated needles and thus less HIV transmission. In the simulation, the authors varied NEP attendance and effectiveness. The population was divided into seven groups, each group representing a different level of risky IDU behavior. Needle-sharing partnerships formed with respect to a proportionate mixing pattern, and dissolved based upon the combined number of partners of a couple. The authors concluded that the most effective means of reducing HIV transmission involved increasing the attendance of NEP
and decreasing needle sharing for the reliable NEP participants. This type of simulation exhibits only deterministic reactive behavior at the individual level, lacking any complex cognitive model or use of historical knowledge.

A geographic information system (GIS) is a program which utilizes or manipulates geographical data as part of its computation. Xiong et al. implemented a GIS-enhanced simulation of IDU use subjected to peer pressure [186]. In a random social network, each agent maintains 5 to 8 contacts per day within his personal group of 20 friends. An agent’s peers may persuade him to adopt or reject drug use. Unfortunately, the initializing parameters are merely assumptions of the authors and lack a basis in literature. As such, the authors performed a sensitivity analysis of the input parameter space and reaffirmed general notions of the impact from reducing needle sharing and promoting education of high-risk behavior.

Unlike the previous projects, Marshall et al. bridges simulation of HIV transmission in the drug using community with sexual transmission networks [118]. Additionally, the authors modeled both IDU and non-injecting drug use (NIDU). Intervention strategies tested include needle exchanges, drug treatment program, HIV testing, and ART. This simulation was calibrated against existing IDU/NIDU prevalence, HIV prevalence, and HIV incidence measures of the New York City area. Instead of building a full social network, the authors model a risk network, where links only exist where there is sexual or injecting behavior between the agents. An agent may interact with a randomly sampled number of contacts at each time step for sexual or injecting interactions. Multiple sexual orientations are supported and probabilistic assortative mixing is applied for link formation to ensure compatible partners. Finding comparable results to published estimates of HIV statistics, the authors expected future utility in using MAS for evaluating hypothetical intervention strategies.

Jiayu et al. applied a two-layer multi-agent simulation with a small-world social network model to HIV transmission in IDU populations of China [91]. Agents participated in needle-sharing networks and reacted to social influences and HIV testing interventions.

Most recently, Wei et al. utilized a complex agent model to study IDU populations subjected to various intervention strategies [177]. Interventions included needle exchange programs, drug counseling, HIV testing, and ART. The authors concluded their initial simulations were consistent with empirical studies.

## 2.4 Proof of Concepts

Here we present projects where the authors have presented an introductory approach to modeling HIV transmission with agents, but have not yet completed the development needed to report significant results. These proof-of-concept implementations tend toward simplistic or unproven rules and abstractions of population dynamics. While the authors may present novel methods of modeling, the lack of validation and verification leaves the actual utility of these approaches unevaluated at present.
In 2004, Teweldemedhin et al. briefly described a multi-agent system comprised of controller, person, environment, and statistical agents [162]. Applying this model to the Republic of South Africa, their model yielded 90 percent accuracy with the South Africa Department of Health figures for the region. However, due to the exceptionally limited details provided, understanding or replicating the results seems unlikely.

Bobashev et al. simulated HIV transmission in needle sharing networks with seniority-based needle use [28]. Senior members, representing those with established access to heroin, would use new needles first and then pass the used needles along to junior members. Each person in the simulation existed within a network of 250 “buddies.” Agent types included drug dealers and drug users. The 5-year mortality of HIV and 0.0008 per-contact risk of IDU HIV transmission probability seem unrealistic. Although the authors presented an interesting proof of concept model, their simulation admittedly lacked proper validation. Additionally there is a noticeable lack of reference to existing work in the field.

Otukei simulated the effect of a quarantine intervention on the spread of HIV [131]. The author assumes a person will be quarantined after infection. For some reason, the author deemed it necessary to impart initial immunity for HIV to a fraction of the population. The lack of implementation details and results analysis suggest the simulation is still in the early stages of development.

Approving of Mei’s work, Tirado-Ramos recently proposed employing Mei et al.’s simulation [121] to study the effect of a laundry list of HIV prevention strategies upon populations with elevated risk of transmission such as prison drug users [166]. However, as of yet, nothing appears to have been developed.

Yergens et al. developed a GIS multi-agent system to model infectious disease spread [191]. While their primary concern was the spread of highly infectious diseases, the authors included HIV status in their agents as they intended to model sub-Saharan Africa regions where HIV prevalence is high. Certain vaccinations, such as the smallpox vaccine, adversely affect HIV positive people. However, the authors did not attempt to model HIV transmission. Person agents interact within and migrate between town agents, displayed on a GIS map. Although only preliminary results were presented, the authors plan to integrate their model with real-time data from the Global Surveillance and Emergency Response System.

Yang presents a proposed coupling of agent-based modeling with GIS [190]. The simulation includes the typical parameters of initial HIV incidence, daily contact rate, and per-contact HIV transmission probability. Additionally, the model also includes a measure of the strength of the infection and indicators of the agent’s hospital and treatment habits. The authors propose that such a model would be a useful decision support platform for guiding AIDS prevention efforts.

Wang et al. modeled a closed population of agents within a grid-based GIS representation [173]. Pairing is governed by simple pair formation and separation probabilities. Agents are able to
move through the map and have a vision parameter governing their interactions with others. HIV transmission may be mitigated by a probability of condom use.

For the purpose of developing a GIS-enabled early warning system for HIV, Xu proposes a multi-agent framework for simulating HIV transmission involving commercial sex workers, clients of sex workers, homosexuals, and the general population [189]. Xu et al. hope to apply this framework for modeling HIV spread in Kunming, China.
CHAPTER 3
MODELING HIV SPREAD

HIV, primarily spread by sexual transmission, is the virus which causes AIDS. Over time, HIV reduces the human immune system’s ability to function properly by destroying the cells necessary for the immune system to function effectively. Eventually, the immune system function is diminished to the point that the host becomes susceptible to opportunistic infections and certain cancers. Unlike most other viruses, the human immune system is unable to clear HIV from the body. Although there is no cure for HIV, treatments exist to prolong life, suppress HIV replication, and diminish the probability of transmission during sexual intercourse. Untreated, HIV infection progresses to AIDS and subsequent death in about 10 and 11 years respectively.

In 1981, doctors in the United States started noticing instances of pneumocystis carinii pneumonia (PCP) and Kaposi’s sarcoma (KS) lesions in previously healthy MSM [60]. On June 5, 1981, Gottlieb et al. first reported on five male patients in the Los Angeles area with PCP and noted “profoundly depressed” numbers of lymphocyte cells [71]. A subsequent publication in December 1981 proposed that an acquired cellular immunodeficiency afflicted homosexual patients experiencing PCP, KS, and multiple viral infections [72].

Early on this immunodeficiency disease was termed gay-related immune deficiency (GRID) as it tended toward afflicting MSM [12]. Other reports used the term “4H disease”, referring to the observed afflicted groups of Haitians, homosexuals, hemophiliacs, and heroin users [117, 47]. Finally, in 1982 the Centers for Disease Control and Prevention (CDC) adopted the term AIDS for referring to this disease, acknowledging that the disease was not limited to the groups encompassed in the earlier terms GRID and “4H disease”.

One of the first to identify the viral cause of AIDS, Luc Montagnier linked human T-cell lymphotropic retrovirus type III (HTLV-III) to the development of AIDS [33]. Converging with the work of Robert C. Gallo, HTLV-III was subsequently termed the human immunodeficiency virus (HIV). HIV is a member of the genus Lentivirus of the Retroviridae family [90]. Lentiviruses are commonly associated with a long incubation period and immune deficiency. While it is typical of retroviral reverse transcriptases not to use exonucleolytic proofreading for preventing transcription errors during viral replication, Roberts et al. discovered that HIV is the least accurate known reverse transcriptase [142]. The authors note that this error-prone transcription process creates a diverse HIV genome, hampering the immune system’s ability to combat HIV and complicating the development of an HIV vaccine.

Since the discovery of HIV in 1981, HIV has spread primarily through sexual contact to the level of an epidemic. As of 2013, an estimated 36 million people globally have died of HIV related causes [181]. Most of southern Africa and parts of eastern Africa are experiencing generalized HIV epidemics whereas concentrated epidemics exist in American continents, Europe, Asia, and
In generalized epidemics the HIV infections have become established in the general population. A generalized epidemic is defined as one where HIV prevalence in pregnant women exceeds 1% consistently [34]. These epidemics are primarily affecting heterosexual populations. Concentrated epidemics, on the other hand, occur when the HIV prevalence does not rise to the level of a generalized epidemic, but exceeds 5% in a vulnerable sub-population such as MSM, IDU, or sex workers. In concentrated epidemics, transmission remains largely confined in these vulnerable sub-populations and their immediate sexual partners.

MSM are disproportionately affected by HIV due to a combination of a high transmission probability of URAI, versatile sexual positioning, societal factors, and other behavioral factors. Unlike the strict sexual role segregation inherent in heterosexual couples, MSM exhibit more versatile sexual roles that overcome the low probability of transmission from the receptive to the insertive partner [26]. Additionally, societal stigma and discrimination against MSM and HIV inhibits HIV status disclosure, discourages individuals from seeking HIV treatment services, hampers the provision of such services, and complicates population studies [27].

### 3.1 AIDS Pathology

As time progresses, HIV passes through three stages—primary HIV infection (PHI), asymptomatic, and AIDS. Each stage represents a distinct period of HIV activity and thus virulence. Immediately after the initial infection, the individual enters the acute or PHI phase of HIV. As the body has not yet developed the capability to keep the virus in check, the virus multiplies rapidly, reproducing itself at the cost of CD4 cells which it destroys during viral replication. The PHI phase marks an especially virulent but short-lived period. During the PHI phase, the person has a 70 percent chance of showing symptoms, giving an early indication of the presence of the virus [136]. However, these symptoms are often overlooked or misattributed [54, 180]. A study in Rakai, Uganda estimated that 43 percent of HIV transmissions occurred within the first 10 weeks of infection [176]. However, Cohen et al. demonstrate enormous variation on the estimated impact reported in studies [48]. Estimates of the percent of HIV infections attributed to PHI ranged from 5 to 50% for Sub-Saharan Africa, 20 to 90% for US MSM, 5 to 20% for US in general, and 0 to 40% for Europe.

It is during the PHI phase that seroconversion occurs. Seroconversion refers to the process of the body developing and producing antibodies to a sufficiently detectable level. There exists a window period where HIV antibody tests may read as non-reactive despite the presence of an HIV infection. Thus, a seronegative HIV test result indicates that either that an HIV infection is not present, or an HIV infection is present but not yet detectable, i.e. seroconversion is not complete. A seropositive HIV test indicates the presence of HIV antibodies, and thus HIV. In common parlance, a person with a reactive and confirmed HIV test is “HIV positive”. As the body develops antibodies, the viral load, and thus virulence, diminishes marking the progression toward
the asymptomatic phase of infection. During the asymptomatic phase, which lasts 9 to 10 years on average but varies greatly among untreated individuals, the virus is replicating at low levels.

HIV primarily infects cells with CD4 receptor molecules. The term “CD4 cell”, also known as “T-helper cell” or “T-cell”, refers to a CD4 T-lymphocyte white blood cell which triggers the immune system to respond to anomalies such as viruses and bacteria. A CD4 count is a lab test to determine the number of CD4 cells in a cubic millimeter of a blood sample. The CD4 count is often used as a proxy to gauge the status of an HIV infection via the health of the immune system.

A normal CD4 count ranges from 800 to 1200 CD4 cells/mm$^3$ of blood [97]. Over time, HIV weakens the immune system, lowering the CD4 count. When the CD4 count drops below 200, the individual has progressed to AIDS. At this point the person becomes susceptible to opportunistic infections and certain cancers that the immune system would normally have been able to handle. Left untreated, AIDS may progress to death within a year.

Today, persons with HIV are routinely treated with a combination of potent antiviral drugs, collectively known as antiretroviral therapy (ART), which are extremely effective at slowing viral replication. ART slows virus progression thus extending a person’s life. ART is often administered when the CD4 count drops below a certain threshold value, typically in the 350-500 range. ART efficacy is dependent on a person’s physiology, stage of infection, and adherence to the drug regimen. While there are clinical trials to test PrEP—the administration of ART in absence of infection to increase resistance to transmission [161]—ART is generally provided subsequent to detection and confirmation of an HIV infection. Thus, HIV testing remains an important factor in the adoption of ART.

Medication treatment regimens post-exposure to HIV can be divided into two groups—post exposure prophylaxis (PEP) and ART. For PEP, drugs are administered shortly after exposure in hopes of preventing HIV infection. ART generally refers to drug regimens administered after an HIV infection is detected.

### 3.2 Transmission

HIV is considered a sexually transmitted disease as it is predominately spread by sexual contact. However, HIV may be transmitted through blood transfusions, shared needle use of IDU, contaminated needlestick, during the birthing process, and during breast-feeding. Table 3.1 reproduces Levy’s table summarizing the estimates of per-contact likelihood of infection and the percentage of HIV infections attributable to different modes of transmission [90]. The transmission estimates for sexual intercourse represent the unprotected rates, i.e., in the absence of condom use.

However, increases in the genital viral load, a measure of the amount of virus in genital secretions, can greatly impact the rates of sexual transmission. A model by Chakraborty et al. estimates a rate of heterosexual HIV transmission per sexual act of 0.03% when the viral load is 1,000 copies of HIV per mm$^3$ but 1% when the viral load is 100,000 [43]. During the early PHI phase of an HIV
infection the viral load soars.

Although unprotected receptive anal intercourse (URAI) represents the highest per-contact risk of sexually transmitted HIV, it only accounts for 5-10% of the world’s HIV infections [90]. However, in developed nations, HIV tends to be a concentrated epidemic primarily affecting populations with elevated risk of transmission such as MSM, IDU, and sex workers. While only estimated as 2% of the U.S. population [41], MSM represented 79% of new male HIV diagnosis and 62% of all HIV diagnosis in the United States for 2011 [40].

Although allowing many other STIs to pass more readily, oral sex is widely considered not to be a major transmission route for HIV [143, 156]. Hawkins et al. note that HIV transmission through oral sex is biologically plausible [84], but few cases are linked to oral transmission and the CDC considers the risk of oral HIV transmission to be low [42].

3.3 Testing

HIV testing refers to tests used to determine the presence of HIV in an individual. Awareness of one’s HIV infection is a critical first step towards receiving treatment. Identifying the presence of an HIV infection may involve detecting the presence of viral RNA, the p24 antigen, or anti-HIV antibody [51]. This is accomplished by tandem testing of a high-sensitivity test and a high-specificity test. High-sensitivity tests for HIV return accurate positive results with few false negatives when HIV is present whereas high-specificity tests for HIV return accurate negative results with few false-positives when HIV is absent. As such, the initial test may produce a few false-positive results that are subsequently caught by the secondary test which has few false positive results [50]. Initial testing is performed by an immunological assay (IA). Upon identifying reactive results, i.e. positive, the results are subsequently confirmed by Western blot or indirect immunofluorescence assay [39].

Non-reactive test results only indicate that HIV is not detectable, not that HIV isn’t present. That is, a test may provide a false-negative result in the case where HIV is present, but the marker for detecting HIV is not yet at detectable levels. As such, one might erroneously interpret a non-reactive IA result as proof of being HIV negative. Regular testing is necessary to confirm the absence of HIV.

Rapid HIV testing kits are available that allow an individual to quickly and visually detect an HIV infection without instrumentation. As of October 29, 2013, the US Food and Drug Administration (FDA) has approved five rapid HIV tests: OraSure Technologies’ OraQuick Advance HIV-1/2 Antibody Test; MedMira’s Reveal G3 Rapid HIV Test; Trinity Biotech’s Uni-Gold Recombigen HIV Test; Bio-Rad Laboratories’ Multispot HIV-1/HIV-2; and Alere’s Clearview Complete HIV 1/2 [171]. A 2005 estimate for cost per rapid test kit ranged from $14 to $25 [75].

For the US, Greenwald et al. estimate that 40-50% of HIV infections are detected within 1 year prior to diagnosis with AIDS [75]. As HIV takes 8-11 years to progress to AIDS, this statistic provides compelling evidence that many individuals are not testing regularly for HIV. Such a high
proportion of late discovery is not entirely surprising as HIV includes a long asymptomatic period. Approaching AIDS, the immune system is severely compromised and HIV becomes symptomatic through increased susceptibility to opportunistic infections. With a late discovery, infected individuals are not aware of their infection until much later in the progression of the virus. As such, there is an increased likelihood of onward HIV transmission and increased morbidity, mortality, and health care costs [148]. Krentz et al. observed that patients presenting with a CD4 count less than 200, i.e. late discovery cases, could produce medical costs in the year following diagnosis twice that of the earlier HIV discovery cases [100]. They attributed the difference in cost primarily to hospital care costs which were reported as 15 times higher for late presenting HIV cases. As of 2006, the CDC advocates routine HIV testing in healthcare settings with patient HIV prevalence of 0.1% or greater for all persons 13 to 64 years of age [32]. Additionally, the CDC recommends that individuals with an elevated risk of contracting HIV should be tested for HIV at least once a year.

With a confirmed HIV diagnosis, a person in the US may elect to initiate ART and reevaluate their sexual practices to reduce the possibility of onward infection. However, in many parts of the world, eligibility is limited by the individual’s current CD4 count. Weinhardt et al. observed from numerous studies that an HIV diagnosis typically results in a person having a significant reduction in unprotected intercourse and a reduction in the number of sexual partners [179].

3.4 Drug Therapies

There is no cure for HIV, however, there are therapies to prolong life and reduce the probability of transmission. Drug therapies for treating or preventing the spread of HIV include PrEP, ART, PEP, and prevention of mother-to-child transmission (PMTCT). These drug regimens differ in the target audience, the presence of an HIV infection, and the type of drugs administered. Figure 3.1 summarizes the types of antiretrovirals. PrEP and PEP are chemoprophylaxis administered to HIV negative individuals at-risk of contracting HIV with the aim of reducing the probability of the individual acquiring an HIV infection. For PrEP, the antiretroviral drugs regimen is targeted toward HIV negative but elevated risk individuals such as those in HIV serodiscordant partnerships. ART generally refers to ARV drug combinations administered to known HIV-positive individuals to prolong their life and reduce onward transmission. In the case of PEP, an antiretroviral drugs regimen is prescribed shortly after an individual believes he is exposed to HIV, such as by contaminated needlestick or via sex with a suspected HIV positive individual. PMTCT is an antiretroviral drugs regimen administered to pregnant women known to be HIV positive in hopes of preventing the transmission of HIV from the mother to her child.
## Table 3.1: HIV Transmission Estimates

<table>
<thead>
<tr>
<th>Type of exposure</th>
<th>Likelihood of infection after single exposure (%)</th>
<th>Global total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sexual intercourse</strong></td>
<td>0.0-1.0</td>
<td>70-80</td>
</tr>
<tr>
<td>Receptive vaginal</td>
<td>0.01-0.32</td>
<td>60-70</td>
</tr>
<tr>
<td>Receptive anal</td>
<td>1.0</td>
<td>5-10</td>
</tr>
<tr>
<td>Insertive anal</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>Insertive vaginal</td>
<td>0.01-0.1</td>
<td></td>
</tr>
<tr>
<td><strong>Injection drug use</strong></td>
<td>0.5-1.0</td>
<td>5-10</td>
</tr>
<tr>
<td><strong>Maternal transmission</strong></td>
<td>12-50</td>
<td></td>
</tr>
<tr>
<td>Pregnancy/delivery</td>
<td>12-50</td>
<td>5-10</td>
</tr>
<tr>
<td>Breast milk</td>
<td>12</td>
<td>Not quantified</td>
</tr>
<tr>
<td><strong>Medical interventions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blood transfusions</td>
<td>&gt;90</td>
<td>3-5</td>
</tr>
<tr>
<td>Blood products</td>
<td>Not quantified</td>
<td>Not quantified</td>
</tr>
<tr>
<td>Organ transplantation</td>
<td>Not quantified</td>
<td>Not quantified</td>
</tr>
<tr>
<td>Artificial insemination</td>
<td>Not quantified</td>
<td>Not quantified</td>
</tr>
<tr>
<td><strong>Health care work (needlestick, etc.)</strong></td>
<td>0.1-1.0</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Source: Table reproduced from the work of Jay A. Levy [90].

![Figure 3.1: Possible Uses of Antiretroviral Drugs](image)
3.4.1 Post-Exposure Prophylaxis

Chemoprophylaxis is the administration of medication for the purpose of disease prevention. For example, cephalosporin is often prescribed as an antimicrobial prophylaxis prior to surgery to prevent surgical site infections [144]. For individuals who believe they were potentially exposed to HIV, PEP can reduce the probability of contracting HIV. Such incidents of exposure include needle sharing by IDUs, unprotected anal intercourse (UAI), as well as occupational needlestick [153]. However, PEP must be administered within 72 hours of exposure to HIV and continued for 28 days [153].

Occupational exposures to HIV, such as from a contaminated needle stick, have been treated through prompt administering of PEP. Outside of occupational hazards, PEP may be administered to combat HIV transmission due to suspected HIV exposure due to sexual activity or injection drug use [153]. Lessons learned from the introduction of PEP include the importance of widespread awareness, program adherence, and availability of the prophylactic option [49].

3.4.2 Pre-exposure Prophylaxis

The ability of PEP to substantially reduce HIV transmission inspired research into the possibility of using ARV pre-exposure in hopes of preventing the infection from establishing itself within a host. HIV PrEP are drugs administered to seronegative individuals to reduce their risk of acquiring an HIV infection. The administration of this drug regimen is typically targeted toward populations perceived to have a substantial risk of acquiring HIV. These populations include MSM, IDU, and seronegative partners of seropositive individuals. In 2011, MSM represented 53 percent of incident HIV infections in the United States [155]. The CDC recommends pairing a PrEP regimen with periodic STI testing and treatment, risk reduction counseling, as well as renal and liver function testing [155].

One such drug, the only one approved by the US FDA, is a once-daily oral medication from Gilead Sciences called Truvada, comprised of 200 mg emtricitabine (FTC) and 300 mg tenofovir disoproxil fumarate (TDF). While the time from initial PrEP dose to maximal protection is not known definitively, the CDC clinical guidelines report that with daily-dosing the medication reaches maximum intracellular concentrations after approximately 20 days in the blood, 7 days for rectal tissue, and 20 days for cervicovaginal tissues [52]. Common side effects of Truvada include gastrointestinal discomfort, dizziness, headaches, and rashes [135].

For this medication, Grant et al. observed a 44% reduction in HIV incidence for a study of 2499 MSM over 3324 person-years that encountered 100 HIV infections [73]. For the subjects with detectable study-drug levels, indicating consistent use, the authors calculated an adjusted relative reduction in HIV risk of 95% with a 95% confidence interval ranging from 70 to 99. Cremin et al.’s estimate of a 75% reduction in acquisition of infection per PrEP protected sex act falls within that confidence interval [53]. However, these authors emphasize that significant reductions in cost are
needed for PrEP to generate large reductions in HIV incidence [53].

Assuming a high degree of efficacy and adherence to PrEP, Abbas et al. projected that a PrEP program costing about $2.0 billion could prevent 3.2 million new HIV infections among heterosexual couples in sub-Saharan Africa over 10 years [1]. In the United States, Paltiel et al. estimated a $753 monthly cost for a TDF/FTC PrEP drug combination, such as Truvada, based upon the monthly wholesale price adjusted to account for Medicaid rebate programs and retail pharmacy dispensing fees [132].

Modeling the administering of PrEP to different at-risk population in New York City, Kessler et al. observed that prioritizing all MSM resulted in a 19% reduction in new HIV infections compared to a 15% reduction if only at-risk MSM are targeted [93]. Additionally, they noted that prioritizing all MSM produced a 79% preventive effect at 15% of the total cost.

As such statistics demonstrate, adherence to the drug regimen can have a significant impact on the risk reduction. However, the benefits of PrEP on preventing HIV transmission can be offset by an increase in risky sexual behavior or exacerbating conditions such as STIs. Reasons for failure to adhere to drug regimen include concerns over increase in unprotected sex, drug toxicity, viral resistance, and false representation in the media. For example, in 2005, Singh and Mills expressed grave concern over the misrepresentation of tenofovir-based PrEP trials by the media and activists [151]. For example, in 2004, the Cambodian Prime Minister halted a trial of PrEP for Cambodian sex workers. Next, in 2005, a similar trial was halted by the Cameroon Minister of Public Health. Protesters of the Cambodia trial expressed concern regarding perceptions of inadequate HIV counseling, lack of insurance coverage for affected participants, and questions of the long-term safety of the tenofovir drug.

3.4.3 Anti-Retroviral Therapy

ART is a combination of ARV drugs administered to inhibit HIV replication. These drugs are divided into five primary classifications: nucleoside reverse transcriptase inhibitors (NRTI), non-nucleoside reverse transcriptase inhibitors (NNRTI), protease inhibitors (PI), entry or fusion inhibitors, and integrase inhibitors [15]. NRTIs and NNRTIs both interfere with the reverse transcriptase used to transcribe HIV RNA into HIV DNA. PIs inhibits the segmentation of HIV genetic material into usable partitions. Entry or fusion inhibitors can prevent HIV from entering a cell through interfering with the receptor sites. Finally, integrase inhibitors prevent the integrase enzyme from embedding the HIV DNA into the host cell’s DNA.

In 1987, the United States approved azidothymidine (AZT) as the first drug for the treatment of HIV. AZT, also known as brand name Retrovir, belongs to the NRTI classification of ARV drugs. Since then, numerous drug and drug combinations have been developed and approved for ART use.

However, problems can arise at all parts of the treatment spectrum [64]. Table 3.2 summarizes many of the factors that can become barriers to treatment. At the society level, the population
must be made aware that HIV exists, how it is transmitted, and that treatment options exist. Of course, HIV testing and ART programs must be available, affordable, and accessible in the society. Affordability can vary depending on medication cost, insurance coverage, negotiated price plans with medication providers, and subsidies from the government or humanitarian organizations. At the personal level, a person must first be aware that they have an HIV infection and the effect that infection will have on their quality of life and life expectancy. The person must be aware that treatment options exist, want to initiate an ART regimen, and be able to afford the ART program. Cultural considerations may influence the societal acceptability or personal desire to seek medical treatment for HIV. Once in an ART program, a person must stay in the program (program adherence) and must take the medication prescribed (medication adherence). On the biological front, the medical personnel treating the infected person must be vigilant against ART toxicity, HIV becoming resistance to the ART drugs prescribed, and serious complications that can arise from co-infections with TB.

### Table 3.2: Factors Influencing ART Effectiveness

<table>
<thead>
<tr>
<th>Societal</th>
<th>Behavioral</th>
<th>Biological</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV/ART Education</td>
<td>HIV Infection Awareness</td>
<td>ART Toxicity</td>
</tr>
<tr>
<td>HIV Testing Availability</td>
<td>ART Awareness</td>
<td>ART Drug Resistance</td>
</tr>
<tr>
<td>ART Availability</td>
<td>ART Program Initiation</td>
<td>ART Drug Combination</td>
</tr>
<tr>
<td>ART Affordability</td>
<td>ART Program Adherence</td>
<td>Co-infection with TB</td>
</tr>
<tr>
<td></td>
<td>ART Medication Regimen Adherence</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cultural Considerations</td>
<td></td>
</tr>
</tbody>
</table>

Counter-intuitively, the presence of people on ART can maintain heightened HIV prevalence levels in a community. This elevated prevalence is explained by the increased survival time of individuals on ART. That is, instead of dying and reducing the HIV prevalence measure, people on ART are surviving and maintaining the HIV prevalence level of the community. However, the per-contact probability of transmission for a person on ART, assuming heterosexual sex, is estimated to be a 92% reduction compared to people not on ART [62]. The actual impact of ART on reducing transmission HIV transmission via anal sex is less clear. Although this decrease in risk is substantial, increased risky behavior has the potential to obviate the benefits of ART. For example, Mei et al. estimated that a 30% increase in risky behavior would eliminate the benefits of ART [122].

### 3.5 Risk Factors

In this section we discuss some of the factors that can increase or decrease the probability of transmission for HIV. Factors such as STIs are known to increase susceptibility or transmission.
Condom use and circumcision of the male insertive partner can decrease the probability of HIV or STI acquisition and transmission. Other behavioral risk factors which may impact the spread of HIV include the presence of concurrent relationships, versatile sexual roles, and seroadaptive behaviors; however, the impact of these risk factors is subject to debate.

### 3.5.1 Sexually Transmitted Infections

As mentioned previously, the genital viral load can greatly impact the probability of transmission during intercourse. STIs and inflammation are known to increase genital viral load [43]. Galvin and Cohen explain that the presence of certain STI can both increase susceptibility to contracting HIV or increase the probability of transmission [63]. For women, the authors note that bacterial vaginosis, herpes simplex virus, human papillomavirus, chlamydia, gonorrhea, genital ulceration, and vaginal discharge can increase HIV shedding and thus viral load. For men, the authors note gonorrhea, trichomonas vaginalis, cytomegalovirus, urethritis, and genital ulcer disease increase HIV shedding in semen. In Table 3.3, we provide a summary of the effects of the presence of various STIs as compiled by Bracher et al. [29].

**Table 3.3: Relative Increase of Susceptibility and Infectiousness Due to Presence of STIs**

<table>
<thead>
<tr>
<th>STI</th>
<th>Susceptibility to HIV when infected with STI</th>
<th>Infectiousness of HIV when infected with STI</th>
<th>Susceptibility of acquiring STI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gonorrhea</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Chlamydia</td>
<td>2.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacterial Vaginosis</td>
<td>1.5</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Syphilis</td>
<td>2.5</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Chancroid</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Herpes</td>
<td>2.75</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This table is a subset of a table of values compiled by Bracher et al. [29].

### 3.5.2 Concurrent Relationships

In 2009, to alleviate concerns of conflicting definitions of concurrency, the the Joint United Nations Programme on HIV/AIDS (UNAIDS) Reference Group defined concurrent relationships as “overlapping sexual partnerships where sexual intercourse with one partner occurs between two acts of intercourse with another partner” [80]. There is considerable debate on the degree to which concurrent relationships contribute to HIV epidemics. Morris et al. contends that concurrent relationships result in a faster growing and more pervasive epidemic [127]. Lurie and Rosenthal caution that
there is insufficient evidence to attribute significant HIV prevalence, pervasiveness, or transmission to concurrency and that theoretical transmission does not equate to actual transmission [114].

From a graph reachability perspective, concurrent relationships do have the potential to provide the HIV virus with greater access to hosts. Concurrent sexual contact networks also have the potential to bridge low-risk and elevated-risk groups. Given the relatively high transmission probability during PHI phase of HIV, there is a possibility that high-transmission PHI infections of HIV could spread through the sexual contact graph in waves. Goodreau et al. have demonstrated through simulation that concurrency coupled with the PHI period of HIV could explain a persistent HIV epidemic [70]. The authors note that removing either concurrency or the high-transmission period of HIV brings transmission levels below the level needed to sustain persistence.

To illustrate the potential effect of PHI and concurrency, consider an individual with \( N \) concurrent partners per year, and \( S \) number of episodes of sex per year, divided equally among the number of concurrent partners. We track whether HIV was transmitted to his partners only within the first 90 days of his infection. If HIV is transmitted, then we create an additional infection to track with \( N - 1 \) partners. All new infections have \( N - 1 \) uninfected partners and 1 infected partner, totaling \( N \) concurrent partners. Thus, we start with one node of degree \( N \) and as infections are added, add additional at risk nodes of degree \( N - 1 \), developing a tree of infections and at-risk individuals.

We assume a 6.64\% probability that HIV will be transmitted per sexual contact during a 90 day PHI period. The probability represents an average of the uncircumcised insertive and receptive probability of transmission [92] multiplied by the risk increase factor of 8 for PHI stage HIV [62].

Episodes of sex for each relationship are spaced equally with the duration provided in Equation 3.1. To account for variability in relationship start times, we initialize the date of last sex for each relationship pair to a random number between \( t - 0 \) and \( t - D \), where \( t \) is the current timestep. Since the progression is based upon the probability of transmission this is a stochastic process. To account for stochastic variation, we repeat the simulation 1000 times and average the results.

\[
D = \frac{365}{S} \quad (3.1)
\]

In Figure 3.2a, we display the results of simulating HIV transmission during PHI assuming 50 annual sex acts over the course of 2 years. We repeat the test for different degrees of concurrent partnerships. As the diagram shows, there is a slight increase in the number of infections transmitted due to concurrency. However, if we double the number of sex episodes and repeat the tests, we see a more dramatic change. Figure 3.2b shows the results of simulating the HIV transmission during PHI assuming 100 annual sex acts. As this figure demonstrates, the number of HIV transmissions becomes exponential as the number of concurrent partnerships increases. This suggests there is a critical ratio of concurrent partnerships to number of sexual acts that can produce exponential growth.
Figure 3.2: Simulating the Effects of PHI Transmission and Concurrency
However, as compelling as this test may be, it represents a highly simplified view of a sexual network without condom use. There are numerous factors that can limit the spread of what may otherwise have been an exponential growth due to PHI and concurrency. For example, not every person in a sexual network will have sex at the same rate and with the same number of concurrent partners.

Additionally, the above test assumed a uniform average of HIV transmission probability and sexual role differences. In Figure 3.3 we repeat the above tests, but alternate sexual roles between transmission and use sexual role specific probability of HIV transmission during PHI. Thus, if the infected partner is the insertive role, then his partners will be a receptive role and be subject to the risk of HIV transmission for the receptive role. We utilize a 8.32% probability for transmission to a receptive sexual role and a 4.96% probability for transmission to an insertive role. These values represent a factor 8 increase due to PHI from the base sexual role probabilities of transmission for uncircumcised individuals [92]. For the first infection, we assume his role is insertive.

As Figure 3.3a demonstrates, compared to Figure 3.2a, there is only a slight decrease due to alternating sexual roles and probabilities of transmission. In the case of 100 sexual contacts per year, Figure 3.3b demonstrates a greater decrease, but still produces similar exponential growth shapes as Figure 3.2b.

Our tests demonstrate the theoretical possibility that concurrency can increase the number of HIV transmissions during PHI. However, there exist limiting factors of HIV spread due to social network saturation with HIV, behavioral variations, and sexual contact network structure. Thus, this section has not established whether or not an actual society has produced exponential HIV growth due to the confluence of PHI and concurrency.

3.5.3 Versatility

A person’s sexual role is an important factor in the assumption of risk. In the context of sexual roles, versatility refers to the ability of a person to engage in both insertive and receptive roles. A versatile person is exposed to both the relatively lower risk of contracting HIV from an insertive role and the relatively higher risk of contracting HIV from a receptive role. Naturally, this practice does not apply to heterosexual couples as the female is always the receptive partner and the male is always the insertive partner. However, for MSM, such role-reversals are possible and can allow HIV to spread by overcoming the low probability of HIV transmission from receptive to insertive partners [26].

Beyrer et al. concluded that “the ability of MSM to be role-versatile also predisposes them to large epidemics” [26]. In the absence of versatility, the authors’ model reduced HIV incidence in MSM by 19 to 55%. Similarly, Goodreau et al. observed the potential for a doubling of HIV prevalence under fully versatile unprotected sexual practices [69]. A fully versatile role is where a person is equally likely to engage in both insertive and receptive roles during sex.
Figure 3.3: Simulating the Effects of PHI Transmission and Concurrency with Alternating Sex Roles

(a) 50 Sexual Episodes Annually

(b) 100 Sexual Episodes Annually
3.5.4 Condoms

When used effectively, condoms can greatly decrease the probability of transmission of STIs. However, the effectiveness of condoms depends on their availability, cost, correct use, and proper storage. The choice to use condoms can be affected by religious doctrine, peer pressure, personal preferences, and impaired judgment due to drug or alcohol use. Condom durability can be compromised by improper storage, prolonged duration of sex, and using an inappropriate lubricant. Condom slippage and failure may be attributed to inexperienced use and use of an incorrect size.

Davis and Weller report heterosexual condom efficacy at 87% [56]. In 1989 Detels et al. reported on the only large longitudinal study of gay condom use which tracked 2915 HIV seronegative men for up to 2 years [57]. They concluded that condom use for anal sex among MSM yielded 70% efficacy. Golombok et al. 1 condom failure per 27 usages [67]. In 1997, Silverman and Gross reviewed literature on anal condom use and found rates of condom breakage between 0.5% to 6% and rates of condom slippage at 3.8% to 5% [150].

In 1999, Stone et al.’s study of 2592 MSM participants reported a condom failure rate for MSM anal intercourse of 16.6 percent [157]. This figure includes both incidents of slippage and breakage. A study of 475 MSM in Atlanta, Georgia by Romieu et al. found condom failures and incomplete condom use to be common among MSM with only 36 percent reporting effective use of condoms despite 82 percent condom use [85]. In addition to condom breakage and slippage, the use of oil lubricants and improperly fitting condoms were reported as contributing factors.

Understandably, Thompson et al. noted that infrequent or inexperienced use of condoms correlated with a higher risk of condom failure [164]. For MSM in the study who only had sex once in the study year, they reported a failure rate of 15%. More experience resulted in a significant decline in condom failure, dropping the failure rate below 1% for men who had used a condom more than 10 times in the study year.

In contrast, Prater’s survey of 944 MSM found that for every increase in the number of male sexual partners by 10, the probability of condom failure increased by 5% [137]. Prater attempts to explain this figure by citing possibly correlations between people who have many sex partners and incidents of vigorous sex or use of inappropriate lubricant types. Prater notes the contradiction with Thompson et al.’s above work [164] and explains that the number of sexual partners does not necessarily correlate with actual sexual experience.

3.5.5 Circumcision

In randomized controlled trials in Sub-Saharan Africa, circumcision has been found to decrease the likelihood of HIV transmission from the female to male partner by 50 to 60 percent [17, 19, 74]. Grulich et al. note that circumcision is believed to protect against transmission due to the HIV receptors in foreskin Langerhans cells, the vulnerability of foreskin to epithelial damage during intercourse, and the risk of ulcerative infections [81].
However, although circumcision has been demonstrated to reduce transmission in HIV discordant heterosexual couples, the effect of circumcision in HIV transmission for MSM couples is not as clearly established. In 2008, Millet et al.'s literature review concluded that while there is a protective though non-significant relationship between circumcision and HIV infection, the benefits of circumcision may only protect against HIV infection if the person primarily or exclusively takes an insertive role during unprotected anal intercourse [123]. More recently, in 2010 Jin et al.'s study on 1427 homosexuals in Sydney, Australia resulted in the per-contact probabilities in Table 3.4 that demonstrated nearly a 6 fold increase in the probably of transmission for insertive role uncircumcised partners compared to circumcised partners [92]. However, the confidence interval on these values is very wide.

Table 3.4: Sydney MSM Per-Contact HIV Transmission Probability

<table>
<thead>
<tr>
<th>Role</th>
<th>Per-act probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insertive partner (circumcised)</td>
<td>0.11%</td>
</tr>
<tr>
<td>Insertive partner (uncircumcised)</td>
<td>0.62%</td>
</tr>
<tr>
<td>Receptive partner (with ejaculation)</td>
<td>1.43%</td>
</tr>
<tr>
<td>Receptive partner (without ejaculation)</td>
<td>0.65%</td>
</tr>
</tbody>
</table>

3.5.6 Seroadaptation

Hoping to reduce personal risk, individuals adapt their behavior based upon their HIV status and the perceived HIV status of their partners. This practice is commonly referred to as serosorting, though Le Talec and Jablonski argue that seroadaptation would be more accurate encompassing terms for this set of behavior [107]. In serosorting, individuals will preferentially seek sex with the people of the same HIV status. Other seroadaptation or seropositioning behavior includes selectively using condoms or limiting sexual roles with serodiscordant partners [66].

Seropositioning involves choosing a sexual role in a serodiscordant pairing based upon the risk of HIV transmission for the sexual role [133]. For example, an HIV negative individual may choose oral or insertive sex in lieu of the higher risk of receptive anal sex. Alternatively, an individual may insist on condom use for serodiscordant partners. All seroadaptive methods are intended to prevent or reduce risk in serodiscordant partnerships that could result in HIV transmission. However, the effectiveness of these practices depend on the frequency of HIV testing, self-awareness of one’s HIV status, individuals asking their partners for their HIV status, and accurate HIV disclosure from both partners. Simulating HIV transmission, Kok determined that in societies with a short time to diagnosis, serosorting decreases HIV prevalence [99]. However, in societies with times to diagnosis greater than 3.5 years, serosorting increases HIV prevalence. While this simulation may be a simplified scenario, this inflection point suggests that the effectiveness of such seroadaptive behaviors may vary regionally based upon the society’s frequency of HIV testing.
Wilson et al. cautioned against endorsing serosorting as the practice can lead to increased HIV transmission in populations of MSM where the undiagnosed HIV prevalence is relatively high [182]. That is, if a person is unaware that he is HIV positive, the practice of serosorting may lead him to form relationships with HIV negative individuals allowing onward transmission of HIV. An individual’s recent HIV negative test result may erroneously be perceived to indicate the absence of an HIV infection. However, a recently introduced HIV infection in the PHI phase is both highly virulent and undetectable through the commonly used antibody tests. Butler and Smith note that as the proportion of recent HIV infections increase in the population the effectiveness of serosorting diminishes [35].

However, in the case of the Amsterdam Cohort Studies dataset, Van der Bij et al. maintain that variations of condom use, not differing rates of serosorting behavior accounts for the rising HIV incidence observed in 2004 to 2006 [172]. They noted a higher degree of UAI among anonymous or nonconcordant partners to which they attributed the rising incidence.
CHAPTER 4
MASHIV OVERVIEW

In this chapter we discuss the components of the Multi-Agent System of HIV (MASHIV) simulator software and compare the HIV Model used in the simulator to the Spectrum model of HIV progression and mortality [62]. MASHIV is a simulator we developed to model HIV transmission among MSM. The simulator has been developed in the JAVA programming language using the JFreeChart graphing package for charts [130]. The simulator has been developed from an object-oriented perspective, encapsulating related variables into objects representing people, relationships, and instances of HIV infections. Additionally, for a given random seed, the simulator will always produce the same results. This stochastic determinism allows multiple users to experience the same result on the same dataset. The user may save and restore the prepared dataset via an XML file. We utilize JAVA’s XML serialization library for saving datasets and results to XML files.

A MAS is a decentralized approach to solving a problem or simulating behavior. A MAS simulation is comprised of agents, available agent actions, rules guiding action use, and an environment. Agents perceive each other and the environment, using their perceptions and personal history to decide on a course of action for accomplishing their goals. An agent’s action may result in changes to the environment, changes to the agent’s internal state, and changes to other agents’ internal states. Agents in a MAS simulation work toward their goals without centralized control allowing global consequences to arise from local decisions among individual agents.

In our multi-agent system simulation, agents represents men who have sex with men (MSM). Agent actions include evaluating another agent for a relationship, forming a relationship, ending a relationship, having sex with another agent, and mitigating risk of contracting HIV. As such, an agent’s goal is to form sexual relationships while minimizing the personal risk of contracting HIV. An optional secondary goal is to minimize risk of infecting a steady partner with HIV. However, as our agents are primarily reactive agents, their ability to minimize such risks is deterministic. While complex agents might invoke a cognitive model of reasoning for minimizing their risk to HIV, our reactive agents respond based upon a set of predefined behaviors.

When two agents engage in sex, HIV may be transmitted from one agent to another. We determine the probability that HIV was transmitted between the agents based upon the presence of HIV, risk reduction factors, and risk exacerbating factors. For example, successful condom use can reduce the probability of transmission whereas a PHI stage HIV infection can increase the risk. A uniform sample is used to translate the calculated probability of transmission into a reality, determining whether HIV is transmitted between the agents. HIV is modeled on an individual basis with each infected agent having its own progression to mortality based upon a CD4 model.

The MASHIV simulator is comprised of four major components—a parameter set described in Section 4.1, the HIV model described in Section 4.2, the agent model described in Section
The parameter set is comprised of the user-specified values that define the population, HIV prevalence, interventions being tested, dating behavior, and simulation settings. The HIV model controls the transmission, progression, and mortality of HIV infections. The agent model governs agent interactions such as dating, relationship formation, HIV interventions, and sexual conduct. The GUI is used to define the parameter set, initialize simulations, and observe the collected statistics.

4.1 Parameter Set

A parameter set defines a simulation and is comprised of operational parameters in Table 4.1, population parameters in Table 4.2, and group-specific parameters. The minimal group parameters, defined in Table 4.3 are the lower bound on age, the upper bound on age, and the size of the population. The ages are distributed uniformly between the minimum age in days and the maximum age in days, exclusive of the maximum age. To allow for a greater diversity of datasets that can be used in our simulator, we allow the user to migrate all population parameters into group-specific parameters. As such, the simulator is capable of being instantiated with a simple dataset comprised of global population parameters, or instantiated with an exceptionally detailed dataset where each group has its own set of defined population parameters. This feature allows users with both sparse and comprehensive literature resources to run simulations matching their depth of detail.

Operational parameters define major settings that govern the initialization of the simulator. Behavioral parameters define initializing parameters and behavioral parameters. Population initialization parameters include items such as the number of HIV infections, the percent of the population that are circumcised, and the sexual roles. Population behavioral parameters include parameters defining an agent’s relationship maintenance, sexual conduct within relationships, and risk reduction strategies. Group-specific parameters define population groups. At a minimum, the group parameters include an age range and population size. However, as noted above, population parameters can be migrated into group-specific parameters.

As noted in Table 4.2, most population parameters are distributable. That is, the user may define either a fixed value, a normal distribution, or a uniform distribution for the values. Fixed values, of course, only have a single value. Normal distributions are defined with a mean and standard deviation. Uniform distributions are defined as a mean value and a difference from the mean to sample uniform values in both positive and negative directions. While this definition of a uniform distribution range differs from the standard parameters (minimum and maximum), using a mean and difference allows for an equivalent definition that is easier to vary for aggregate runs. Distributed parameters are only sampled at initialization to fix individual agent’s behavioral constants. Distribution applies to variables where there is a possibility of individual variation for the parameter value. For example, the preferred number of parameters would vary from person to person, and thus can be based on a distribution. On the other hand, parameters that result in a
boolean assignment, such as circumcision, are not distributable.

Table 4.1: Operational Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>numYears</td>
<td>Integer</td>
<td>The number of years to simulate, based upon one iteration per day.</td>
</tr>
<tr>
<td>seed</td>
<td>Long</td>
<td>The random seed.</td>
</tr>
<tr>
<td>probabilisticTransitions</td>
<td>Boolean</td>
<td>True if probabilistic CD4 group transitions are enabled; False if duration-based CD4 group transitions are enabled.</td>
</tr>
<tr>
<td>probabilisticDating</td>
<td>Boolean</td>
<td>True if dating is based on daily probability of formation, False if dating is based upon strict adherence to durations.</td>
</tr>
<tr>
<td>symptomaticDetection</td>
<td>Boolean</td>
<td>True if HIV can be discovered through symptomatic PHI; False otherwise.</td>
</tr>
<tr>
<td>replaceRemoved</td>
<td>Boolean</td>
<td>True if dead agents and agents aged 65+ are replaced with 15 year old demographic duplicates, False if dead agents are merely removed.</td>
</tr>
<tr>
<td>steadyUAI</td>
<td>Boolean</td>
<td>True if once a steady partnership has UAI, all future sex will be UAI; False otherwise.</td>
</tr>
</tbody>
</table>

Each agent in the population has a preferred integer number of concurrent steady relationships. Parameter $R_C$ defines the percentage of a population desiring concurrent relationships, that is, more than one relationship at a time. A random uniform sample on the percentage $R_C$ is used to determine whether the agent will have more than one relationship concurrently or practice serial monogamy. For the case of serial monogamy, the agent will only have 1 relationship at a time. For the case of concurrent relationships, we use the value or distribution defined by the $N_S$ parameter to determine the number of preferred concurrent relationships for the agent with a minimum of 2.

An agent has a separate time track for each of their preferred number of concurrent steady relationships. In a track, relationships will form and end without overlap. Concurrent relationships exist when a relationship in one track overlaps, in time, a relationship in a different track.

Each agent has his own preferred duration for steady relationships defined at initialization by the parameter or distribution $D_S$. Each day we check to see if the relationship should end. In testing, we discovered that asking both agents each day whether they want to end the relationship produced a double jeopardy problem causing the relationships to terminate prematurely. To alleviate this problem, we alternate which agent we ask each day if the relationship should end. This allows both agents’ relationship duration preference to influence the end of the relationship.

For duration-based dating, the length of the steady relationships are fixed for each agent. An agent will want to end the relationship exactly on their preferred relationship length. Thus, relationships will end on the minimum value of the preferred relationship lengths of the two agents.
Table 4.2: Population Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Distributable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HIV &amp; HIV Risk</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N_H$</td>
<td>No</td>
<td>The number of initial HIV infections.</td>
</tr>
<tr>
<td>$P_{PHI}$</td>
<td>No</td>
<td>Percent of the initial HIV infections that are PHI stage.</td>
</tr>
<tr>
<td>$P_{Circ}$</td>
<td>No</td>
<td>The percent of the population that is circumcised.</td>
</tr>
<tr>
<td>$P_{NoTest}$</td>
<td>No</td>
<td>The percent of the population that never tests for HIV.</td>
</tr>
<tr>
<td>$T_{HIV}$</td>
<td>Yes</td>
<td>The number of HIV tests per year.</td>
</tr>
<tr>
<td>$SS_+$</td>
<td>Yes</td>
<td>The percentage of serosorting for an HIV+ person.</td>
</tr>
<tr>
<td>$SS_-$</td>
<td>Yes</td>
<td>The percentage of serosorting for an HIV- person.</td>
</tr>
<tr>
<td>$K_D$</td>
<td>Yes</td>
<td>The percentage decrease in unprotected sex following discovery of HIV.</td>
</tr>
<tr>
<td>$K_L$</td>
<td>Yes</td>
<td>The length of time in years that $K_D$ is in effect.</td>
</tr>
<tr>
<td>$P_A$</td>
<td>Yes</td>
<td>The percentage of PrEP adherence for the population.</td>
</tr>
<tr>
<td>$P_C$</td>
<td>No</td>
<td>The percentage of PrEP coverage for the population.</td>
</tr>
<tr>
<td>$P_{ART}$</td>
<td>Yes</td>
<td>The annual probability of an agent with a known infection to initiate ART.</td>
</tr>
<tr>
<td><strong>Relationships</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_C$</td>
<td>No</td>
<td>The percentage of the population having concurrent relationships.</td>
</tr>
<tr>
<td>$N_S$</td>
<td>Yes</td>
<td>The preferred number of concurrent steady relationships.</td>
</tr>
<tr>
<td>$D_S$</td>
<td>Yes</td>
<td>The mean duration of steady relationships.</td>
</tr>
<tr>
<td>$S_S$</td>
<td>Yes</td>
<td>The annual target number of sex acts for all steady relationships.</td>
</tr>
<tr>
<td>$N_C$</td>
<td>Yes</td>
<td>The number of 1-day casual sexual encounters each year.</td>
</tr>
<tr>
<td><strong>Sexual Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_R$</td>
<td>No</td>
<td>The percent of population preferring a receptive sexual role.</td>
</tr>
<tr>
<td>$R_I$</td>
<td>No</td>
<td>The percent of population preferring an insertive sexual role.</td>
</tr>
<tr>
<td>$R_V$</td>
<td>No</td>
<td>The percent of population preferring a versatile sexual role.</td>
</tr>
<tr>
<td>$P_{Always,Steady}$</td>
<td>No</td>
<td>The percent of the population that always uses a condom in steady relationships.</td>
</tr>
<tr>
<td>$P_{Always,Casual}$</td>
<td>No</td>
<td>The percent of the population that always uses a condom in casual relationships.</td>
</tr>
<tr>
<td>$P_{Never}$</td>
<td>No</td>
<td>The percent of the population that never uses a condom.</td>
</tr>
<tr>
<td>$C_S$</td>
<td>Yes</td>
<td>The percent of condom use for steady relationships.</td>
</tr>
<tr>
<td>$C_C$</td>
<td>Yes</td>
<td>The percent of condom use for casual relationships.</td>
</tr>
</tbody>
</table>
Thus, in a simple relationship model, the relationships in each relationship track would line up neatly together without gaps. For probabilistic dating, both the initiation and the length of the steady relationships are determined by normal hazard functions.

Sexual acts for steady relationships are defined annually with the parameter $S_S$ and are divided equally among the different steady relationship tracks with a minimum of one per track. We define an average length of time between sex as $L_S = 365/(S_S/N_S)$. For duration-based dating, sex will occur in relationships at equal divisions of $L_S$ in each track. For probabilistic dating, we use a normal hazard function for determining when sex occurs using $\mu = L_S$ and $\sigma = \mu/2.0$.

In addition to the steady relationship tracks, there is a single relationship track for casual sexual encounters. Casual relationships are meant to model fleeting sexual encounters with only one sexual act. In our model, casual relationships exist for a minimum of 1 day and cease to exist on the day after the first instance of sex.

### Table 4.3: Minimal Group Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_L$</td>
<td>The lower bound on the age for the population. (Inclusive)</td>
</tr>
<tr>
<td>$A_U$</td>
<td>The upper bound on the age for the population. (Exclusive)</td>
</tr>
<tr>
<td>$N$</td>
<td>The size of the population.</td>
</tr>
</tbody>
</table>

## 4.2 HIV Model

In this chapter, we define the parameters governing HIV transmission, disease progression, and HIV-related mortality. We discuss the adaptation of these parameters to daily probabilities and present an analysis of the resulting HIV model. For the purpose of our simulation, we define serodiscordant couples to be the pairing of one seropositive individual and a seronegative individual. Seroconcordant couples are defined as either both seronegative partners or both seropositive partners.

### 4.2.1 HIV Initialization

We initialize the HIV prevalence of the population according to a user-specified level for the first year. The user may specify either global HIV prevalence or sub-population HIV prevalence. If the user specifies global HIV prevalence, $N_H$, then the number of HIV infections specified are distributed uniformly, without duplication, to the entire population of agents. If the user specifies population level HIV prevalence, then the number of HIV infections in the population $N_{H_i}$ are
distributed uniformly, without duplication, to only the sub-population of agents $N_i$. If the “initialHIVareAsym” parameter is false, then all of the initial HIV infections are started at the PHI phase of the HIV infection; if true then all of the initial HIV infections for our model are started at the first day of the asymptomatic phase of the HIV infection. Using initial infections that are all PHI is useful if the number of infections is low and one wishes to study the possibility of an epidemic becoming established. However, due to the short duration of PHI, the actual proportion of HIV infections existing in the PHI is likely small. Thus, initializing all initial HIV infections to the asymptomatic stage is an approximation of the distribution of HIV infection stages. Under probabilistic transitions, although the initial asymptomatic infections all start at the same time, their progression will soon diverge from uniformity.

4.2.2 HIV Transmission

The probability of sexual HIV transmission is affected by many factors. For our model, these factors include the stage of HIV infection, circumcision of the insertive partner, condom use, age group, sexual role, ART use, and PrEP use. For our simulations we assume that there is only one strain of HIV in the population. As such, seroconcordant partners, partners with the same HIV status, cannot transmit HIV to each other.

There are three stages of the HIV infection—primary HIV infection (PHI), an asymptomatic period, and progression toward AIDS. The effect of PHI and AIDS stages on the probability of transmission are defined as multiplicative factors relative to the baseline probability of transmission during the asymptomatic period. Compared to the asymptomatic stage per-contact transmission risk, Wawer et al. estimated a 8 to 12 fold greater probability of transmission during PHI; 4 to 8 fold greater during AIDS [176]. However, the default value for the Goals model PHI relative infectiousness of PHI for countries include values ranging 8 to 22 and higher [25]. Hollingsworth et al. criticize Wawer’s methodology, deriving instead a 26 fold increase in infectiousness during PHI and 7 fold increase during AIDS [87]. The authors emphasize that the contribution of the stages of infection to HIV transmission are dependent not just upon the relative infectiousness of each stage, but also the sexual behavior of the population and variations in stage duration.

For our simulation, we are modeling PHI over a period of 3 months in a simplified manner assuming a uniform increase in the daily probability of transmission throughout the duration of PHI. As such, we are more interested in an average number than the peak infectiousness during PHI, which could vary between individuals.

The probability that a seronegative agent $A_j$ will contract HIV from a seropositive agent $A_i$ at timestep $t$ can be summarized in the Equation 4.1. The probability of transmission is a product of the condom risk reduction cofactor $M_C$, the transmission risk of the seropositive partner $T(A_i)$, and the reception risk of the seronegative partner $R(A_j)$.

$$P(t, A_i, A_j) = M_C \times T(A_i) \times R(A_j)$$ (4.1)
Condoms are estimated to prevent the onward transmission of HIV approximately 82 percent of the time [85]. Using that percentage, a Bernoulli sample determines whether the condom failed as described in Equation 4.2. If the condom did not fail, then we assume there is a zero percent chance of HIV transmission between a serodiscordant couple for that sexual act. If a condom did fail, then it does not effect the probability of transmission. Thus, the reduction cofactor for condoms, \( M_C \), is 0 when the condom is successful, and 1 when the condom breaks with a probability of success \( R_C \). If the condom is successful, we assume there is zero chance of HIV transmission. We use a Bernoulli trial to model the success of the condom. A Bernoulli trial is a simple random experiment which returns either a success or a failure result. Success occurs with a fixed probability \( p \). For a given probability \( p \), a Bernoulli trial is equivalent to performing a uniform random sample \( x \) from the interval \([0,1]\) and returning “true” if \( x < p \). For our equations, we interpret a success result as a 1 value and a failure result as a 0 value. Equation 4.3 defines the Bernoulli function we use to perform a Bernoulli trial.

\[
M_C = 1 - \text{Bern}(R_C) \quad (4.2)
\]

\[
\text{Bern}(p) = \begin{cases} 
1 & : x \leq p \quad x \sim \text{Uniform}(0,1) \\
0 & : x > p
\end{cases} \quad (4.3)
\]

The transmission risk, Equation 4.4, is a product of the probability of transmission for the seronegative partner’s sexual role \( P(A_j) \), the HIV stage cofactor for the seropositive partner \( A_i \), and the transmission reduction cofactor for ART.

\[
T(A_i) = S(A_i) \times Art(A_i) \quad (4.4)
\]

Due to the low probability of oral HIV transmission, for simulation we only consider HIV transmission via URAI and unprotected insertive anal intercourse (UIAI). We assume a per-contact transmission risk of 0.5 percent for URAI and 0.065 percent for UIAI [42].

The reception risk, Equation 4.5, is a product of the PrEP reduction factor \( R_{PrEP} \) and the circumcision reduction factor \( C_i \).

\[
R(A_j) = P(A_j) \times \text{PrEP}(A_j) \times \text{Circ}(A_j) \quad (4.5)
\]

If the seronegative partner \( A_j \) is in the insertive sexual role, then circumcision has the potential to reduce the probability of contracting HIV from agent \( A_i \). Obviously, if \( A_j \) is in the receptive role, then circumcision would provide \( A_j \) no benefit. Equation 4.6 accounts for the effect of reduced risk due to circumcision of the insertive partner. The equation returns 1 if \( A_j \) is in the receptive role, implying no reduction in risk. Otherwise, the equation returns the multiplicative risk reduction factor \( R_{Circ} \).
Equation 4.7 defines the reduction of risk of transmission for the seropositive individual. If the individual $A_i$ is not on ART, then the function returns 1, implying no effect from ART. If the individual $A_i$ is on ART, then the function returns $(1 - R_{ART})$, producing the multiplicative reduction factor for ART.

$$\text{Art}(A_i) = \begin{cases} 
1 & : A_i \text{ is not on ART.} \\
(1 - R_{ART}) & : A_i \text{ is on ART.}
\end{cases} \quad (4.7)$$

The stage of the HIV virus of agent $A_i$ impacts the probability of transmission. Equation 4.8 accounts for increased virulence of PHI and AIDS stage by returning multiplicative factors. If $A_i$ is in PHI stage, the equation returns $M_{PHI}$, which is currently set to 8. As the numbers for probability of transmission assume asymptomatic stage transmission rates, this function returns 1 if $A_i$ is in the asymptomatic stage. For the AIDS stage, the function returns $M_{AIDS}$, which is currently set to 4.

$$S(A_i) = \begin{cases} 
M_{PHI} & : A_i \text{ is in the PHI stage of infection.} \\
1 & : A_i \text{ is in the asymptomatic stage of infection.} \\
M_{AIDS} & : A_i \text{ is in the AIDS stage of infection.}
\end{cases} \quad (4.8)$$

Since we are calculating the risk to seronegative agent $A_j$ from seropositive agent $A_i$ in this equation, we are concerned with the probability of transmission based upon the sexual role of $A_j$. Equation 4.9 defines the sexual role risk for agent $A_j$.

$$P(A_j) = \begin{cases} 
P_I & : A_j \text{ is the insertive sexual role.} \\
P_R & : A_j \text{ is the receptive sexual role.}
\end{cases} \quad (4.9)$$

If agent $A_j$ is on PrEP, the risk of $A_i$ transmitting HIV to agent $A_j$ is further reduced by the $PrEP(A_j)$ value from Equation 4.10. However, this requires the agent $A_j$ to have taken his PrEP medication. We assume that there is zero reduction in risk if the agent’s time of last PrEP dose $t_D$ is more than 1 day away from the current timestep $t$. Hence, the function returns a reduction factor of 1 if the agent is not on PrEP or if the last dose of PrEP is more than 1 day away. Otherwise, the function returns the multiplicative risk reduction factor for PrEP $1 - R_{PrEP}$.

$$\text{PrEP}(A_j) = \begin{cases} 
1 & : A_j \text{ is not on PrEP} \\
1 & : A_j \text{ is on PrEP, and } (t - t_D) \geq 1 \\
1 - R_{PrEP} & : \text{is on PrEP, and } (t - t_D) < 1
\end{cases} \quad (4.10)$$

Table 4.4 defines the HIV transmission and risk reduction parameters used in the equations above. The reduction factor for circumcision, $R_{Circ}$ is calculated by taking the ratio of uncircumcised UAI to circumcised UAI from Jin et al.’s paper [92].
Table 4.4: HIV Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Definition</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_I$</td>
<td>0.0062</td>
<td>Probability of transmission per unprotected anal insertive act with asymptomatic HIV</td>
<td>[92]</td>
</tr>
<tr>
<td>$P_R$</td>
<td>0.0104</td>
<td>Probability of transmission per unprotected anal receptive act with asymptomatic HIV</td>
<td>[92]</td>
</tr>
<tr>
<td>$M_{PHI}$</td>
<td>8</td>
<td>The factor by which the probability of transmission is increased during the Primary HIV infection (PHI) stage.</td>
<td>[25]</td>
</tr>
<tr>
<td>$M_{AIDS}$</td>
<td>4</td>
<td>The factor by which the probability of transmission is increased during the AIDS stage.</td>
<td>[25]</td>
</tr>
<tr>
<td>$R_C$</td>
<td>0.82</td>
<td>Proportion of condom success (didn’t slip or break).</td>
<td>[185]</td>
</tr>
<tr>
<td>$L_{PHI}$</td>
<td>90 days</td>
<td>The duration of the primary HIV infection (PHI) stage.</td>
<td>[25]</td>
</tr>
<tr>
<td>$R_{Pr,EP}$</td>
<td>0.75</td>
<td>The proportion reduction in risk of receiving HIV by full adherence to pre-exposure prophylaxis (PrEP).</td>
<td>[53]</td>
</tr>
<tr>
<td>$R_{ART}$</td>
<td>0.92</td>
<td>The proportion reduction in risk of transmitting HIV by full adherence to anti-retroviral therapy (ART).</td>
<td>[62]</td>
</tr>
<tr>
<td>$R_{Circ}$</td>
<td>0.11/0.62</td>
<td>The proportion reduction in risk of transmitting HIV when circumcised.</td>
<td>[92]</td>
</tr>
</tbody>
</table>

From the above equations and the parameter values in Table 4.4, we produce the table of per sexual contact HIV transmission probabilities in Table 4.5. The columns in the table represent the viral stage and ART status of the HIV-positive partner whereas the rows represent the configuration of the HIV-negative partner. A value in the table signifies the per sexual contact probability that HIV will be transmitted from the HIV-positive partner to the HIV-negative partner.

These probabilities are the probability of transmission without the use of a condom. Thus, for an insertive HIV-negative partner the probabilities are UIAI. For a receptive HIV-negative partner the probabilities are URAI. We omit rows in this table for configurations which do not produce a change in the probability values. For example, the probability of transmission during URAI is not affected by the circumcision status of the receptive partner.
Table 4.5: Probabilities of HIV Transmission

<table>
<thead>
<tr>
<th></th>
<th>HIV+ Partner not on ART</th>
<th>HIV+ Partner on ART</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PHI</td>
<td>Asymp.</td>
</tr>
<tr>
<td>Insertive HIV- Partner</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UIAI</td>
<td>0.04960</td>
<td>0.00620</td>
</tr>
<tr>
<td>w/PrEP</td>
<td>0.00992</td>
<td>0.00124</td>
</tr>
<tr>
<td>w/Circumcision</td>
<td>0.00880</td>
<td>0.00110</td>
</tr>
<tr>
<td>w/Circumcision &amp; PrEP</td>
<td>0.00176</td>
<td>0.00022</td>
</tr>
<tr>
<td>Receptive HIV- Partner</td>
<td></td>
<td></td>
</tr>
<tr>
<td>URAI</td>
<td>0.08320</td>
<td>0.01040</td>
</tr>
<tr>
<td>w/PrEP</td>
<td>0.01664</td>
<td>0.00208</td>
</tr>
</tbody>
</table>

4.2.3 Disease Progression & Mortality

From the time of infection, the agent begins his progression toward mortality. The HIV progression is divided into three stages— the primary infection, the asymptomatic period, and symptomatic progression toward AIDS death. Subsequent to HIV being introduced to a body’s bloodstream, the PHI phase begins. This period is marked by a rapid ascent to a peak of high viral load and heightened probability of transmission. As the body develops antibodies to combat the virus, the viral load drops considerably marking the end of the PHI and the beginning of the relatively long asymptomatic period. HIV exploits the CD4 cells of the host’s immune system to replicate leading to an eventual collapse of the immune system and thus resulting in susceptibility toward other infections and certain cancers. The final stage, AIDS, is defined by a CD4 level of 200 or below.

Our model treats the PHI as a fixed length of time lasting 90 days [62]. We assume that no one dies of HIV-related causes during the PHI period. After PHI, we utilize a CD4 compartment model to govern HIV progression and mortality. Our CD4 compartment model is an adaptation of the CD4 compartment model developed by Stover et al. for use in Spectrum and the Estimates and Projections Package (EPP) [158]. The UNAIDS Reference Group on Estimates, Modelling and Projections guides the development of Spectrum and EPP which are bundled together and often used in conjunction to model the course and consequences of HIV epidemics. EPP is a curve fitting model that fits a prevalence trend to a set of surveillance data. Spectrum utilizes EPP to estimate country-specific HIV incidence values. Spectrum provides a wider system of policy models to provide analytical tools necessary to evaluate, plan, and advocate for health programs [62]. Spectrum provides a full age and gender structured model of HIV. Spectrum is a key component in the UNAIDS goal of providing regular HIV estimates for all countries with populations exceeding 250,000.

For modeling HIV progression, John Stover created the CD4 compartment model, which was later incorporated into EPP. The CD4 compartment model is comprised of CD4 group compart-
ments, inter-compartment transition rates, mortality rates specific to each compartment, as well as a separate set of compartments and mortality rates for individuals on ART [158]. Each compartment represents a specific range of CD4 cell counts. For example, compartment “350-500” represents individuals having between 350 and 500 CD4 cells/mm$^3$ of blood. Compartments representing CD4 counts above 200 represent asymptomatic stage HIV infection; CD4 counts less than 200 represent AIDS stage HIV infection. Agents proceed down the chain of compartments eventually subjected to an HIV-related death. Although the original model of Stover et al. supports the possibility of allowing back-progression up the chain of CD4 compartments, the default setting in the Spectrum software prevents that possibility. Thus, for the purpose of our simulation, we assume only forward compartment progression.

The CD4 model, illustrated in Figure 4.1, contains two sets of compartments—one for agents without treatment and one for agents on ART. For each CD4 compartment in the non-ART track, the model specifies both a transition rate to the subsequent CD4 compartment and a mortality rate. For the annual mortality rates that exceed 1.0, all members of the group are expected to expire before the year has ended. In Figure 4.1 $p$ signifies the probability of fast-progression, $\mu$ is a non-ART compartment’s mortality probability, $\alpha$ is an ART compartment’s mortality probability, and $\lambda$ is a compartment’s progression probability. The probabilities are numbered according to the natural indexed order of compartment progression with 1 signifying CD4 $>$500 and 7 signifying CD4 $<$50.

If the agent initiates ART treatment, they are transitioned to the ART side of the model. An agent on ART stays in the ART compartment giving their CD4 at initiation. There is no transitioning between these ART compartments except by death. For each compartment in the ART tracks, only a mortality rate is specified. As such, the agent is subject to the mortality rate of the CD4 count at which they had initiated ART. The original model by Stover et al, uses user-provided information on ART services to calculate transition parameters from the non-ART side to the ART side [159]. However, these ART adoption rates are not appropriate for our simulation as they treat populations in aggregate and require the user to supply detailed country-specific data on ART services.

**HIV Progression Speed**

Post PHI, at the beginning of the asymptomatic phase, the model determines whether an HIV infection will be a slow-progression or fast-progression. An age-specific ratio, provided in Table 4.6, divides new infections into slow-progressors beginning at “CD4 $>$500” and those fast-progressors starting in the “CD4 350-500” compartment. Thus, fast-progressors skip a CD4 group, shortening the path toward HIV-related death. This bifurcation of new infections is used to model HIV infections which do not fully recover near-normal CD4 levels after the initial PHI stage. Additionally, these fast-progressors reduce the combined average length of time from infection to death bringing
Figure 4.1: The CD4 Compartment Model
it more in line with established mortality rates.

Table 4.6: Distribution of New Infections By CD4 count

<table>
<thead>
<tr>
<th>CD4 Count</th>
<th>15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45+</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;500</td>
<td>64.3</td>
<td>60.7</td>
<td>58.5</td>
<td>55.2</td>
</tr>
<tr>
<td>350-500</td>
<td>35.7</td>
<td>39.3</td>
<td>41.5</td>
<td>44.8</td>
</tr>
</tbody>
</table>

Disease Progression

Table 4.7 contains the average number of years spent in each CD4 compartment. For duration-based transitions we transition at the expected mean compartment transition times. For probabilistic transitions, these mean durations are adapted into daily probability of transitions. Daily probabilities are calculated using a normal distribution hazard function with mean set to the transition time and the standard deviation set to a fraction multiple of the mean. This calculation is explained further in Section 4.2.4.

Table 4.7: Average Number of Years in a CD4 Category

<table>
<thead>
<tr>
<th>CD4 Count</th>
<th>15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45+</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;500</td>
<td>8.53</td>
<td>6.78</td>
<td>5.45</td>
<td>4.7</td>
</tr>
<tr>
<td>350-500</td>
<td>4.49</td>
<td>4.17</td>
<td>2.82</td>
<td>1.87</td>
</tr>
<tr>
<td>250-349</td>
<td>3.4</td>
<td>2.21</td>
<td>1.72</td>
<td>1.17</td>
</tr>
<tr>
<td>200-249</td>
<td>1.97</td>
<td>0.92</td>
<td>0.8</td>
<td>0.55</td>
</tr>
<tr>
<td>100-199</td>
<td>4.67</td>
<td>1.57</td>
<td>1.48</td>
<td>1.05</td>
</tr>
<tr>
<td>50-99</td>
<td>2.87</td>
<td>0.69</td>
<td>0.69</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Mortality

Each compartment in the CD4 model has a set of parameters that define that compartment’s annual probability of mortality based upon age ranges. We initialize our model using the Spectrum software’s default mortality rates for the Asia region. Table 4.8 provides the annual mortality rates without ART whereas Table 4.9 provides the annual mortality rates for persons on ART. Note that these are rates and not probabilities. As such, values greater than 1 indicate that the entire population in that compartment is expected to die in less than a year. We discuss the adaptation of these values into daily probabilities in Section 4.2.4.
Table 4.8: Mortality Without ART

<table>
<thead>
<tr>
<th>Age Group</th>
<th>CD4 Count</th>
<th>15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45+</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;500</td>
<td>0.005</td>
<td>0.004</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>350-500</td>
<td>0.011</td>
<td>0.010</td>
<td>0.013</td>
<td>0.013</td>
<td></td>
</tr>
<tr>
<td>250-349</td>
<td>0.026</td>
<td>0.026</td>
<td>0.036</td>
<td>0.032</td>
<td></td>
</tr>
<tr>
<td>200-249</td>
<td>0.061</td>
<td>0.069</td>
<td>0.096</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>100-199</td>
<td>0.139</td>
<td>0.185</td>
<td>0.258</td>
<td>0.203</td>
<td></td>
</tr>
<tr>
<td>50-99</td>
<td>0.321</td>
<td>0.499</td>
<td>0.691</td>
<td>0.513</td>
<td></td>
</tr>
<tr>
<td>&lt;50</td>
<td>0.737</td>
<td>1.342</td>
<td>1.851</td>
<td>1.295</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.9: Mortality With ART

<table>
<thead>
<tr>
<th>Age Group</th>
<th>CD4 Count</th>
<th>15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45+</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-6 months on treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;500</td>
<td>0.005000</td>
<td>0.003531</td>
<td>0.005000</td>
<td>0.005000</td>
<td></td>
</tr>
<tr>
<td>350-500</td>
<td>0.011491</td>
<td>0.009503</td>
<td>0.01398</td>
<td>0.012624</td>
<td></td>
</tr>
<tr>
<td>250-349</td>
<td>0.015750</td>
<td>0.016738</td>
<td>0.018557</td>
<td>0.031872</td>
<td></td>
</tr>
<tr>
<td>200-249</td>
<td>0.016560</td>
<td>0.017608</td>
<td>0.019537</td>
<td>0.056437</td>
<td></td>
</tr>
<tr>
<td>100-199</td>
<td>0.020620</td>
<td>0.021938</td>
<td>0.024427</td>
<td>0.070647</td>
<td></td>
</tr>
<tr>
<td>50-99</td>
<td>0.025800</td>
<td>0.027468</td>
<td>0.030667</td>
<td>0.088787</td>
<td></td>
</tr>
<tr>
<td>&lt;50</td>
<td>0.037290</td>
<td>0.039738</td>
<td>0.044507</td>
<td>0.129027</td>
<td></td>
</tr>
<tr>
<td>7-12 months on treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;500</td>
<td>0.005000</td>
<td>0.003531</td>
<td>0.005000</td>
<td>0.005000</td>
<td></td>
</tr>
<tr>
<td>350-500</td>
<td>0.010770</td>
<td>0.009503</td>
<td>0.012517</td>
<td>0.012624</td>
<td></td>
</tr>
<tr>
<td>250-349</td>
<td>0.011110</td>
<td>0.014538</td>
<td>0.012917</td>
<td>0.031872</td>
<td></td>
</tr>
<tr>
<td>200-249</td>
<td>0.011450</td>
<td>0.014988</td>
<td>0.013327</td>
<td>0.038827</td>
<td></td>
</tr>
<tr>
<td>100-199</td>
<td>0.012940</td>
<td>0.016938</td>
<td>0.015117</td>
<td>0.044087</td>
<td></td>
</tr>
<tr>
<td>50-99</td>
<td>0.014570</td>
<td>0.019078</td>
<td>0.017077</td>
<td>0.049847</td>
<td></td>
</tr>
<tr>
<td>&lt;50</td>
<td>0.017580</td>
<td>0.023018</td>
<td>0.020687</td>
<td>0.060467</td>
<td></td>
</tr>
<tr>
<td>Greater than 12 months on</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;500</td>
<td>0.003980</td>
<td>0.003531</td>
<td>0.004367</td>
<td>0.005000</td>
<td></td>
</tr>
<tr>
<td>350-500</td>
<td>0.004040</td>
<td>0.005288</td>
<td>0.004447</td>
<td>0.012624</td>
<td></td>
</tr>
<tr>
<td>250-349</td>
<td>0.004170</td>
<td>0.005468</td>
<td>0.004607</td>
<td>0.013187</td>
<td></td>
</tr>
<tr>
<td>200-249</td>
<td>0.004310</td>
<td>0.005638</td>
<td>0.004767</td>
<td>0.013647</td>
<td></td>
</tr>
<tr>
<td>100-199</td>
<td>0.004890</td>
<td>0.006408</td>
<td>0.005467</td>
<td>0.015717</td>
<td></td>
</tr>
<tr>
<td>50-99</td>
<td>0.005350</td>
<td>0.007238</td>
<td>0.006237</td>
<td>0.017977</td>
<td></td>
</tr>
<tr>
<td>&lt;50</td>
<td>0.006710</td>
<td>0.008788</td>
<td>0.007647</td>
<td>0.022137</td>
<td></td>
</tr>
</tbody>
</table>
4.2.4 Adaptation of CD4 Model

The initializing constants for the CD4 model compartment transitions and mortality transitions are given as average durations, and annual probabilities respectively. However, our model uses daily timesteps. Thus, we must adapt the CD4 model to accommodate daily timesteps which requires making certain assumptions about the distributions of the transition probabilities.

Transition Adaptation

The inter-compartment transitions are defined as the average number of years residing in the compartment as provided in Table 4.7. For our simulation we allow for either duration-based or probability-based transitions. Duration-based transitions are useful when we want to eliminate variance in transition time while studying other aspects of the model. However, simulating large numbers of agents all starting their infection at the same time can produce undesirable results stemming from all agents transitioning in lockstep through their CD4 compartments. Instead, probabilistic transitions allow variance in transition durations, but require we make assumptions on the distribution of the probabilities.

For duration-based transitions, we convert the average number of years in the compartment to the average number of days in the compartment by multiplication by 365. During simulation, when the in-compartment duration exceeds this average value of days, the agent’s CD4 model transitions to the next compartment. As such, under duration-based transitions, all infections will have the exact same progression times through the CD4 compartments.

For probability-based transitions, we must convert the average number of years in the compartment to a daily probability of transition out of the compartment. For each timestep, if a random uniform sample is less than the transition probability, then the transition occurs and the CD4 group of the agent is updated. Note that the <50 CD4 group has no transition probability as it is the last group possible and is only subject to mortality. Additionally, as noted above, ART CD4 groups do not experience CD4 group transitions. This is a limitation of the original model we are basing our implementation upon. To obtain the daily probability of transition, we must make an assumption as to the distribution of the transition probabilities. We examined both using a uniform daily probability and a probability based on a normal hazard function. Using a constant uniform daily probability is analogous to treating HIV progression as a Markov chain with time-homogeneous transition probabilities whereas the normal hazard approach is analogous to a Markov chain with time-inhomogeneous transition probabilities [167]. As such, both approaches satisfy the Markov property with the compartment transition probabilities only depending on the present compartment state without regard to the sequence of events that precede it.

For a uniform daily probability, we used Fleurence and Hollenbeak’s equation to adapt annual rates to daily probabilities [61]. In Equation 4.11, the number of years in the compartment $n_{\text{years}}$ is converted to a annual rate by taking its reciprocal. Next, the annual rate is converted to a daily rate.
by division by 365. Subsequently, the daily rate is converted to a daily probability approximation of the rate. Table 4.10 summarizes the resulting transition probabilities from processing Table 4.7 through Equation 4.11.

\[ p_{\text{daily}} = 1 - \exp^{-\left(\frac{1.0/n_{\text{years}}}{365}\right)} \]  

\[ (4.11) \]

Table 4.10: Uniform Daily Probabilities

<table>
<thead>
<tr>
<th>CD4 Count</th>
<th>15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45+</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;500</td>
<td>3.2114E-4</td>
<td>4.0401E-4</td>
<td>5.0258E-4</td>
<td>5.8275E-4</td>
</tr>
<tr>
<td>250-349</td>
<td>8.0548E-4</td>
<td>1.2389E-3</td>
<td>1.5916E-3</td>
<td>2.3389E-3</td>
</tr>
<tr>
<td>200-249</td>
<td>1.3898E-3</td>
<td>2.9735E-3</td>
<td>3.4188E-3</td>
<td>4.9689E-3</td>
</tr>
<tr>
<td>100-199</td>
<td>5.8649E-4</td>
<td>1.7435E-3</td>
<td>1.8495E-3</td>
<td>2.6059E-3</td>
</tr>
<tr>
<td>50-99</td>
<td>9.5415E-4</td>
<td>3.9627E-3</td>
<td>3.9627E-3</td>
<td>5.4645E-3</td>
</tr>
</tbody>
</table>

Unfortunately, using a uniform daily probability does not produce a desirable distribution of transitions in our opinion. Figure 4.2a demonstrates the distribution of transition times for 10,000 individuals when using the uniform daily probability of transition based upon a 1 year mean duration. This graph only represents a single hypothetical transition, not the full CD4 model. Note that a significant number of people transition in the first few days, which is not realistic. If we applied uniform transition probabilities to the entire CD4 model, then we would have an unrealistic number of individuals who rocket through the CD4 model compartments into the final highest mortality compartment and onto a quick death. This would overestimate the mortality and underestimate the likelihood of the individual spreading HIV.

If what we expect is a natural bell curve distribution pattern of transition durations, centered on the mean, then we cannot use a fixed uniform daily transition probability. As such, we implemented a normal distribution hazard function to use for determining the daily probability of transition. Hazard functions are often used in survival analysis to, for example, estimate the transition probability for quality of life states and mortality [167].

Equation 4.12 defines the normal distribution hazard function. The hazard function \( H(x) \) represents the probability that the person will transition at time \( x + \epsilon \) given that the person has survived up until time \( x \). The normal hazard function is the ratio of the normal probability density function (PDF) to complement of the normal cumulative distribution function (CDF). Due to limits of numerical precision on computers, the normal hazard function becomes more difficult to compute as the hazard ratio approaches 0/0, which is undefined. For completeness, when the normal hazard function exceeds \( \mu + 5\sigma \), we approximate the function by linear extrapolation from the points \( \mu + 5\sigma \) to \( \mu + 5\sigma + 1 \). However, although this is an approximation, our tests have shown that it is unlikely that many people have not transitioned by \( \mu + 5\sigma \).
\[ H_N(x, \mu, \sigma) = \frac{f(x, \mu, \sigma)}{1 - F(x, \mu, \sigma)} \]
\[ = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]
\[ = 1 - \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{x-\mu}{\sigma \sqrt{2}} \right) \right] \] (4.12)

In Figure 4.2b we demonstrate the distribution of transition times when using the modified normal hazard function assuming a mean of \( \mu = 1 \) year and standard deviation \( \sigma = \mu / 2.0 \). As the figure shows, the hazard function produces a normal bell of transitions centered about the mean.

Figure 4.3 compares the percent of the population that has transitioned over time between the uniform daily and normal hazard approaches. The lines stop when the simulated population of 10,000 people have all transitioned. Using the uniform daily probability of transition, it takes almost 11 years for 100% of the population to transition with approximately 63% transitioning before the mean. Using the normal hazard approach, it took just over 3 years for 100% of the population to transition with 50% of the population transitioning by the mean. As such, we believe that the normal hazard model provides a more suitable approach to modeling transition probabilities.

Our normal hazard method requires us to define a standard deviation for the normal distribution we base the hazard upon. For this, we chose to utilize a standard deviation that is proportional to the mean. We use a fixed constant to multiply times the mean to produce the standard deviation. This process produces a standard deviation that is of the same relative scale as the mean. This mean multiplier is used for all transition normal hazard functions.

Through a combination of iterative experimentation and a simple hill-climbing algorithm, we determined that a mean multiplier of 0.6532 resulted in the closest match between our CD4 compartment model and Spectrum. Table 4.11 summarizes the resulting standard deviations we use based upon multiplying the mean durations by 0.6532. As is evident, this produces standard deviations which are proportional to the mean, providing scales of durations which are of a proportional magnitude of time.

**Table 4.11: Standard Deviation in Years for CD4 Categories**

<table>
<thead>
<tr>
<th>CD4 Count</th>
<th>15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45+</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;500</td>
<td>5.57</td>
<td>4.43</td>
<td>3.56</td>
<td>3.07</td>
</tr>
<tr>
<td>350-500</td>
<td>2.93</td>
<td>2.72</td>
<td>1.84</td>
<td>1.22</td>
</tr>
<tr>
<td>250-349</td>
<td>2.22</td>
<td>1.44</td>
<td>1.12</td>
<td>0.76</td>
</tr>
<tr>
<td>200-249</td>
<td>1.29</td>
<td>0.60</td>
<td>0.52</td>
<td>0.36</td>
</tr>
<tr>
<td>100-199</td>
<td>3.05</td>
<td>1.03</td>
<td>0.97</td>
<td>0.69</td>
</tr>
<tr>
<td>50-99</td>
<td>1.87</td>
<td>0.45</td>
<td>0.45</td>
<td>0.33</td>
</tr>
</tbody>
</table>
Figure 4.2: Comparison of Uniform and Normal-Hazard Transition Patterns
Figure 4.3: Comparison of Uniform and Normal-Hazard Transition Rates
Mortality Adaptation

For each CD4 compartment, age-based annual mortality rates are specified for both HIV infections without ART and HIV infections with ART. The mortality rates are specified for four different age ranges. Table 4.8 defines the mortality rate for HIV infections without ART. Table 4.9 defines the mortality rate for HIV infections with ART. These rates must be adapted for use in our multi-agent system which uses day timesteps.

Unfortunately, the normal hazard approach we used for CD4 group transitions is not appropriate for modeling daily mortality. As is evident from Figure 4.3, the normal hazard under-represents transitions below the mean. The normal hazard approach works well when you expect a normal distribution of events about the mean. Mortality is specified in terms of an annual mortality rate. Assuming the annual mortality rates are constant, which is a dubious assumption, then the mean survival time would be constant and is given by calculating its reciprocal [149].

Although calculating the inverse of the mortality rates in Table 4.8 is trivial, we reproduce the resulting table of mean survival years for a person not on ART in Table 4.12 to illustrate a point. Obviously, no person is going to survive 250 years. As such, for the upper CD4 compartments, mortality from HIV is very low. However, as the disease progresses, the agent proceeds into compartments with higher mortality rates. In the most severe case, a person with a CD4 count <50 and a 45+ age group, the mortality rate 1.295 translates to a mean survival time of 0.772 years. Thus, for mortality rates of greater than one, the agent is expected to die in less than a year. Thus, for most of these mortality rates we do not expect to have events occurring on both sides of the theoretical means in Table 4.12. To complicate the matter, we assume the mortality rates are based upon the duration spent in the CD4 compartment. When the agent’s infection transitions to a new CD4 compartment, the in-compartment duration resets to zero. Hence, for most of these mortality rates, the normal hazard approach would greatly underestimate mortality.

Table 4.12: Hypothetical Mean Years Survival

<table>
<thead>
<tr>
<th>CD4 Count</th>
<th>15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45+</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;500</td>
<td>200</td>
<td>250</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>350-500</td>
<td>90.909</td>
<td>100</td>
<td>76.923</td>
<td>76.923</td>
</tr>
<tr>
<td>250-349</td>
<td>38.462</td>
<td>38.462</td>
<td>27.778</td>
<td>31.250</td>
</tr>
<tr>
<td>100-199</td>
<td>7.194</td>
<td>5.405</td>
<td>3.876</td>
<td>4.926</td>
</tr>
<tr>
<td>50-99</td>
<td>3.115</td>
<td>2.004</td>
<td>1.447</td>
<td>1.949</td>
</tr>
<tr>
<td>&lt;50</td>
<td>1.357</td>
<td>0.745</td>
<td>0.540</td>
<td>0.772</td>
</tr>
</tbody>
</table>

Therefore, instead of the normal hazard function that we used for compartment transitions, we use an constant daily probability of mortality as defined by Equation 4.11 using the reciprocal
annual rate for an approximation of the mean mortality duration. Note that this approach is
equivalent to using an exponential hazard function, where the \( \beta \) in Equation 4.13 is the reciprocal
of the \( p_{\text{daily}} \) from Equation 4.11.

\[
H_E(x, \beta) = 1/\beta
\]  

(4.13)

4.3 Agent Model

In the simulation, an agent represents a single person. Each agent belongs to a population which
defines key demographic characteristics as well as the range of behavioral responses. Over time, each
agent develops its own unique history and internal state based upon his relationships, advancing
age, sexual acts, and exposure to infection. As such, the agent adapts its behavior based upon its
internal state within the confines of the simulation parameters.

At the start of a simulation, the initialization parameters, outlined in Table 4.2, are used to
set the agent’s initial internal state, defined in Table 4.13. At initialization and during simulation,
HIVInstance objects, defined in Table 4.14, are created and assigned to the agents which are infected
with HIV. During simulation, the agents internal state is updated through three processes—internal
updates, relationship maintenance, and the dating module. Internal updates include age advancing,
HIV testing, HIV viral progression, and PrEP dosing. Relationship maintenance involves having
sex with existing relationships and determining when the relationship will end. The dating module
is used to form new relationships when an agent determines that it wants to date. An overview
diagram of the algorithm is provided in Figure 4.4.

To run a simulation, the user must define at least one population group. The population must
have a valid age range and a useful population size. We consider a useful population size to be on
the range of 1000 to 100,000. While the simulation is not limited to these values, the heap space
limit of JAVA will impose its own limitations. We routinely run our simulation tests with a heap
size range of 4048 MB to 8096 MB.

Additionally, the user must specify several key parameters from the set of available population
parameters defined in Table 4.2. The user must specify the mean duration of steady relationships
\( D_S \), the number of initially HIV positive individuals \( N_H \), and the percentages of insertive \( R_I \),
receptive \( R_R \), and versatile sexual roles \( R_V \). The user may choose to also specify a number of
annual casual sexual encounters \( N_C \). Our model prohibits steady relationships from re-forming
with the same partners, but re-forming is allowed for casual relationships.

To study the effects of concurrent relationships, the user must also specify the number of
preferred concurrent relationships \( N_S \), and set a non-zero value for the percentage of the population
practicing concurrency \( R_C \). If more detailed information is available, the user may specify normal
or uniform distributions for many of the parameters allowing for a wider range of initial states for
the agents.
Figure 4.4: Algorithm Diagram
4.3.1 Partner Finding

Partner finding, illustrated in Figure 4.5, is the process of pairing up agents who want to form a new relationship. At each timestep $t$, a pool of all agents wishing to date, $D_t$, is formed. We iterate through all of the agents and ask each one if it is ready to date. When an agent initially determines that it wants to date, it triggers an internal dating mode change for that agent. That is, while the agent is in free-mode, then each timestep the agent will calculate whether it wants to date and if so, whether it wants to search for a steady relationship or a casual relationship. Once the agent determines that it wants to date, it transitions to dating-mode and will remain wanting to date until it has formed a new relationship. For example, if an agent determines that it wants to form a casual relationship, then it will search for a casual relationship until it forms one. Once a relationship is formed, the agent’s dating mode resets to a non-searching mode. If the agent desires both a steady relationship and a casual relationship, preference is given in the search for a steady partnership due to the presumption that steady relationships will be longer. The determination of the dating mode is based upon the agent’s dating history, the casual relationship schema, and the steady relationship schema. The structure for a relationship schema is provided in Table 4.16. In our model, a person is limited may have more than one steady relationship at a time. Casual relationships are minimal length relationships that exist until the first instance of sex occurs. We allow for two different calculation modes of agents deciding whether they want to date—duration-based and probabilistic.

In duration-based dating mode, an agent wants to date for a relationship type when the time since the last formation of that relationship type exceeds the mean number of days between formations of that relationship type. For a given agent $A$ and relationship schema $S$, we determine the mean days between relationship formations by Equation 4.14. If the agent has a steady relationship, even if the time between steady relationship formations has been exceeded, a new relationship will not form until the old one has ended. Additionally, the agent will not enter into a search for a new
steady relationship until the old steady relationship has ended.

\[
\mu_D(A, S) = \begin{cases} 
365/F & : A \text{ has no steady relationship.} \\
365/F_{\text{Steady}} & : A \text{ has a steady relationship.}
\end{cases} \quad (4.14)
\]

In probabilistic dating mode, an agent \( A_i \) wishes to date if his current number of partners is less than his preferred number of partners and a uniform random sample satisfies the daily probability of dating. There are two daily probabilities of dating—one for steady relationships, and one for casual relationships. Since steady relationships are presumed to be of longer durations, we give preference to forming steady relationships.

As it would be both intractable computationally and an impractical dating scenario in reality, an agent does not evaluate the entire dating population \( D_t \) for a new potential relationship. Instead, each dating agent \( A_i \) is provided with a fixed-size random subset, \( D_{t,A_i} \). To prevent unnecessary attempts to form relationships with sexually incompatible partners, all agents selected for \( D_{t,A_i} \) must have complementary preferred sexual roles to agent \( A_i \). That is, if the agent prefers an insertive sexual role, then the new partners must be either receptive role or versatile role. Agents preferring a receptive role can only pair with agents preferring an insertive or versatile role. Versatile preference agents may pair with a versatile, insertive, or receptive preferring agent.

If the user has elected to specify a non-zero previous partnership mixing parameter, \( W_{PP} \), then we add all currently dating past partners of \( A_i \) to his dating pool \( D_{t,A_i} \). Thus, an agent will have the potential to make new relationships or resume previous casual relationships.

We iterate randomly through the set of agents in the dating pool, giving each agent that has not formed a new pair a chance to form a pair with a subset of the dating pool. An agent evaluates all members of their dating pool for compatibility and attempts to form a new relationship. As each new relationship is formed, the available dating pool is updated, removing the newly paired agents from the pool \( D_t \). As such, each agent in the dating pool may form at most one new relationship per timestep.

### 4.3.2 Dating Evaluation

The United States has demonstrated strong assortative pairing by race [106, 129]. In assortative pairing, a person is more likely to choose another person who is similar in demographic characteristics such as age, race, etc. Disassortative pairing occurs when a person is more likely to choose another person who is dissimilar in characteristics. We have implemented a method of partner evaluation that allows for assortative and disassortative pairing.

When an agent is dating, he is presented with a set of potential new partners and must evaluate each partner to find the most compatible partner. While it would be ideal to have full preference data encompassing all possible combinations of evaluation parameters, such data is intractable to collect as each additional variable increases the amount of data collection exponentially. For
example, if there are $N$ different demographic groups, then there are $N^2$ different values to collect. Thus, we have implemented a simplified, symmetric evaluation function. As the evaluation of partners is symmetric, for agents $A_i$ and $A_j$ the value of the evaluation function $Eval(A_i, A_j)$ is equal to $Eval(A_j, A_i)$. The evaluation function is defined in Equation 4.15.

\[ Eval(A_i, A_j) = W_{NP} \ast \frac{1}{\left| NPP(A_i) - NPP(A_j) \right| + 1} \]
\[ + W_{Age} \ast \frac{1}{\left| Age(A_i) - Age(A_j) \right| + 1} \]
\[ + W_{Pop} \ast SamePop(A_i, A_j) \]
\[ + W_{PP} \ast Previous(A_i, A_j) \]
\[ + C_{SS}(A_i) \ast W_{SS} \ast SameHIV(A_i, A_j) \]  

(4.15)

Depending on the mixing weights, defined in Table 4.17, each mixing parameter can become either assortative or disassortative. If the weight is greater than 1, then the mixing parameter is assortative. If the weight is less than 1, then the mixing parameter is disassortative. Under assortative mixing, agents with similar values are more likely to pair. Under disassortative mixing, agents with different values are more likely to pair.

An agent evaluates each potential partner according to Equation 4.15. This equation combines the weighted mixing parameters for promiscuity, age, population group, previous partnerships, and serosorting. The equation uses the additional functions defined in Table 4.18. The promiscuity mixing factor is based upon each agent’s preferred number of partners. The preferred number of partners is an optional parameter that can be defined for populations. The age mixing factor depends on the difference between the two agents’ ages in years. The population group factor is a simple weight depending on whether or not the two agents belong to the same population group. The weight for previous partners, $W_{PP}$, is only added if the pair of agents have previously had a casual relationship. Note that if the past relationship was a steady relationship, we do not add the constant in that case as we do not allow recurring steady partners in our model.

The serosorting factor is only added if the agent $A_i$ is practicing serosorting on this timestep. This is determined by a Bernoulli trial for the agent as described in Equation 4.16. We allow for differing rates of serosorting for HIV positive and HIV negative individuals, as illustrated in Equation 4.17. If the individual $A_i$ knows he is HIV positive, then he will add the serosorting weight to evaluations with the probability $SS_+$. Otherwise, the individual will use the serosorting weight for HIV negative individuals, $SS_-$. 

\[ C_{SS}(A_i) = 1 - Bern(SS(A_i)) \]  

(4.16)

\[ SS(A_i) = \begin{cases} 
SS_+ & \text{if } A_i \text{ is HIV} \\
SS_- & \text{if } A_i \text{ is HIV} 
\end{cases} \]  

(4.17)
4.3.3 Pair Formation & Dissolution

An agent may form either a casual or steady partnership. An agent may have multiple steady relationships at a time, but may have only one casual relationships at a time. Casual relationships are meant to be brief sexual encounters. Only one relationship of either type can be formed per day. The user defines the number of casual partnerships per year, \( N_C \), the mean length of steady relationships, \( D_S \), and the number of concurrent relationships allowed, \( C \). A casual relationship has minimal length, terminating upon the first occurrence of sex.

As we need to know when to form new relationships, we calculate the expected number of days between the formation of new relationships. For steady relationships, we assume the duration between new steady relationships \( B_S \) is equal to the relationship duration \( B_D \) divided by the number of concurrent relationship tracks, \( C \). Thus, as soon as one steady relationship ends, one immediately starts.

For duration-based dating, we maintain a variable to track the number of days since the last formation of each relationship type. Relationships are formed when the duration since the last partnership formation exceeds \( B_C \) or \( B_S \). When a relationship is due to be formed, the dating function that determines the agent’s desire to date will indicate the appropriate relationship type sought. Relationships are terminated when the duration of the relationship exceeds the expected relationship duration \( D_C \) or \( D_S \).

For probabilistic-based dating, we use a normal hazard functions to determine the duration between forming relationships and the duration of the actual formed relationship. For these distributions, we assume a standard deviation equal to half the mean. For the duration between forming relationships, we used a mean of either \( \mu_s = B_S \) or \( \mu_c = B_C \) depending on whether the relationship is steady or casual. This distribution is used within the dating function that determines whether the agent is wishing to date and, if so, which relationship type is sought.

For determining when to end an existing relationship, we used a normal hazard function with a mean of either \( \mu_s = D_S \) or \( \mu_c = D_C \) depending on whether the relationship is steady or casual. Since a relationship includes a relationship schema from each partner, we alternate use of each partner’s relationship schema to determine whether the relationship should end. This allows both schema expected durations to influence the relationship duration without the double jeopardy effect that would result from testing for relationship termination for both schemas at each timestep.

4.3.4 Sexual Roles

For the population, the percentage of sexual role preferences are defined for insertive, receptive, and versatile preferred roles. We assume that only agents with compatible sexual roles may form a relationship. Thus, a versatile agent may pair with insertive, receptive, or versatile agents. Receptive agents may only pair with insertive or versatile agents. Insertive agents may only pair with receptive or versatile agents. For each sexual act, versatile roles are converted into the appropriate
matching roles as summarized in Table 4.19.

While our model does support a degree of serosorting in relationship formation and condom use, we do not support seropositioning. In seropositioning, an agent changes his sexual role based upon the perceived HIV status of his partner [133]. For our model, insertive and receptive roles never change. Versatile roles will change into insertive or receptive roles during sex in response to being paired with versatile agents, but not in response to HIV status.

4.3.5 Sexual Frequency

We consider sexual acts as occurring within a sexually active relationship between two agents, whether it be a single day casual relationship, or a multiple year steady relationship. An agent may have sex with each of his relationship partners a maximum of once per day. We track the last timestep that a pair of agents has attempted sex to prevent agents from multiple attempts per day for the same relationship. Regardless of whether we are simulating under duration-based dating or probabilistic dating, we do not prohibit an agent from having sex with multiple relationships per day. However, the occurrence of multiple sex per day is expected to be rare in our simulation.

Regardless of sexual frequency, at the start of a relationship the couple will attempt to have sex each day until they succeed or the relationship ends. While this might not seem realistic, one could interpret the start of a relationship as the start of a sexual portion of the relationship. As such, the non-sexual part of the relationship likely started prior to the initiation of sex. Requiring sex at the start of a sexual relationship is necessary to accommodate casual relationships which may be used to model single day sexual dalliances or episodes of anonymous sex. Depending on the user settings, it is possible that a pair of agents will never have sex within their relationship. This would occur, for example, if an agent designated as one who always wears condoms had paired with an agent who never wears condoms. In that case, the two agents would never reconcile condom negotiations in a manner that would result in sex.

Apart from the initial sex of a relationship, future sex is governed by the sexual frequency preferences of the agents in the relationship. Steady and casual relationships each have their own sexual frequency rates $S_{\text{Steady}}$ and $S_{\text{Casual}}$ respectively. These rates are defined in terms of the number of sexual acts for the relationship type per year. Thus, the number of sexual acts are divided equally among the expected number of relationships per year. We determine when the next sexual act will occur based upon an expected length of time between sex, defined for steady relationships in Equation 4.18 and for casual relationships in Equation 4.19. To clarify, $S_{i,\text{Steady}}$ represents the sexual frequency of agent $A_i$ for a steady relationship, $D_{i,\text{Steady}}$ is the mean duration in days for steady relationships of agent $A_i$, and $N_{i,\text{Steady}}$ is the expected number of steady relationships formed by agent $A_i$ annually. These equations support both the case where sexual frequency is defined globally and the case where sexual frequency is defined at the population level.
\[ L_{\text{Steady}}(A_i) = \frac{D_{i,\text{Steady}} S_{i,\text{Steady}}}{N_{i,\text{Steady}}} \]

\[ L_{\text{Casual}}(A_i) = \frac{D_{i,\text{Casual}} S_{i,\text{Casual}}}{N_{i,\text{Casual}}} \]

If the user specifies sexual frequency as a global parameter, then for a given relationship type, steady or casual, both partners will always have the same annual sex rate and thus both want sex at the same timesteps. However, if the user specifies sexual frequency as a population specific parameter, it is possible that two agents will have different annual sex rates for the same relationship type. This would occur if the agents belong to two different populations with differing sexual frequency rates. To accommodate the possibility of different sexual frequency desires between the partners, we assume that the agent with the lower sexual frequency rate will govern the mean duration between sex. Thus, for agents \( A_i \) and \( A_j \) in a casual relationship, if \( L_{i,\text{Casual}} \geq L_{j,\text{Casual}} \), then the length of time between sex will be at least \( L_{i,\text{Casual}} \).

For duration-based dating, each relationship maintains an integer counter defining the number of days until the next sexual contact. Upon successful sexual contact, the counter is initialized to either \( L_{\text{Steady}} \) or \( L_{\text{Casual}} \) depending on whether the relationship type is steady or casual. At each timestep the counter is decremented by one. When the counter reaches zero, the partners will engage in sex if the relationship is still active. If an attempt at sex is unsuccessful, such as due to a lack of agreement on condom use, we do not reset the sex delay counter. As such, the couple will try to have sex again on the next day.

For probabilistic-based dating, we use a normal hazard function with mean initialized to either \( \mu_s = L_{\text{Steady}} \) or \( \mu_c = L_{\text{Casual}} \) depending on whether the relationship is steady or casual. For this distribution, we assume a standard deviation equal to half the mean. We chose this standard deviation as it provided similar results to the duration-based model while still providing some variation about the mean. The normal hazard function ensures a bell-shaped distribution of times between sex centered at the mean duration. We evaluate the normal hazard function with the duration in days since the last sex event in this relationship. To calculate the duration since last sex, we store the day that the last sex occurred and calculate the difference from the current day timestep from the last sex day. When an attempt at sex is successful, we store the current timestep day as the new day of last sex. If an attempt at sex is unsuccessful, we do not change the day of last sex. Note that, unlike the duration-based dating, probabilistic dating does not guarantee the couple will try for sex again the next day. However, as the duration since last sex increases, the hazard function will ensure that the couple will try to have sex again with increasing frequency until they successfully negotiate sex.
4.3.6 Condom Use

The decision to use condoms is affected by the type of relationship, the perceived serostatus of each partner, the sexual role of each partner, as well as personal preferences regarding the necessity of condom use. For sex to occur within a relationship at timestep \( t \), the condom preference at timestep \( t \) for both partners must match. Based upon the user preferences for the simulation, there may be a set of agents who always use a condom and a set of agents who never use a condom. Membership in either of these two sets dictating condom use will override all other considerations regarding the decision to use a condom. For the cases where condom use is not already fixed, we determine condom use based upon known HIV discordance, the presence of any negotiated safety arrangements, and the probability of condom use in the relationship schemas.

Davidovich et al. proposed that relationships where UAI is either perceived as a symbol of trust or is strongly desired would benefit from negotiated safety arrangements [55]. Such a safety arrangement may include establishing valid seroconcordance, discussing condom use, and minimizing risk to the steady relationship from casual relationships. In a similar vein of thought, for our model, the user may designate a percent of agents as those who practice negotiated safety arrangements. This feature is intended to model the decreased use of condoms within steady relationships as well as the efforts of the partners to reduce the risk of compromising the health of their steady partner during casual dalliances. If the agent \( A_i \) is set to use negotiated safety, then he will only use condoms with casual partners. Within a steady relationship, an agent practicing negotiated safety will cease the use of condoms with the steady partner after the first instance of unprotected sex within the steady relationship. Note that it is possible that not both partners in a relationship will be practicing negotiated safety. Additionally, the relationship contains a relationship schema for each agent that defines the relationship type’s condom use percentage, \( C_R \).

Excluding the special cases where an agent always uses a condom or never uses a condom, Equation 4.20 calculates the boolean preference of agent \( A_i \) to use a condom with agent \( A_j \) inclusive of the effects of steady relationships, negotiated safety arrangements, and the base probability of condom use. Equation 4.21 defines the base probability of condom use with an optional risk reduction adjustment further defined in Equation 4.22. When active, the risk reduction adjustment increases the probability of condom use by \( K_D \) percent. This reduction starts when the agent discovers his infection and stops when the duration known, \( durKnownPoz(A_i) \), exceeds the risk reduction period \( K_L \). The risk reduction adjustment is used to model a period of increased condom usage subsequent to a person discovering he has contracted HIV. If the agent does not know he has an HIV infection, then he doesn’t reduce his risk of transmitting HIV through increased condom usage.

\[
desiresCondom(A_i, A_j) = (NS(A_i) \land hasSteady(A_i) \land \neg areSteady(A_i, A_j)) \\
\lor (probCondom \land \neg (NS(A_i) \land areSteady(A_i, A_j) \land hadUAI(A_i, A_j)))
\]  

(4.20)
\[ \text{probCondom} = \text{Bern}(C_R \times \text{riskReduc}(A_i)) \]  

\[ \text{riskReduc}(A_i) = \begin{cases} 
1 & : -\text{knownHIV}(A_i) \\
1 + K_D/100 & : \text{knownHIV}(A_i) \land \text{durKnownPoz}(A_i) < K_L 
\end{cases} \]

4.3.7 HIV Testing & Discovery

HIV testing and disclosure are important components in forestalling epidemics. Accurate knowledge about a partner’s HIV status is necessary to make educated choices about the level of risk to assume with a partner. We model HIV discovery in two ways–symptomatic discovery and regular testing.

There are two times when an agent can become symptomatic, during PHI and during AIDS. An agent has a 70% chance of being symptomatic during the PHI phase of an HIV infection [136]. We determine whether the agent was symptomatic by a simple uniform sample against the probability of being symptomatic. At the end of the PHI phase, if the agent was symptomatic, we assume a correct diagnosis of HIV will be made 17 percent of the time [180]. Failing this early method of HIV discovery, the user may only discover his infection by regular HIV testing or progression to AIDS.

After PHI, an infected agent enters a long asymptomatic period during which HIV may be detected by HIV testing. If duration-based transitions are enabled, then the agent will test for HIV exactly on multiples of the mean number of days between testing. The mean days between HIV tests, \( \mu_T \), is calculated by dividing 365 by the expected number of tests per year, \( T_{HIV} \). Alternatively, for probabilistic testing we use a normal hazard function to establish a normally distributed distribution of duration till the next HIV test. For this distribution we assume a mean of \( \mu_T \) and a standard deviation of \( \sigma_T = \mu_T \).

However, it is important that individuals on PrEP discontinue PrEP use if they become infected with HIV. Therefore, PrEP is typically administered as part of monitored healthcare program that includes regular HIV testing. All agents who are designated as receiving PrEP will take an HIV test on the first timestep of the simulation to ensure that no HIV infected will receive PrEP. Additionally, in our simulation all agents taking PrEP will take an HIV test quarterly each year, which equates to an HIV test every 91 days.

Eventually the agent becomes symptomatic again during the final progression toward AIDS since the body becomes susceptible to opportunistic infections. We assume that all infected agents will know they have HIV at CD4 counts less than 200.

4.3.8 ART

ART can substantially prolong an HIV infected person’s life. However, a person must know they have HIV before they start ART. For agents who know they have HIV, we assume a fixed daily
probability of initiating ART. We use the inverse of the annual ART rate $P_{\text{ART}}$ to approximate the mean years in Equation 4.11. This produces a uniform daily probability of initiating ART for a person who is aware they are infected with HIV. Note that, as discussed previously, this is equivalent to utilizing an exponential hazard function. Once an agent has started ART, we assume perfect medication and program adherence. That is, we assume the agent will take the required medication to attain the ideal HIV transmission reduction and that the agent will not leave the ART program except in the case of death.

4.3.9 PrEP

The agents who take PrEP are assigned at the beginning of the simulation to individuals who are HIV negative. If the user specifies PrEP coverage on the total population, then the percent PrEP coverage ($P_C$) is applied uniformly across all agents. If the user specifies PrEP coverage as a population-specific parameter, then the user must specify a percent PrEP coverage $P_C$ for each population. In that case, the percent PrEP coverage is applied only to the agents in that population.

Each day, there is a $P_A$ percent probability that the agent will take his daily dose of PrEP. For each agent on PrEP, we perform a random uniform sample each day to determine whether or not the PrEP dose was taken or missed. Whether or not the PrEP dose was taken, an agent on PrEP incurs the daily cost of PrEP, $P_D$. For our simulation, PrEP only reduces the probability of acquiring HIV if the daily dose was taken within 1 day of sex and the agent is not already HIV positive.

4.3.10 Mortality

In our model, Agents may die of HIV-related causes or be aged out at the standard mortality for men of 80 years [138]. Aged-out agents are removed from the simulation. Owing to our expectation that simulations will be performed on shorter timescales, we have not implemented background mortality rates which would removed agents for deaths due to non-HIV causes prior to age 80.

4.4 GUI

We have designed a graphical user interface (GUI) to the simulator that allows the user to define simulation parameter sets, run a parameter set, run batches of parameter sets, display collected metrics on the simulation, and analyze the simulator function. The GUI allows for running a parameter set in both a graph-rich interactive mode as well as running the simulator in a more streamlined mode for testing ranges of parameter values in batch.
4.4.1 Parameter Set Editor

The “Parameter Set Editor”, displayed in Figure 4.6, allows the user to define and adjust parameter sets that define the range of agent behavior in simulations. For defining groups, the user may add a group using the “Add Group” button or the user may select a column with the mouse and click the “Remove Group” button to remove a group. The minimum number of groups necessary is one group.

Percentage parameters are defined on the 0 to 100 range. If the sexual role percentages for insertive, receptive, and versatile sexual role do not sum to 100 they are scaled proportionally to 100.

Using the “Edit Parameters” button and resulting panel, displayed in Figure 4.7, the user may move population parameters between simple global definition and more specific group-specific definition. As such, a parameter can have a simple global value, or each group can have its own value for the parameter. This allows for greater diversity of types of data sets that are possible to define. Additionally, this panel allows the user to adjust whether a parameter is defined in terms of a single fixed value, a uniformly distributed value, or a normally distributed value. Only behavioral parameters are eligible for distribution. For example, the length of casual relationships is distributable but the percent of the population that is circumcised is not distributable. Any parameters moved from the global set to the group-specific set will be automatically initialized with the last value from the global set.

4.4.2 Model Analysis

The Analysis menu in the MASHIV interface provides numerous graphs and tables for analyzing the HIV model and the user’s parameter set. We’ve stored Spectrum mortality and CD4 progression values for ready comparison and graph generation. These allow us to regenerate commonly referenced graphs when testing new configurations of the HIV model.

4.4.3 Interactive Simulator

The interactive mode provides the user with the ability to initiate a simulation run and receive real-time results as the simulation progresses. The user may select from a diverse selection of common metrics to graph that ranges from epidemiological metrics to parameter set verifications. The interactive simulator interface is a useful debugging tool to ensure that the parameter set produces the desired behavior before sending the parameter set to the aggregate simulation runner.

To enable the interactive simulator, the user must first either define a new parameter set, or load a parameter set from a pre-saved XML file format which our program generates. Next, the user activates the interactive simulator interface from the menu selection. Presented with the interactive simulator window, the user may wish to select additional graphs via the “Graphs” button which
Figure 4.6: Set Editor
Figure 4.7: Parameter Editor
displays the interface provided in Figure 4.9. With the desired graphs selected, the user may elect to run the simulation to its completion with the “Run” button, or he may advance the simulation a set number of steps using the “Step Forward” button in combination with the number and time scale selections to the right. While the simulation is occurring, the user may use the “Stop” button to pause the simulation. As the simulation progresses, the graphs the user has selected will update with data collected from the simulator.

At the end of the simulation, the user may elect to save a log of all HIV transmissions, time of infection, and virus stage of the infection source. Such a log may prove useful if input to more advanced graph visualization software such as Gephi [22].

4.4.4 Aggregate Simulator

The aggregate mode provides a means of testing ranges of initial conditions and aggregating the results. This is useful for observing the sensitivity of the simulation to a range of input values. First, the user must either open or define a new parameter set in the parameter set editor. With a parameter set loaded, the user opens the aggregate set editor from the “Run” menu. The aggregate set editor, displayed in Figure 4.10, loads and imports the default parameter values from the open parameter set.

The layout of the Aggregate Set Editor matches the global and population division of parameters specified by the user in the parameter set editor previously. Values that the user does not wish to vary remain the same value as imported from the parent parameter set. The user chooses which parameter values to test by toggling the checkboxes. When a checkbox is toggled, two new input fields are presented along with a drop-down list to select the type of distribution. The user may select parameter variations of min/max, uniform distribution, and normal distribution. Min/max allows for sampling parameter values in a linear fashion at equal divisions. For the “uniform” selection, values are selected between the mean value and plus or minus the mean value. For the “normal” selection, values are sampled from the normal distribution with the specified mean and standard deviation. The labels of the two input fields change as to indicate which form of input values are expected from the variation selected by the user.

An aggregate set defines multiple parameter sets, one for each set of varied parameters, and thus multiple simulations. The number of samples specified in the aggregate set determines the number of samples from each varied parameter. The number of random seeds determines how many different random seeds will be tested for the same parameter values. Thus, the total number of parameter sets involved in an aggregate set is the number of samples times the number of random seeds. We designed the simulator to process these simulations each in a different thread to speed up the processing. The number of threads is fixed in the code.

When the user is ready to run the aggregate set through the simulations it defines, the user may opt to run multiple aggregate sets in batch or run a single aggregate set. If the user wishes to do a
Figure 4.8: Interactive Simulator
Figure 4.9: Data Collection Selector
Figure 4.10: Aggregate Editor
batch processing of multiple aggregate sets, he can press the “Run All” button. At that point the GUI will display a data collector interface similar to Figure 4.9. Once the user has chosen which data he wishes to collect and presses the “Save” button, the program will initialize the aggregate simulator interface, displayed in Figure 4.11, and automatically trigger the simulators to begin working on parallel threads, displayed in Figure 4.12. As the simulations proceed, the status of each work thread is displayed via progress bars on the right with a total progress bar at the top.

When the aggregate simulations complete, the program will automatically save the aggregate results object to an XML file in the same directory as the parent file. Then, the program will proceed to the next aggregate set and generate a new aggregate simulation window. Each aggregate simulation window generated is titled appropriately to avoid confusion.

If the user wishes to perform these aggregate simulations individually, they may select an aggregate set from the list in the aggregate set editor and click “Run”. This brings up the aggregate simulation window in Figure 4.11. Via the “Select Data Collected” button, the user may select the data he wishes to collect during the simulation.

When the aggregate simulations for an aggregate set are complete, the results are presented such as in Figure 4.12. Since we are using the JFreeChart library for graphs, the user may right-click on the graphs and save the graph as a PNG file [130]. Clicking the “Resize Graphs” checkbox will toggle the layout of graphs between the resized mode and the full-size mode. The full-size mode allows for larger images to be saved to file whereas the resized mode is better for gaining an overview of all the graphs.

If the user clicks the “Report” button, a table similar to Figure 4.13 will displayed. This table presents the parameter values in the individual sets as well as key metric values collected. For convenience, we calculate a weight for each parameter set and sort the parameter sets by the weight values. The weight is calculated by determining the mean squared error between mean of the simulation prevalence values and the prevalence values of each parameter set. These weights are useful when we are performing large stochastic variation tests. In such tests, the aggregate set has many different random seeds, but does not vary individual parameters. If the user selects a row from the table and clicks “Run Selected”, the corresponding parameter set to the row will be sent to initiate an interactive mode simulator interface. Thus, the user may use the aggregate simulation results to determine the parameter set that produces the closest results to the mean of the stochastic variation and then send the results back to the interactive simulator for more in-depth analysis.

If the user presses the “Analyze” button on the aggregate simulator then the analysis tool will be displayed such as in Figure 4.14. This tool is used to calculate the Pearson correlation coefficient between parameters varied in the simulations and data collected from the simulations. The Pearson correlation coefficient measures the correlation between two sets of values and produces a value between -1 and 1. Values close to 0 indicate there is little correlation whereas values close
to -1 and 1 indicate strong negative correlation and strong positive correlation respectively. To use the tool, the user selects a parameter from the first list, selects the desired collected value from the second list and presses the “Analyze” button. The program then displays the data points collected and the corresponding attempt at a linear line fit to the points. The results box displays the Pearson correlation coefficient, called “R”, as well as the slope of the line. The slope may be useful for estimating values between the tested data points.

If the user presses the “Export Results” button on the aggregate simulator, the program will prompt the user for a filename and will subsequently export the data collected in CSV format to the specified filename. This is a one-way data export. If the user wishes to resume analysis later, he may use the “Save Results” button to save the aggregate results object to an XML file representation. This is the same file format as used when the “Run All” batch mode saves the aggregate results object.

There are two ways to resume reviewing the results of an aggregate results object file. The user may click the “Load Result Set” from the file menu of the main GUI and select the file. The selected file will be displayed in a new aggregate runner window complete with all the graphs and analytical abilities. Alternatively, if the user has used the “Run All” batch mode, then when the user selects a row in the aggregate set editor, the program will check to see if the expected file name of the results file exists in the directory. If the file exists, the “View Results” button will enable in the aggregate editor window. Clicking the “View Results” window will open the aggregate runner window and display the results.

Thus, our aggregate simulator allows an easy to use interface to define parameter variations, select data to collect, display the results, analyze the results, and resume displaying the results.

4.5 Data Collection

The MASHIV simulator includes many types of data that the user can select for collection during simulations. The types of data that can be collected are simulator state, metric distributions, and annual metrics. Simulator state values give the user insight into the state of the simulator’s function. These include iteration timing and the frequency of certain key function calls such as the number of requests for dating. Metric distributions display histograms such as the distribution of relationship durations. Annual metrics are single data points aggregated annually such as the HIV prevalence, HIV incidence, and average concurrency.

The UNAIDS group defined a measure of concurrency as the proportion of people engaging in concurrent relationships to the population being studied. The numerator of this measure included people who maintained two or more relationships where first sexual contact began more than six months prior, and the most recent sexual contact occurred less than six months prior [80]. This definition of concurrency as relationships with intersecting durations is problematic when applied to closed populations such as the one in the MASHIV simulation. During simulation there is a
Figure 4.11: Aggregate Simulations in Progress
Figure 4.12: Completed Aggregate Simulations Interface
Figure 4.13: Aggregate Report Tool
Figure 4.14: Aggregate Analysis Tool
definite possibility that a casual relationship early in the simulation could recur much later in the population. Thus, in such a closed population, such a metric would trend towards 100% concurrency as the simulation progressed. While Morris and Kretzschmar have defined numerous metrics for measuring properties of concurrency in social networks[127], their metrics use relationship existence as a proxy for sexual activity.

For the concurrency average, we calculate an estimate of the number of concurrency partnerships within a window of the past 365 days. While many concurrency metrics are primarily concerned with overlapping relationships, the key factor that can produce HIV transmission is sexual contact. Thus, we utilize a concurrency measure that is based upon the number of unique sexual partners within a time window.

Each agent maintains a list of the last time he had sex with each past partner. When the concurrency measure is collected from a source agent, we iterate over this list of all past partners and create a count of all the agents that the source agent has had within the past 365 days from the timestep of the calculation. For the concurrency average, we obtain the concurrency measure from every living agent and divide by the number of living agents.

4.6 Comparison with Spectrum

To evaluate our HIV model, we compare the output of our model with that produced by the Spectrum software. While MASHIV is an individual-based simulation, Spectrum treats populations in aggregate, migrating fractions of the population between the compartments in the CD4 model. In order to compare the two models, we must disable some of the complicating factors present in Spectrum that would occlude the mortality and CD4 transitions. To do this, we build a country dataset in Spectrum with only 10,000 males and 1 female. The inclusion of the female is only to prevent calculation errors that result in Spectrum when there are zero females. Additionally, we set the international migration and the ratio of male births to zero to prevent additions to the population. As we are only interested in comparing HIV related mortality, we increase the life expectancy in Spectrum to the maximum of 100 years, and select the “UN General” regional model life table. While this doesn’t entirely eliminate background mortality, it does provide a 35 year window where the number of deaths is nearly zero in the absence of HIV infections. We initialize the epidemic by setting a 100% incidence in the “Direct Incidence Input” tool in Spectrum.

Spectrum and MASHIV both define CD4 compartment transition mean duration and mortality rates divided into four different age groups–15 to 24, 25 to 34, 35 to 44, and 45 to 54. In MASHIV, as a person ages, they progress into the corresponding age group, utilizing that group’s set of mean compartment transition durations and mortality rates. To compare Spectrum and MASHIV, we examine their CD4 compartment progression and mortality progression. For examining mortality progression, we compare the two models’ annual AIDS deaths and cumulative AIDS deaths. We examine these statistics for men in the age range of 15 to 49. We examine three cases, differing
only by the initial age of infection. All persons for both models will start their infection at the same age. We explore HIV progression both without ART and then with ART.

4.6.1 HIV Progression Speed

HIV progression is divided into fast and slow progressors. Both MASHIV and Spectrum use the same percentages to divide their population into the two groups, but the manner in which these percentages are applied differs. Post PHI, agents in MASHIV are divided into the two groups according to the percentages in Table 4.6. Using those percentages, a uniform random sample is used to determine whether the agent becomes a fast or slow progressor. In Spectrum, percentages of the aggregate population of infected people are transitioned into groups designated for fast progressors and slow progressors.

Figure 4.15 demonstrates the effect of the progression rate on the expected mortality duration. The figures compare the duration-based transition mode of MASHIV, the probabilistic transition mode of MASHIV, and the Spectrum values. We created the Spectrum values by generating a population of 10,000 infections in Spectrum at different ages and calculated statistics on the number of years till death. In our opinion, MASHIV shows reasonable parity with the Spectrum values for mortality according to both fast and slow progression. However, we did notice a curious increase in the mean age of mortality for later ages in Spectrum. To explain Spectrum’s unexpected increase in the mean age of mortality, we created Figure 4.16 which illustrate how infections started at the ages of 50 to 65 experience a much wider distribution of mortality durations. We believe this anomalous behavior is indicative of the non-realistic effect that treating populations in aggregate can sometimes have on simulated populations. However, this effect is likely magnified by our unrealistic population simulation with Spectrum which only had a single closed population group comprised of an entirely infected population.

4.6.2 CD4 Compartment Progression

For examining CD4 compartment progression, we look at a population with 10,000 infections, a starting age of 15, and a data collection age range of 15 to 49. Figure 4.17a shows the Spectrum software’s resulting CD4 compartment counts for such a population; Figure 4.17b shows the MASHIV software’s CD4 compartment counts. The Spectrum software transitions people between compartments by moving a fixed fraction of the aggregate compartment population to the next compartment. This results in a convex or exponential distribution to the size of the CD4 >500 compartment. However, the MASHIV software transitions are based upon a normal hazard function which varies the probability of transition over time as to produce a normal distribution of transitions about the expected mean compartment duration. This results in a concave or normal distribution to the size of the CD4 >500 compartment. Why is this difference important? The CD4 >500 compartments represents the longest duration compartments with the lowest mortality
Figure 4.15: Comparison of the Effect of Progression Rate on Mortality for Different Ages of Initial Infection

(a) Slow-progression

(b) Fast-progression
Figure 4.16: Spectrum Mortality Distribution Based Upon Infection Start Age

(a) Slow-progression

(b) Fast-progression
rates of non-ART compartments. If people are transitioned too quickly, they will progress to higher mortality compartments and more rapidly be eliminated from the population due to a HIV-related death. This would lead to an overestimation of mortality, and produce a shorter estimation of the mean number of years till death.

In both Spectrum and MASHIV, for each CD4 group, we generate 10,000 infections with an infection starting of 15 years and observe the percent of the infections that have transitioned from their initial CD4 group over time. In Table 4.21 we summarize the results of these trials. At the mean years, we would ideally see 50% of individuals transitioned out of the initial CD4 compartment. However, by using the standard uniform probability, Spectrum overestimates the transition rate before the mean, consistently producing approximately 63% transitions at the mean. If the model transitioned exactly 50% at the mean, then the ratio of the time to reach 50% transitions divided by the expected mean duration will equal 1. Given that Spectrum overestimates the transitions before the mean, this ratio of the 50% duration to the mean duration is naturally less than 1. Our normal hazard function approach, however, is designed to ensure 50% of the transitions occur at the mean, allowing for some stochastic variation. While Spectrum overestimates the percent transitioned by 13% on average, MASHIV underestimates the percent transitioned by less than 4% on average. Thus, for the purpose of our CD4 compartment transitions, we deem our normal hazard approach superior to the uniform probability approach used in the Spectrum software.

4.6.3 HIV Mortality without ART

The mortality rates we use to inform our model are provided in four age groups–15 to 24, 25 to 34, 35 to 44, and 45+. Thus, it is important to observe the mortality distributions that span these groups. For each age in the range, we simulate 1,000 HIV infections with a slow-progression rate. We chose slow-progression rates for study since the Spectrum calculation of median time from infection to AIDS death does not appear to be influenced by the distribution of slow and fast progressions in the Spectrum software interface.

In Figure 4.18 we show the mortality distribution for the age group 15-24. Figure 4.19 is the mortality distribution for the age group 25-34. Figure 4.20 is the mortality distribution for the age group 35-44. Figure 4.21 is the mortality distribution for the age group 45+. The line over each histogram represents the Spectrum distribution of deaths for a similar initial configuration with approximately 1,000 HIV slow-progression infections for age in the specified age range.

Table 4.22 provides the median times till AIDS death for the four age groups. The Spectrum software values represent the median time till AIDS death for the default HIV model for countries in the Asia region. The MASHIV values are calculated during the previously mentioned computation that generated Figures 4.18, 4.19, 4.20, and 4.21. In probabilistic mode, MASHIV produces median values below the Spectrum median in the 15-24 age group and above the Spectrum median in the remaining age groups. The bulk of the difference can be attributed to the difference in methods.
Figure 4.17: Comparison of CD4 Distributions Between Spectrum and MASHIV
Figure 4.18: MASHIV Mortality Distribution for Ages 15-24

Figure 4.19: MASHIV Mortality Distribution for Ages 25-34
Figure 4.20: MASHIV Mortality Distribution for Ages 35-44

Figure 4.21: MASHIV Mortality Distribution for Ages 45-56
for assigning CD4 compartment transitions. MASHIV uses a normal hazard function to model
daily probabilities of transmission whereas Spectrum uses a daily rate of transition imposed on the
aggregate population.

Part of the difference may arise from the method Spectrum uses to calculate their median time
till death for these age groups. However, as Spectrum is not an open source software, we do not
know the manner in which they calculate this median.

Next, we examine 10,000 men starting their infection at age 15. For this case, Figure 4.22
displays the annual deaths, cumulative deaths, and HIV population count. These graphs demon-
strate a strong correlation between the results of the Spectrum and MASHIV models as for HIV
progression toward mortality.

In our second case, we examine 10,000 men starting their infection at age 25. For this case,
Figure 4.23 displays the annual deaths, cumulative deaths, and HIV population count. Again, these
graphs show a strong correlation between the Spectrum and MASHIV in regard to HIV progression
toward mortality. However, in this case, MASHIV appears to underestimate the number of deaths
between the early and later years when the graphs intersect.

In our final case, we examine 10,000 men starting their infection at age 35. For this case, Figure
4.24 displays the annual deaths, cumulative deaths, and HIV population count. Compared to Spec-
trum, MASHIV initially underestimates the number of deaths compared to Spectrum. However,
as the number of years exceeds 7.5, MASHIV overestimates the number of HIV deaths.

4.6.4 HIV Progression with ART

Prior to the initiation of ART, the HIV progression is composed of CD4 progression and mortality.
However, when ART begins in the CD4 compartment model, the CD4 progression halts. Subse-
quently, the mortality rates affecting the individual are based solely upon the CD4 compartment
at which they initiated ART. As such, the validation of the HIV progression with ART is primarily
the validation of the mortality rates affecting ART which has been previously discussed. Mortal-
ity for ART is split into three time phases–0 to 6 months, 6 to 12 months, and greater than 12
months. Each time phase defines a period of different mortality rates. In Figure 4.25 we present
an example comparison of Spectrum and MASHIV with regard to HIV progression under ART. In
both scenarios, 100% of the population is designated to initiate ART when the CD4 level reaches
the CD4 250-349 range.

4.7 Benchmarks for Normal Hazard Functions

In this section we examine the benefit of pre-calculating and storing the daily probabilities produced
by the normal hazard function we used to determine when compartment transitions occur in the
CD4 model. Unlike the exponential hazard function daily mortality probabilities that are constant
Figure 4.22: Spectrum and MASHIV Comparison with Starting Age at 15
Figure 4.23: Spectrum and MASHIV Comparison with Starting Age at 25
Figure 4.24: Spectrum and MASHIV Comparison with Starting Age at 35
Figure 4.25: Comparison of Spectrum and MASHIV CD4 Distributions with 100% ART at CD4 250-349
over time, the normal hazard probabilities are time-variant and require a costly estimation of an integral in the error function. Additionally, it is natural to consider pre-calculation these probabilities since every infection will require the same daily probabilities, though individual infections will only use as many probabilities as the number of days necessary to produce a transition. For a crude estimate of the number of how many values to store, we store 3.5 times the mean number of days or 3650 days (10 years) whichever is less. The maximum of 10 years is a natural limit to assume since the model defines a different transition mean for each 10 year period. Hence, for all except the last age group, ages 45 plus, it is impossible for an agent to utilize more than 10 years’ worth of daily probabilities within a single age group. While we have observed that pre-calculating and storing these daily probabilities produces a marked speed up, in this section we quantify how much improvement of performance is actually gained.

Firstly, we establish a baseline measurement without the pre-calculation of daily transition probabilities. Thus, every time a daily probability of transition is needed, it will be a fresh calculation. We create 20,000 HIV infections directly, without the complications of dating or the agent objects. All infections are set to be slow-progression with the asymptomatic period initiating with a CD4 count $>500$. We update the HIV progression for a total of 30 years’ worth of daily timesteps.

Next, we generate 20,000 new infections with the same properties as the baseline test, reset the random number generator to the same initial value as the earlier test, and enable the use of stored daily probabilities. With these conditions in place, we repeat the test on the 20,000 infections. In theory, the reset of the random number generator should ensure the mortality pattern and number of probabilities used is consistent between the two tests. The results of the tests are summarized in Table 4.23. By comparing the durations of the two tests, we calculated a 73.9% reduction in time by using the cached values. Additionally, the number of probabilities used matched exactly between the two tests confirming identical mortality and transition usage. While the calculation of the stored daily probabilities does require several seconds that are not included in the test timing, the stored values are generated at the start of the MASHIV program and can be referenced by all simulations run in the user’s session.

Given the above test results, we conclude that pre-calculation of these daily probabilities is definitely worthwhile. While the initial calculation does consume both time and heap space, we consider the expense to be negligible due to the frequency of use and resultant speed increase.
### Table 4.13: Agent State

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>Integer</td>
<td>A unique index for the agent.</td>
</tr>
<tr>
<td>age</td>
<td>Integer</td>
<td>The age (in days) of the agent.</td>
</tr>
<tr>
<td>alive</td>
<td>Boolean</td>
<td>True if the agent is alive; False if the agent has died.</td>
</tr>
<tr>
<td>knownInfected</td>
<td>Boolean</td>
<td>True if the agent knows he is HIV+; False if the agent believes he is HIV-</td>
</tr>
<tr>
<td>hivStatus</td>
<td>HIVInstance</td>
<td>Null if the agent is HIV-; Instantiated to an HIVInstance object if the agent is HIV+</td>
</tr>
<tr>
<td>sexualRole</td>
<td>Insertive, Receptive, Versatile</td>
<td>The preferred sexual role of the agent.</td>
</tr>
<tr>
<td>currentRelationships</td>
<td>Relationship[]</td>
<td>A list of all current relationships.</td>
</tr>
<tr>
<td>allPartnerIds</td>
<td>Integer[]</td>
<td>A set containing all partner ids for both current and past relationships.</td>
</tr>
<tr>
<td>onPrEP</td>
<td>Boolean</td>
<td>True if the agent is assigned to be on PrEP; False otherwise.</td>
</tr>
<tr>
<td>negotiatedSafety</td>
<td>Boolean</td>
<td>True if the agent practices negotiated safety arrangements with steady partners; False otherwise.</td>
</tr>
<tr>
<td>datingMode</td>
<td>Casual, Steady, None</td>
<td>None if the agent is not seeking any relationships; Casual if the agent is seeking a casual relationship; Steady if the agent is seeking a steady relationship.</td>
</tr>
<tr>
<td>condomModeSteady</td>
<td>Always, Never, Sometimes</td>
<td>Condom use policy for steady relationships. Always if the agent will always use a condom; Never if the agent never uses a condom; Sometimes if the Agent uses a condom based upon his population’s probability of condom use.</td>
</tr>
<tr>
<td>condomModeCasual</td>
<td>Always, Never, Sometimes</td>
<td>Condom use policy for casual relationships. Always if the agent will always use a condom; Never if the agent never uses a condom; Sometimes if the Agent uses a condom based upon his population’s probability of condom use.</td>
</tr>
</tbody>
</table>
### Table 4.14: HIV Instance

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>stage</td>
<td>PHI, Asymptomatic, AIDS</td>
<td>The stage of HIV Infection.</td>
</tr>
<tr>
<td>cd4Group</td>
<td>&gt;500, 350-500, 250-349, 200-249, 100-199, 50-99, &lt;50</td>
<td>The CD4+ group of the infection.</td>
</tr>
<tr>
<td>onART</td>
<td>Boolean</td>
<td>True if the agent is on ART; False if the agent not on ART.</td>
</tr>
<tr>
<td>symptomatic</td>
<td>Boolean</td>
<td>True if the agent’s PHI period is symptomatic; False otherwise.</td>
</tr>
<tr>
<td>infectionSource</td>
<td>Agent</td>
<td>The agent which introduced this agent’s infection.</td>
</tr>
<tr>
<td>infectionSourceStage</td>
<td>PHI, Asymptomatic, AIDS</td>
<td>The stage of the HIV infection that caused this agent’s infection.</td>
</tr>
</tbody>
</table>

### Table 4.15: Relationship Object

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>startDay</td>
<td>Integer</td>
<td>The day the relationship started.</td>
</tr>
<tr>
<td>endDay</td>
<td>Integer</td>
<td>The day the relationship ended.</td>
</tr>
<tr>
<td>partners</td>
<td>Agent[]</td>
<td>The agents in the relationship.</td>
</tr>
<tr>
<td>sexRoles</td>
<td>SexualRole[]</td>
<td>The sexual role of each partner, either Insertive, Receptive, or Versatile.</td>
</tr>
<tr>
<td>schemas</td>
<td>RelationshipSchema[]</td>
<td>The relationship schemas of each partner for this relationship.</td>
</tr>
<tr>
<td>steady</td>
<td>Boolean</td>
<td>True if the relationship is a steady relationship; False otherwise.</td>
</tr>
<tr>
<td>unprotected</td>
<td>Boolean</td>
<td>True if the partners have engaged in UAI during this relationship; False otherwise.</td>
</tr>
</tbody>
</table>
Table 4.16: Relationship Schema

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Parameter</th>
<th>Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_C$</td>
<td>probCondomUse</td>
<td>Double</td>
<td>The proportion of condom use.</td>
</tr>
<tr>
<td>$F$</td>
<td>rateFormation</td>
<td>Double</td>
<td>The number of relationships of this type each year.</td>
</tr>
<tr>
<td>$F_{Steady}$</td>
<td>rateFormationWhenSteady</td>
<td>Double</td>
<td>The number of relationships of this type each year when the agent already has a steady relationship.</td>
</tr>
<tr>
<td>$S$</td>
<td>sexRate</td>
<td>Double</td>
<td>The number of sex acts per year for this type of relationship.</td>
</tr>
</tbody>
</table>

Table 4.17: Mixing Weights

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W_{NP}$</td>
<td>Double</td>
<td>The weight for differences in the number of preferred partners.</td>
</tr>
<tr>
<td>$W_{Age}$</td>
<td>Double</td>
<td>The weight for age difference.</td>
</tr>
<tr>
<td>$W_{Pop}$</td>
<td>Double</td>
<td>The weight for assortative preference towards one’s own population group.</td>
</tr>
<tr>
<td>$W_{SS}$</td>
<td>Double</td>
<td>The weight for serosorting based on perceived HIV status.</td>
</tr>
<tr>
<td>$W_{PP}$</td>
<td>Double</td>
<td>The weight for preferring a casual partner who has previously been used.</td>
</tr>
</tbody>
</table>

Table 4.18: Mixing Functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Type</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$NPP(A_i)$</td>
<td>Integer</td>
<td>Returns the number of preferred partners of agent $A_i$.</td>
</tr>
<tr>
<td>$Age(A_i)$</td>
<td>Integer</td>
<td>Returns the age in days of agent $A_i$.</td>
</tr>
<tr>
<td>$SamePop(A_i, A_j)$</td>
<td>Boolean</td>
<td>True if agents $A_i$ and $A_j$ are both from the same population group; False otherwise.</td>
</tr>
<tr>
<td>$Previous(A_i, A_j)$</td>
<td>Boolean</td>
<td>True if agent $A_j$ is a previous casual partner of $A_i$; False otherwise.</td>
</tr>
<tr>
<td>$SameHIV(A_i, A_j)$</td>
<td>Boolean</td>
<td>True if agents $A_i$ and $A_j$ both have the same HIV status; False otherwise.</td>
</tr>
</tbody>
</table>
Table 4.19: Sexual Roles

<table>
<thead>
<tr>
<th>Agent Role</th>
<th>Partner Role</th>
<th>Agent Role at Sexual Act</th>
</tr>
</thead>
<tbody>
<tr>
<td>Versatile</td>
<td>Receptive</td>
<td>Insertive</td>
</tr>
<tr>
<td>Versatile</td>
<td>Insertive</td>
<td>Receptive</td>
</tr>
<tr>
<td>Versatile</td>
<td>Versatile</td>
<td>50% Receptive, 50% Insertive</td>
</tr>
<tr>
<td>Receptive</td>
<td>Insertive</td>
<td>Receptive</td>
</tr>
<tr>
<td>Insertive</td>
<td>Receptive</td>
<td>Insertive</td>
</tr>
</tbody>
</table>

Table 4.20: Helper Functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Type</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>hasSteady($A_i$)</td>
<td>Boolean</td>
<td>True if $A_i$ has a steady relationship; False otherwise.</td>
</tr>
<tr>
<td>areSteady($A_i, A_j$)</td>
<td>Boolean</td>
<td>True if $A_i$ and $A_j$ are in a steady relationship; False otherwise.</td>
</tr>
<tr>
<td>hadUAI($A_i, A_j$)</td>
<td>Boolean</td>
<td>True if $A_i$ and $A_j$ have had unprotected intercourse in this relationship; False otherwise.</td>
</tr>
<tr>
<td>knownHIV($A_i$)</td>
<td>Boolean</td>
<td>True if $A_i$ has tested HIV-positive, False otherwise.</td>
</tr>
<tr>
<td>durKnownPoz($A_i$)</td>
<td>Double</td>
<td>If knownHIV($A_i$) == True, then returns duration (in years) since day of HIV-positive HIV test; 0 otherwise.</td>
</tr>
</tbody>
</table>

Table 4.21: CD4 Progression with Infection Starting Age of 15

<table>
<thead>
<tr>
<th>CD4 Count</th>
<th>50% Transitioned Duration / Mean</th>
<th>% Transitioned at Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spectrum</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;500</td>
<td>0.469</td>
<td>63.34</td>
</tr>
<tr>
<td>350-500</td>
<td>0.649</td>
<td>62.98</td>
</tr>
<tr>
<td>250-349</td>
<td>0.473</td>
<td>62.83</td>
</tr>
<tr>
<td>200-249</td>
<td>0.292</td>
<td>63.05</td>
</tr>
<tr>
<td>100-199</td>
<td>0.587</td>
<td>63.02</td>
</tr>
<tr>
<td>50-99</td>
<td>0.003</td>
<td>62.93</td>
</tr>
<tr>
<td><strong>MASHIV</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;500</td>
<td>1.05</td>
<td>46.88</td>
</tr>
<tr>
<td>350-500</td>
<td>1.04</td>
<td>47.49</td>
</tr>
<tr>
<td>250-349</td>
<td>1.05</td>
<td>46.61</td>
</tr>
<tr>
<td>200-249</td>
<td>1.05</td>
<td>47.03</td>
</tr>
<tr>
<td>100-199</td>
<td>1.05</td>
<td>46.96</td>
</tr>
<tr>
<td>50-99</td>
<td>1.04</td>
<td>47.95</td>
</tr>
</tbody>
</table>
Table 4.22: Comparison of Median Time From Infection to AIDS Death

<table>
<thead>
<tr>
<th>Software</th>
<th>Age Group</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15-24</td>
<td>25-34</td>
<td>35-44</td>
<td>45+</td>
</tr>
<tr>
<td>Spectrum</td>
<td>16.4</td>
<td>12.0</td>
<td>9.20</td>
<td>7.40</td>
</tr>
<tr>
<td>MASHIV (Probabilistic-based)</td>
<td>14.56</td>
<td>13.32</td>
<td>11.09</td>
<td>10.26</td>
</tr>
<tr>
<td>MASHIV (Duration-based)</td>
<td>16.85</td>
<td>13.90</td>
<td>11.15</td>
<td>10.62</td>
</tr>
</tbody>
</table>

Table 4.23: Test Results

<table>
<thead>
<tr>
<th>Test</th>
<th># Probabilities Used</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fresh Calculations</td>
<td>109,928,307</td>
<td>35 seconds</td>
</tr>
<tr>
<td>Stored Calculations</td>
<td>109,928,307</td>
<td>9 sec</td>
</tr>
<tr>
<td>Reduction</td>
<td></td>
<td>73.9%</td>
</tr>
</tbody>
</table>
Table 4.24: Analysis of Caching Usage

<table>
<thead>
<tr>
<th>CD4 Group</th>
<th># Days to 95%</th>
<th>Fraction of Mean @ 95%</th>
<th>% use @ 10yr</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Start Age of Infection: 15</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;500</td>
<td>NA</td>
<td>NA</td>
<td>59.80</td>
</tr>
<tr>
<td>350-500</td>
<td>2936</td>
<td>1.79</td>
<td>98.95</td>
</tr>
<tr>
<td>250-349</td>
<td>2541</td>
<td>2.05</td>
<td>99.89</td>
</tr>
<tr>
<td>200-249</td>
<td>1449</td>
<td>2.02</td>
<td>100.00</td>
</tr>
<tr>
<td>100-199</td>
<td>2876</td>
<td>1.69</td>
<td>99.08</td>
</tr>
<tr>
<td>50-99</td>
<td>1640</td>
<td>1.57</td>
<td>100.00</td>
</tr>
<tr>
<td>&lt;50</td>
<td>1528</td>
<td>NA</td>
<td>99.92</td>
</tr>
<tr>
<td><strong>Start Age of Infection: 25</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;500</td>
<td>NA</td>
<td>NA</td>
<td>75.34</td>
</tr>
<tr>
<td>350-500</td>
<td>2812</td>
<td>1.85</td>
<td>99.42</td>
</tr>
<tr>
<td>250-349</td>
<td>1664</td>
<td>2.06</td>
<td>100.00</td>
</tr>
<tr>
<td>200-249</td>
<td>688</td>
<td>2.05</td>
<td>100.00</td>
</tr>
<tr>
<td>100-199</td>
<td>1090</td>
<td>1.90</td>
<td>100.00</td>
</tr>
<tr>
<td>50-99</td>
<td>472</td>
<td>1.87</td>
<td>100.00</td>
</tr>
<tr>
<td>&lt;50</td>
<td>802</td>
<td>NA</td>
<td>100.00</td>
</tr>
<tr>
<td><strong>Start Age of Infection: 35</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;500</td>
<td>NA</td>
<td>NA</td>
<td>89.79</td>
</tr>
<tr>
<td>350-500</td>
<td>2121</td>
<td>2.06</td>
<td>100.00</td>
</tr>
<tr>
<td>250-349</td>
<td>1280</td>
<td>2.04</td>
<td>100.00</td>
</tr>
<tr>
<td>200-249</td>
<td>602</td>
<td>2.06</td>
<td>100.00</td>
</tr>
<tr>
<td>100-199</td>
<td>1008</td>
<td>1.87</td>
<td>100.00</td>
</tr>
<tr>
<td>50-99</td>
<td>451</td>
<td>1.79</td>
<td>100.00</td>
</tr>
<tr>
<td>&lt;50</td>
<td>583</td>
<td>NA</td>
<td>100.00</td>
</tr>
<tr>
<td><strong>Start Age of Infection: 45</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;500</td>
<td>3597</td>
<td>2.10</td>
<td>95.49</td>
</tr>
<tr>
<td>350-500</td>
<td>1418</td>
<td>2.08</td>
<td>100.00</td>
</tr>
<tr>
<td>250-349</td>
<td>878</td>
<td>2.06</td>
<td>100.00</td>
</tr>
<tr>
<td>200-249</td>
<td>412</td>
<td>2.05</td>
<td>100.00</td>
</tr>
<tr>
<td>100-199</td>
<td>758</td>
<td>1.98</td>
<td>100.00</td>
</tr>
<tr>
<td>50-99</td>
<td>351</td>
<td>1.92</td>
<td>100.00</td>
</tr>
<tr>
<td>&lt;50</td>
<td>844</td>
<td>NA</td>
<td>100.00</td>
</tr>
</tbody>
</table>

NA indicates the value is not available. This occurs if the 95% limit is not reached, or in the case of CD4 <50, there is no mean since a transition will never occur.
In this chapter we test the impact of model parameters on certain collected metrics such as the cumulative number of HIV infections or deaths over the course of the simulation. We perform tests accounting for not only parameter variation but also stochastic variation originating from the probabilistic nature of the MASHIV simulator. We perform two sets of tests, one with duration-based dating and one with probabilistic dating.

For the duration-based dataset, we utilize both duration-based HIV transitions and duration-based dating. Thus, the HIV would progress at uniform rates for all infections and dating events occur at regular intervals. Dating events include relationship formation, relationship dissolution, and sex within a relationship. Although this set is primarily based upon fixed intervals between events, stochastic variation does occur at initialization, during relationship selection, and due to the effects of daily mortality.

For the probabilistic dataset, we utilize both probabilistic HIV transitions and probabilistic dating. This set uses our normal hazard function to establish when transitions in the CD4 compartment model occur as well as for dating events. These settings allow for a normal distribution of HIV progression and relationship events.

### 5.1 Dataset Definition

For the tests in this chapter we utilize a simplified population dataset with a single population of 10,000 agents as defined in Table 5.1. Parameters not defined in Table 5.1 are zero valued or otherwise disabled. Test cases in this section are variations upon this dataset created by using the aggregate simulator tool in MASHIV. Where applicable, we calculate a Pearson correlation coefficient for the parameter being varied against the simulator metric tabulated. In such cases, we also attempt to fit a line to the data points generated.

### 5.2 Stochastic Variation

Stochastic, or probabilistic, variation occurs when we utilize a different seed value to initialize the random number generator. For a given seed value, JAVA will produce a deterministic pseudo-random sequence of random samples. In our model, we use random number generators to assign the initial distribution of demographic characteristics, determine when mortality occurs, and, under the probabilistic modes, determine when events such as CD4 progression and dating occur. Since these simulation properties depend on the distribution of random numbers, we must consider to what degree the results vary based upon the choice of a random seed alone. Thus, in this section
Table 5.1: Simulation Parameters

<table>
<thead>
<tr>
<th>Group</th>
<th>Lower Age</th>
<th>Upper Age</th>
<th># People</th>
<th># HIV</th>
</tr>
</thead>
<tbody>
<tr>
<td># 1</td>
<td>15</td>
<td>45</td>
<td>10,000</td>
<td>1,000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value (Sd)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Operational Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td># Years</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Symptomatic Discovery</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>Replace Removed Agents</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>% PHI Initial</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Steady UAI</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td><strong>Sexual Roles</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Insertive</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>% Receptive</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>% Versatile</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Casual Relationships</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td># / Year</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>% Condom Use</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td><strong>Steady Relationships</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Duration</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>Sex / Year</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>% Condom Use</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td><strong>Mixing Weights</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Sero-sorting</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td><strong>Other Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preferred # Partners</td>
<td>1.181 (2.210)</td>
<td>[79]</td>
</tr>
<tr>
<td>HIV Testing Rate</td>
<td>1.0</td>
<td></td>
</tr>
</tbody>
</table>
we study the effects of the random number generator by using a fixed dataset and repeating the simulation with 1000 different random seeds.

In the duration-based set, the sources of probabilistic variation arise from the initial random sampling which assigns population attributes to the agents, the initial assignment of HIV infections, partner selection for relationships, and mortality from HIV. Mortality for each CD4 compartment is governed by an exponential hazard function which produces a daily probability that remains fixed for the duration of time within the CD4 compartment. For the probabilistic set, the impact of probabilistic variation may be more pronounced here than for a duration-event driven dataset which utilizes duration-based CD4 compartment transitions and duration-based dating.

5.2.1 Duration-Based Mode Results

In Figure 5.1a and Figure 5.1b we present the HIV prevalence and HIV mortality respectively for the test set run with 1000 different random seeds with duration-based dating mode. The graphs include confidence bounds as well as the maximum and minimum values observed for each year of the simulation. The graphs demonstrate that there is not considerable variation due to the choice of random seed. The prevalence varies less than 5% about the mean and cumulative number of deaths varies less than 10% about the mean by the end of the simulation.

5.2.2 Probabilistic Mode Results

In Figure 5.2a and Figure 5.2b we present the HIV prevalence and HIV mortality respectively for the test set run with 1000 different random seeds with probabilistic dating. The graphs include confidence bounds as well as the maximum and minimum values observed for each year of the simulation. As the graphs demonstrate, for this level of sexual activity, there is relatively low variance about the mean. The number of infections varies from approximately 1500 to 1600 by the end of the 15 years, producing a difference in HIV prevalence of about 1%. However, we suspect that for simulations with larger degrees of sexual activity, the bounds will widen on the variance.

5.3 Nascent Epidemics

In this section, we reduce the number of initial infections to low levels and observe the growth of the epidemic over a 20 year period. In the first case we start the epidemic with 10 infections, which equates to 0.1% HIV prevalence for the population. In the second case we start the epidemic with 100 infections, which equates to 1% HIV prevalence.

5.3.1 Epidemic with 10 Infections

For the duration-based dating mode, the epidemic started with 0.1% HIV prevalence resulted in 14% HIV prevalence after 20 years. For probability-based dating, the same initial conditions
Figure 5.1: HIV Prevalence and Mortality of the Stochastic Variation Test with Duration-based Dating
Figure 5.2: HIV Prevalence and Mortality of the Stochastic Variation Test with Probabilistic Dating
resulted in 6% HIV prevalence after 20 years. The HIV prevalence for these two cases is displayed in Figure 5.3. The HIV related deaths are displayed in Figure 5.4. Figure 5.5 displays the HIV prevalence distribution according to the agents’ preferred number of concurrent partners. This graph demonstrates that agents with an higher number of concurrent partners are at increased risk of contracting HIV.

5.3.2 Epidemic with 100 Infections

For the duration-based dating mode, the epidemic started with 1.0% HIV prevalence resulted in 57% HIV prevalence after 20 years. For probability-based dating, the same initial conditions resulted in 35% HIV prevalence after 20 years. The HIV prevalence for these two cases is displayed in Figure 5.6. The HIV related deaths are displayed in Figure 5.7. Figure 5.8 displays the HIV prevalence distribution according to the agents’ preferred number of concurrent partners. Again, this graph demonstrates that agents with an higher number of concurrent partners are at increased risk of contracting HIV.

5.4 Parameter Variation

In this section, we vary parameter values for our base dataset with 10% initial HIV prevalence and observe the effects on the number of infections after 15 years.

5.4.1 PHI Percentage

In this test, we vary the percentage of the initial HIV infections that are in the PHI stage of the HIV infection. The PHI stage has a higher percentage of HIV transmission but only lasts 90 days in our simulation. Due to this short duration, PHI stage initial infections are distributed evenly throughout the first year of the simulation. For this test, we vary the percentage of the initial HIV infections from 0 to 60 at intervals of 10%. We limited our tests to 60% initial PHI as we deemed it unlikely that a population with 10% HIV prevalence would have over 60% of their infections in the PHI stage.

For duration-based dating, Figure 5.9a shows the relationship between the cumulative number of HIV infections at 15 years to the initial percentage of PHI stage infections. The similar graph for the same scenario in probabilistic dating mode is shown in Figure 5.9b. Each column of dots represents results for the same % PHI, but different random seeds. For duration-based dating, we calculated a Pearson correlation coefficient of 0.53 and a line-fit slope of 3.056. For probabilistic dating, we calculated a Pearson correlation coefficient of 0.649 and a line-fit slope of 4.14. Thus, while there is an influence from the initial PHI percentage, the overall effect is not substantial for this set of initial conditions.
Figure 5.3: HIV Prevalence for Nascent Epidemic of 10 Infections
(a) Duration-Based Dating Mode

(b) Probability-Based Dating Mode

Figure 5.4: HIV-related Deaths for Nascent Epidemic of 10 Infections
Figure 5.5: HIV Prevalence by Preferred # Partners for Nascent Epidemic of 10 Infections
(a) Duration-Based Dating Mode

(b) Probability-Based Dating Mode

Figure 5.6: HIV Prevalence for Nascent Epidemic of 100 Infections
Figure 5.7: HIV-related Deaths for Nascent Epidemic of 100 Infections
Figure 5.8: HIV Prevalence by Preferred # Partners for Nascent Epidemic of 100 Infections
(a) Duration-Based Dating Mode

(b) Probability-Based Dating Mode

Figure 5.9: Effect of PHI Percentage on the Cumulative Number of Infections at 15 Years
5.4.2 Circumcision

In the following set of tests we observe the effect of different levels of population circumcision on the HIV incidence of the population. We vary the percent of the population that is circumcised from 0% to 100%. We perform this sequence of tests for both duration-based dating and probabilistic dating.

For duration-based dating, Figure 5.10a shows the relationship between the cumulative number of HIV infections at 15 years to the percentage of the population circumcised. The similar graph for the same scenario in probabilistic dating mode is shown in Figure 5.10b. Each column of dots represents results for the same % circumcision, but different random seeds. For duration-based dating, we calculated a Pearson correlation coefficient of -0.997 and a line-fit slope of -44.451. For probabilistic dating, we calculated a Pearson correlation coefficient of -0.997 and a line-fit slope of -38.994. As expected, these tests demonstrate that the reduced probability of transmission we use for circumcision does in fact produce a reduction in the number of HIV infections.

5.4.3 PrEP

In the following set of tests we observe the effect of different levels of PrEP coverage and PrEP adherence on the HIV incidence of the population. We perform three sequences of tests—30% adherence, 50% adherence, and 70% adherence. For each sequence of tests on an adherence level, we test values of PrEP coverage from 0% to 100%, at equal intervals of 10%. We perform this sequence of tests for both duration-based dating and probability-based dating. All tests are repeated with 100 different random seeds to test for stochastic variation.

Figure 5.11, Figure 5.12, and Figure 5.13 show the relationship between the cumulative number of HIV infections at 15 years to the PrEP coverage for 30% PrEP adherence, 50% PrEP adherence, and 70% PrEP adherence respectively. While the probability-based dating mode produces more infections than the duration-based mode, both versions produce similar Pearson correlation coefficients, provided in Table 5.2. Additionally, as expected, these tests demonstrate that the reduced probability of contracting HIV we use for PrEP does in fact produce a reduction in the number of HIV infections and that greater degrees of PrEP coverage do produce a corresponding reduction in the number of HIV infections.

5.4.4 Condom Use

In the following set of tests we observe the effect of different levels of condom use on the HIV incidence of the population. We vary the probability of condom use from 0% to 100% at equal intervals of 10%. We perform this sequence of tests for both duration-based dating and probability-based dating. All tests are repeated with 100 different random seeds to test for stochastic variation.

For duration-based dating, Figure 5.14a shows the relationship between the cumulative number
Figure 5.10: Effect of Circumcision on the Cumulative Number of Infections at 15 Years

(a) Duration-Based Dating Mode

(b) Probability-Based Dating Mode
Figure 5.11: Effect of PrEP on the Cumulative Number of Infections at 15 Years with 30% PrEP Coverage

(a) Duration-Based Dating Mode

(b) Probability-Based Dating Mode

111
Figure 5.12: Effect of PrEP on the Cumulative Number of Infections at 15 Years with 50% PrEP Coverage

(a) Duration-Based Dating Mode

(b) Probability-Based Dating Mode
Figure 5.13: Effect of PrEP on the Cumulative Number of Infections at 15 Years with 70% PrEP Coverage

(a) Duration-Based Dating Mode

(b) Probability-Based Dating Mode
### Table 5.2: Pearson Correlation Coefficients and Slope for PrEP Tests

<table>
<thead>
<tr>
<th></th>
<th>30% Coverage</th>
<th>50% Coverage</th>
<th>70% Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration-based dating</td>
<td>-0.969</td>
<td>-0.990</td>
<td>-0.996</td>
</tr>
<tr>
<td>Probabilistic dating</td>
<td>-0.980</td>
<td>-0.994</td>
<td>-0.996</td>
</tr>
<tr>
<td>Slope</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration-based dating</td>
<td>-15.416</td>
<td>-29.358</td>
<td>-44.332</td>
</tr>
</tbody>
</table>

of HIV infections at 15 years to the percentage of condom use. The similar graph for the same scenario in probability-based dating mode is shown in Figure 5.14b. Each column of dots represents results for the same % condom use, but different random seeds. For duration-based dating, we calculated a Pearson correlation coefficient of -0.945 and a line-fit slope of -70.745. For probabilistic dating, we calculated a Pearson correlation coefficient of -0.971 and a line-fit slope of -75.046. As expected, these tests demonstrate that the reduced probability of transmission expected from condom use does in fact produce a reduction in the number of HIV infections.

#### 5.4.5 Risk Reduction with Discovery

In the following set of tests we observe the effect of different levels of risk reduction on the HIV incidence of the population. Risk reduction is applied by increasing the probability of condom use by the same percentage as the risk reduction specified in the dataset. This risk reduction behavior is limited to 1 year from the discovery of HIV for each HIV infected agent. We test percentages of risk reduction from 0% to 50%, at equal intervals of 5%. HIV discovery is accomplished through requiring agents perform an annual HIV test. We perform this sequence of tests for both duration-based dating and probabilistic dating. All tests are repeated with 100 different random seeds to test for stochastic variation.

For duration-based dating, Figure 5.15a shows the relationship between the cumulative number of HIV infections at 15 years to the percentage of risk reduction post HIV discovery. The similar graph for the same scenario in probability-based dating mode is shown in Figure 5.15b. Each column of dots represents results for the same % risk reduction, but different random seeds. For duration-based dating, we calculated a Pearson correlation coefficient of -0.033 and a line-fit slope of 0.211. For probability-based dating, we calculated a Pearson correlation coefficient of 0.012 and a line-fit slope of 0.075. These tests only demonstrate a slight reduction in the number of overall infections correlated to the percentage of risk reduction triggered by HIV discovery. As the asymptomatic phase is the longest, HIV discovery through routine HIV testing likely occurs during the asymptomatic phase. Alternatively, HIV may be discovered when HIV becomes symptomatic in the final progression to AIDS. However, the progression from AIDS to death is relatively short.
Figure 5.14: Effect of Condom Use on the Cumulative Number of Infections at 15 Years

(a) Duration-Based Dating Mode

(b) Probability-Based Dating Mode
compared to the length of the asymptomatic phase. Thus, we conclude that the risk reduction tested here likely occurs within the relatively lower virulence asymptomatic phase, and thus doesn’t result in a great reduction in transmission in the one year window of risk reduction we utilize.

5.4.6 Serosorting

In this section we test the effect of different degrees of serosorting preference on the number of infections. Serosorting is the practice of choosing one’s partner or level of risk to assume with a partner based on the perceived or stated HIV status of the partner. Wilson et al. observed that the effectiveness of serosorting depends on the percent of the population that accurately knows their HIV status [182]. Such low HIV status awareness could be caused by either low HIV testing in the population or a PHI-driven fast spreading HIV epidemic. For this experiment, we vary the percentage of the time that the population practices serosorting. To allow their serosorting preference to affect their dating decisions, we add a mixing weight of 1.0 for serosorting. Thus, serosorting agents will prefer partners who have the same HIV status. We vary the percentage of the time that the population practices serosorting from 0 to 100% by 10% intervals with 100 different random seeds at each value. Three trials are performed, one with only seropositive serosorting, one with only seronegative serosorting, and one with both seropositive and seronegative serosorting. In seropositive serosorting, HIV positive individuals have a percentage of the time that they prefer pairing with other HIV positive individuals. For this case, when a random uniform sample determines that the seropositive individual wishes to date another seropositive individual, then during the evaluation of people for relationships, greater weight will be given toward people who are seropositive. In seronegative serosorting, HIV negative individuals have a percentage of the time that they prefer pairing with other HIV negative individuals. For the case where we have both seropositive and seronegative serosorting, the values for the percentages of each type of serosorting are always the same value.

In Figure 5.16, Figure 5.17, and Figure 5.18, we present the cumulative number of infections with duration-based dating applied to our three scenarios for serosorting. In Figure 5.19, Figure 5.20, Figure 5.21, we present the similar set of graphs for the serosorting tests with probabilistic-based dating. Table 5.3 summarizes the Pearson correlation coefficients and slope estimates for the serosorting tests for both the duration-based dating and probabilistic dating scenarios.
Figure 5.15: Effect of Discovery-based Risk Reduction on the Cumulative Number of Infections at 15 Years
Figure 5.16: Duration-based with Seropositive Serosorting

Figure 5.17: Duration-based with Seronegative Serosorting
Figure 5.18: Duration-based with Both Positive and Negative Serosorting

Figure 5.19: Probabilistic with Seropositive Serosorting
Figure 5.20: Probabilistic with Seronegative Serosorting

Figure 5.21: Probabilistic with Both Positive and Negative Serosorting
Table 5.3: Pearson Correlation Coefficients and Slope for Serosorting Tests

<table>
<thead>
<tr>
<th></th>
<th>Sero+</th>
<th>Sero-</th>
<th>Sero+ And Sero-</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pearson Correlation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration-based dating</td>
<td>-0.936</td>
<td>-0.958</td>
<td>-0.956</td>
</tr>
<tr>
<td>Probability-based dating</td>
<td>-0.957</td>
<td>-0.958</td>
<td>-0.974</td>
</tr>
<tr>
<td><strong>Slope</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration-based dating</td>
<td>-12.175</td>
<td>-14.724</td>
<td>-26.193</td>
</tr>
</tbody>
</table>
In this chapter, we present a simplified dataset for a subset of the MSM population in Bangkok, Thailand and apply the dataset to our MASHIV simulator. The dataset is not intended to be a perfect representation of the MSM HIV epidemic in Bangkok, Thailand. However, we believe that demonstrating the development and processing of our dataset will prove illuminating as to the capabilities of our simulator and the impact of model variables.

1984 marked the first discovered case of AIDS in Thailand—a Thai male who developed AIDS overseas and returned to Thailand [163]. Fifteen of the first seventeen reported AIDS and AIDS Related Complex cases in Thailand were MSM [168]. By 1988, surveillance discovered HIV spreading among IDUs [154] where HIV prevalence skyrocketed from 1% to 44% in a single year [170]. Phylogenetic analysis of HIV infections in Thailand bolstered the theory that there exists two HIV subtype epidemics in Thailand—one for IDU based transmission and another for sexually transmitted cases [184].

Today, the HIV epidemic in Thailand is classified as a generalized epidemic with 1.1% HIV prevalence of adults at reproductive age [20]. A generalized HIV epidemic occurs when HIV prevalence in the general population’s sexual networks is sufficient to drive the epidemic. In the decade preceding 2006, heterosexual HIV transmission accounted for 89% of new infections; IDU transmission accounted for 4.4% [184].

While the overall population HIV prevalence in Thailand is only about 1.1%, at-risk subpopulations such as IDU and MSM have much greater HIV prevalence, especially in the major urban areas. The country snapshot summary for Thailand estimates that HIV prevalence is 16.7 times higher among MSM than in the general population [89]. In 2008, Li et al. reported on a survey of 456 male bisexuals and 1,125 MSM-only individuals divided among Phuket, Bangkok, and Chiang Mai [111]. Li et al., in summarizing existing literature, estimated adult males that engage in MSM behavior represent 3.3% to 16% percent of Thailand’s adult male population. A behavior specific term, MSM includes Thailand’s bisexuals, homosexuals, transgender males, and male sex workers with male clients. HIV prevalence estimates for Bangkok’s MSM population have risen from 17.3% in 2003 to 30.8% in 2007 demonstrating a concerning surge in the number of infections [78]. Table 6.1 provides a list of studies of MSM HIV prevalence in Bangkok. We exclude from this table studies that focused primarily on bisexual male sex workers.

Li et al. reported that 66.7% of the bisexual respondents reported 6 or more partners in the past 3 months compared to the MSM-only who reported 27.4% [111]. In regard to other risk factors, consistent condom use in the past 3 months among bisexual males with male partners was 77.6% compared to MSM-only with 62.9%.

Additionally, Tieu et al. surveyed 335 heterosexual men in Bangkok and found only 13.4%...
Table 6.1: MSM HIV Prevalence in Bangkok

<table>
<thead>
<tr>
<th>Year</th>
<th>% HIV+</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1985</td>
<td>1</td>
<td>[152]</td>
</tr>
<tr>
<td>1989</td>
<td>3.1</td>
<td>[152]</td>
</tr>
<tr>
<td>2003</td>
<td>17.3</td>
<td>[78]</td>
</tr>
<tr>
<td>2005</td>
<td>28.3</td>
<td>[78]</td>
</tr>
<tr>
<td>2006</td>
<td>17</td>
<td>[116]</td>
</tr>
<tr>
<td>2007</td>
<td>30.8</td>
<td>[78]</td>
</tr>
</tbody>
</table>

of respondents were circumcised [165]. Acceptance of adult circumcision as a method of reducing HIV risk proved to be very low, though Tieu et al. noted greater acceptance after educating the participants in the study. While circumcision can reduce the per contact risk of HIV transmission, increased risky behavior can undermine the protective benefits [37].

6.1 Base Model

In this section we describe the parameter values we have chosen for a simplified Bangkok dataset. A parameter set for MASHIV requires we define population groups and behavioral parameters.

6.1.1 Population Groups

Our base model for Bangkok, Thailand is based primarily off Van Griensven et al.’s summary of a MSM study for the year 2003 [78]. In this study, the authors classified their study group into three age groups–15 to 22, 23 to 28, and above 29. Van Griensven et al.’s age group statistics are presented in Table 6.2. Additionally, as we have data on the HIV prevalence for these age groups, we utilize population-defined initial HIV infections for our model. For our simulation, we utilize a simplified population model of a single group of 10,000 MSM divided proportionately among the three age groups above with an upper limit of 45 years on the eldest age group. The adapted population and HIV prevalence values for our simulation are presented in Table 6.3.

Table 6.2: 2003 Population Groups

<table>
<thead>
<tr>
<th>Age Group</th>
<th>% of Population</th>
<th>% HIV+</th>
</tr>
</thead>
<tbody>
<tr>
<td>15-22</td>
<td>28.4</td>
<td>12.9</td>
</tr>
<tr>
<td>23-28</td>
<td>38.2</td>
<td>17.5</td>
</tr>
<tr>
<td>29+</td>
<td>33.5</td>
<td>20.8</td>
</tr>
</tbody>
</table>

Tieu et al. observed 13.4% of the 335 survey respondents in his study in Bangkok were circumcised [165]. In 2013, Van Griensven reported 10.3% of the 1744 participants were circumcised [76]. We choose to utilize Van Griensven’s circumcision estimate due to the larger sample size.
In regard to HIV testing, Mansergh et al. observed in a study of 927 MSM in Bangkok a 13.5 month mean between HIV testing [116]. This figure equates to an annual testing rate of 0.89 for our model. Additionally, Van Griensven et al. observed that 47.8% of the cohort in the 2007 had never tested for HIV [78]. We interpret this rate for our model as the percentage of people who will never test for HIV, which may overestimate the number of people who will never test.

### 6.1.2 Sexual Roles

We utilize the sexual role percentages for 2003 reported by Van Griensven et al. [78] and reproduce the values in Table 6.4. Since some people do not engage in anal sex, the reported values do not exactly total 100. Hence, we corrected the values proportionately to total 100.

#### Table 6.4: Sexual Roles

<table>
<thead>
<tr>
<th>Role</th>
<th>Percentage (Reported)</th>
<th>Percentage (Corrected)</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Insertive</td>
<td>46.8</td>
<td>47.8</td>
</tr>
<tr>
<td>% Receptive</td>
<td>17.6</td>
<td>18.0</td>
</tr>
<tr>
<td>% Versatile</td>
<td>33.5</td>
<td>34.2</td>
</tr>
</tbody>
</table>

### 6.1.3 Sexual Partners

To estimate the number of sexual partners for 2003, we use Van Griensven et al.’s 2013 analysis of the same cohort for the years from 2006 to 2011 [76], provided in Table 6.5. To use these values in MASHIV, we must make certain assumptions about the distribution of the values in the real world and adapt the values to accommodate our assumptions.

#### Table 6.5: Number of Sexual Partners in 4 months

<table>
<thead>
<tr>
<th># Partners</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4.5</td>
</tr>
<tr>
<td>1-5</td>
<td>55.4</td>
</tr>
<tr>
<td>≥6</td>
<td>40.1</td>
</tr>
</tbody>
</table>
First, we assume that Van Griensven et al.’s values refer to a count of the number of unique sexual partners in the stated four month period and that the actual values follow a normal distribution. Thus, we use Van Griensven et al. values to inform an estimation of an annual normal distribution of the number of sexual partners. To estimate the distribution, we use the values in Table 6.5 to define data points of a normal distribution CDF. With these points we perform a combination of greedy and iterative searching to find an approximation to the mean and standard deviation of a normal distribution that produces similar CDF values. For the CDF data points of (0,4.5) and (5,59.9), we obtain a mean of 4.355 and a standard deviation of 2.57. However, the data points in the table are for a 4 month period, thus we triple the mean and standard deviation to obtain a mean of 13.065 and a standard deviation of 7.706. However, this distribution represents the combined steady and casual partners for respondents.

For casual partners specifically, Holtz et al. discovered from a survey of 1541 MSM in Bangkok that over a 4 month period the number of casual partners for respondents were 25.6% for 0 partners, 50.7% for 1 to 5 partners, and 23.7% for 6 or more partners. [88]. Thus, calculating CDF data points similar to above and utilizing our normal distribution discovery utility we derive a mean of 2.391 and standard deviation of 3.648 for the 4 month period. Tripling this to obtain annual values we obtain a mean of 7.173 and a standard deviation of 10.944.

For our model, we still need an estimate for the average length of steady relationships. Unfortunately there is very little information available on the average length of MSM relationships in Bangkok or other urban areas of Thailand. In lieu, we have chosen to utilize an estimate for the Sydney Australia MSM relationship lengths. In 1997, Kippax et al., reporting on a survey of Sydney MSM, derived the relationship lengths we’ve reproduced in Table 6.6 [95]. In the final column we’ve produced a running total of percentage of relationship comprised by the row’s relationship duration or less. Using this cumulative column and the maximal number of days representing the row’s relationship length, we use our normal distribution discovery utility to derive a mean of 947.6 days and a standard deviation of 1103 days.

<table>
<thead>
<tr>
<th>Length</th>
<th>% of Total</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 months-1 year</td>
<td>20.9</td>
<td>20.9</td>
</tr>
<tr>
<td>1-2 years</td>
<td>31.9</td>
<td>52.8</td>
</tr>
<tr>
<td>3-5 years</td>
<td>22.3</td>
<td>75.1</td>
</tr>
<tr>
<td>&gt;5 years</td>
<td>24.9</td>
<td>100</td>
</tr>
</tbody>
</table>

### 6.1.4 Concurrent Relationships

There are very few surveys which collect detailed information on concurrent relationships of MSM in Bangkok, Thailand. In 2012, Grierson and McNally reported on the findings of the Bangkok
Sauna Study, a survey of 726 MSM attending 6 saunas in Bangkok [79]. One question of this study asked the participants how many regular partners they had at the current time, providing insight into the instantaneous level of concurrency of the population. Of the 15.5% of the population having 2 or more partners, 63.4% had 2 partners, 19.0% had 3 partners, 5.2% had 4 partners, and 12.4% had 5 or more partners. We translated these values into CDF data points and utilized our normal distribution discovery tool to discover a mean of 1.181 and standard deviation of 2.210. Thus we now have an approximation for the percentage of the population who have concurrent relationships and, for those people, a distribution of the number of concurrent relationships sustained.

### 6.1.5 Sexual Contacts

Through a similar process to our normal approximation of the number of sexual partners, we approximate the number of annual sexual acts. To inform this approximation, we use Van Griensven et al.’s surveyed percentages for sexual frequency [77] as values in a normal distribution’s cumulative distribution function. Table 6.7 contains the values we derived from Van Griensven et al. Unfortunately, the percentages reported are responses to a six option sexual frequency questionnaire and do not define true overlapping percentage regions. For example, the percentages for “> once month”, “once month”, and “< once month” should total 100%, but they instead total 63.9%. Therefore, to estimate the CDF values we needed to combine appropriate categories that describe the same region of the CDF curve. For example, Van Griensven et al. reported 21.4% and 26.9% of respondents reported having sex once per month and less than once per month respectively. Summing those two values we obtain the percentage of respondents having sex less than or equal to once per month. We use the number of sexual contacts per year and the cumulative percentages in our greedy algorithm normal distribution discovery utility and derive a mean of 13.8 and a standard deviation of 42.948.

<table>
<thead>
<tr>
<th># Per Year</th>
<th>Cumulative % Respondents Math</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ 12</td>
<td>21.4 + 26.9</td>
<td>48.3</td>
</tr>
<tr>
<td>≤ 52</td>
<td>17.4 + 15.6 + (21.4 + 26.9)</td>
<td>81.3</td>
</tr>
<tr>
<td>≤ 364</td>
<td>100-1.3</td>
<td>98.7</td>
</tr>
</tbody>
</table>

### 6.1.6 Condom Use

For our simulation of Bangkok, we specify the proportion of the population that are consistent condom users and the probability of condom use for sexual contacts of individuals who do not always use condoms. In 2011, Girault stated that, over the a 4 month period, unprotected sex was reported at 54% for steady partners; 49.7% for casual partners [65]. Therefore, in that 4 month period, 46%
of individuals with steady partners always used condoms and 50.3% of individuals with casual partners always used condoms. These values are likely larger than the actual percentages of the population that always use condoms. The surveyed responses likely include people who sometimes do not use condoms but in that particular 4 month period always used condoms. However, we will assume for steady partnerships that 46% of men always use condoms; for casual relationships we estimate 50.3% of men always use condoms.

For the individuals who sometimes use condoms, we need to estimate the probability of condom use for steady and casual partners. The country snapshot for Thailand in 2012 stated that the estimate for the percentage condom use for MSM at their last encounter, including both casual and steady contacts, was 84.5% in 2010 [89]. In a population, the percent of people using condoms at their last sexual contact is comprised of both those who always use condoms and those who sometimes use condoms. Equation 6.1 reflects this relationship. Reorganizing to solve for the probability of condom use, we get Equation 6.2. This equation holds assuming that the percentage of condom use at last sexual contact for inconsistent condom users is equal to the probability of condom use of a single sexual contact. Using Equation 6.2 with the above values for consistent condom users and the overall condom use at last sexual contact, we yield 68.8% probability of condom use for casual partners in Equation 6.3. For steady partners, we assume that steady partners who are not consistent condom users will cease condom use with their partners at some point early in the relationship. Thus, for steady partners who inconsistently use condoms, we assume 0% condom use.

\[ Pct_{LastTime} = Pct_{Always} \times 1.0 + (1 - Pct_{Always}) \times Prob_{CondomUse} \]  

(6.1)

\[ Prob_{CondomUse} = \frac{Pct_{LastTime} - Pct_{Always}}{1 - Pct_{Always}} \]  

(6.2)

\[ Prob_{CondomUse, Casual} = \frac{84.5/100 - 50.3/100}{1 - 50.3/100} = \frac{0.845 - 0.503}{1 - 0.503} = \frac{0.342}{0.497} = 0.688 \]  

(6.3)

6.1.7 PHI Percentage

As we are initializing our population with 17.3% HIV prevalence rate for 2003, we need to specify the percentage of those initial infections that are in the PHI phase. Assuming a closed population without migration, all new infections for 2003 will have passed through the PHI phase at some point in that year. We estimate the HIV incidence for 2003 by calculating the average annual difference
in HIV prevalence between two years we have data for 2003 and 2005. We then calculate the ratio of this annual percentage change in prevalence to the HIV prevalence in 2003 to produce an approximation to the percentage of HIV infections in 2003 that were in PHI phase. This calculation is illustrated in Equation 6.4. To prevent all of the short-lived PHI stages of the HIV infections from occurring at the same time, the MASHIV simulator distributes the start time of the PHI-designated infections to be at uniformly distributed times throughout the first year of the simulation.

\[
\%PHI = 100 \times \frac{Prev_{05} - Prev_{03}}{Prev_{03}} \times \frac{2.0}{28.3 - 17.3} \times \frac{17.3}{5.5} \\
= 100 \times \frac{17.3}{5.5} \\
= 31.79\%
\] (6.4)

### 6.1.8 Parameter Summary

Table 6.8 summarizes the set of parameters for our Bangkok, Thailand dataset. These values are obtained or derived from the literature cited in the relevant sections above. We choose to utilize the duration-based dating model to limit excessive variation in the number of sexual contacts which might interfere with our analysis of concurrent relationships.

### 6.2 Simulator Runs

In this section we present the simulation runs we have performed for the Bangkok dataset we have developed. The dataset and simulation results contained herein should be considered as preliminary. We expect that further data collection, model development, simulation processing, and analysis will continue to improve the usefulness of the Bangkok dataset and MASHIV simulation results. All runs are performed with 100 random seeds per parameter value except for the stochastic variation test which used 1000 random seeds.

#### 6.2.1 Concurrent Relationships

In this section we attempt to see if there is a correlation between the number of concurrent partners and the number of cumulative HIV infections produced. To accomplish this, we fix the annual sexual frequency and vary the number of relationships and the number of concurrent relationships. However, due to the design of the MASHIV simulator the number of sexual acts per year may still vary to some degree. For example, variation in the number of sex acts per year is affected by condom use negotiation. In tests, we have noticed there is still a slight variation in the annual sex frequency with a slight positive correlation to the number of concurrent relationships.
Figure 6.1a demonstrates that the sexual frequency is roughly constant throughout the simulation years. A closer analysis in Figure 6.1b, examining the relationship between the percentage of the population practicing concurrency and the sexual frequency in the final simulation year, yields a Pearson correlation of -0.13 and slope of -0.005. Thus, although there is some variation in the sexual frequency, the overall effect does not appear to be substantial. These variations are likely due to different dating preferences such as when agents of disparate desired sexual frequencies pair together.

Figure 6.2a demonstrates the relationship between the proportion of the population practicing concurrency and the concurrency measure in 180 days. People who practice concurrency have a minimal preference of 2 concurrent steady relationships. As expected, there is a strong linear relationship between these two values, producing a Pearson correlation coefficient of 0.995 and a slope of 0.005. There is also a definite relationship between the concurrency proportion and the cumulative number of infections at the end year yielding a Pearson correlation coefficient of 0.866 and a slope of 1.5. Figure 6.2b demonstrates this correlation. It is worth noting, however, the actual change in the number of cumulative infections between 0% concurrent relationships and 100% concurrent relationships is not large, producing less than an additional 200 infections, or 2% of the original population.

One aspect which may account for the lack of a substantial change in the number of infections is that the overall population HIV prevalence pattern does not adequately match the expected HIV prevalence for Bangkok in the 15 year period modeled. Van Griensven et al. noted an increase from 17.3% HIV prevalence to 28.3% in a 2 year period [78]. However, our model, started at the same 17.3% HIV prevalence, yields slight growth with an overall decline in HIV prevalence. This suggests that our Bangkok dataset, HIV model, and agent model need further refinement.

6.2.2 Stochastic Variation

In this section, we examine stochastic variability of the initial Bangkok dataset arising from different random seeds. Using the same dataset, specified in Table 6.8, we run the simulation with 1,000 different random seeds. For a given random seed, the simulation will always return the same final simulation results. A random seed initializes the random number generator that determines the distribution of the initial population characteristics, the choice of dating pools an agent chooses dates from, and the result of evaluating probabilities that determine whether HIV is transmitted between individuals.

Figure 6.4a demonstrates little growth in the HIV epidemic. According to Van Griensven et al., we should expect 17.3% HIV prevalence to grow to 28.3% in a 2 year period [78], assuming our initial conditions match the population in her study, which we have striven to produce. This suggests either the sexual frequency or the risk of the sexual contacts is not in line with the study population. One area of concern is that, for the non-PHI initial infections, we initialize these initial
Figure 6.1: Analysis of Concurrency and Sexual Frequency

(a) Sexual Frequency

(b) Concurrency vs. Sexual Frequency in End Year
Figure 6.2: Analysis of Concurrency

(a) Concurrency Proportion vs. Concurrency Measure

(b) Concurrency vs. Cumulative Infections
Figure 6.3: Prevalence and Mortality for Concurrency Test

(a) HIV Prevalence

(b) HIV Mortality
infections to the start of the asymptomatic period. Additionally, our current dataset does not include parameters governing the adoption or HIV mitigating benefits of ART. Thus, the agents all progress through the asymptomatic phase from the same starting point in time, which does not represent a realistic distribution of infection states. However, the normal-hazard function that governs HIV progression ensures a distribution of the progression rates and mortality due to HIV. Additional research is needed to specify a more representative distribution of initial infection states and to build this ability into the MASHIV simulator.

6.2.3 Initial PHI

While the PHI stage of HIV is noted for heightened risk of onward HIV transmission, the duration of this stage is relatively short, on the order of weeks to months. For our Bangkok dataset, we estimated that 31.79% of the initial infections would be in the PHI stage. However, the proportion of PHI stage infections in the initial year could be higher or lower depending on other assumptions. Thus, in this section we test varying the initial percentage of PHI stage infections from 0 to 50%. We limit our parameter exploration at 50% as high percentages of PHI would represent a very fast epidemic growth. For example, if 50% of the initial infections are in PHI phase, then the epidemic will have doubled within 1 year.

Figure 6.5a displays the relationship between the percentage of initial PHI infections and the cumulative number of infections which yields a Pearson correlation coefficient of 0.312 and a slope of 0.482. Thus, there is a slight positive relationship from the initial percentage of initial infections starting in the PHI stage of HIV. There appears to be a somewhat stronger relationship between the PHI percentage and the cumulative number of deaths, displayed in Figure 6.5b, which yields a Pearson correlation coefficient of 0.914 and a slope of 3.231.

6.2.4 PHI Multiplier

Our model assumes a uniform factor by which PHI stage HIV increases the risk of HIV transmission above the baseline HIV transmission risk during the asymptomatic phase. While we assume a PHI factor of 8 based upon the Spectrum software default value, earlier estimates for this factor ranged from 31 to 141 [23]. In this section we vary PHI multiplication factor from 1 to 100 to observe the impact different PHI factors have on the epidemic in Bangkok.

Figure 6.6a displays the relationship between the PHI risk factor and the cumulative number of infections which yields a Pearson correlation coefficient of 0.966 and a slope of 12.709. Thus, there is a strong positive relationship from the PHI factor which determines the increased risk of HIV transmission during the PHI stage of HIV. There is a similar strong correlation with the PHI factor and the cumulative number of deaths, displayed in Figure 6.6b, which yields a Pearson correlation coefficient of 0.955 and a slope of 6.288.
Figure 6.4: Prevalence and Mortality for Stochastic Variation Test

(a) HIV Prevalence

(b) HIV Mortality
Figure 6.5: Analysis for Percent Initial PHI Test

(a) Initial % PHI vs. Cumulative Infections

(b) Initial % PHI vs. Cumulative Deaths
Figure 6.6: Analysis of PHI Multiplier Test

(a) PHI Factor vs. Cumulative Infections

(b) PHI Factor vs. Cumulative Deaths
6.2.5 Sexual Frequency

Sexual frequency is a key factor in the spread of HIV. In particular, the sexual contacts occurring when a person is in PHI stage of a HIV infection could result in fast onward transmission of HIV due to the heightened risk of transmission during the PHI stage. In this section, we vary the mean number of annual sexual contacts desired by each agent while maintaining the same standard deviation as the original dataset standard deviation value. We vary the mean number of sexual acts from 10 to 100.

Figure 6.7a demonstrates the concurrency measure in the 180 period is roughly the same for all simulation runs in this section, varying between 1.2 and 1.6. Displayed in Figure 6.7b, the relationship between sexual frequency and the concurrency measure is roughly flat, as expected, with a Pearson correlation coefficient of 0.471 and a slope of 0.001. Although the coefficient indicates a slight correlation, we believe this is attributed to the outliers in runs with a mean of 10 sex acts per year. The concurrency measure is calculated by determining the number of sexual contacts within a 6 month window. Thus, if the number of sexual contacts is too low, then this will depress the concurrency measure.

Figure 6.8a demonstrates the effect of different values for the mean number of desired sexual acts annually on the measured number of sexual acts annually at the end year of the simulation. As expected, there is a strong linear relationship, yielding a Pearson correlation coefficient of 1.0 and a slope of 0.74. Ideally, the slope would be 1.0. However, the effects of dating appear to depress the number of observed sexual acts. This disconnect could be attributed to disparity in the condom preferences or preferred number of sexual acts of paired agents. It might be prudent to increase the number of sexual contacts by a factor to compensate for the effects of the dating model on sexual frequency.

As expected, Figure 6.8b demonstrates the strong positive correlation between the number of sexual acts and the cumulative number of infections, producing a Pearson correlation coefficient of 0.997 and a slope of 27.734. Thus, errors made in estimating or modeling the number of sexual acts can have a strong impact on the overall size of the epidemic.
Figure 6.7: Analysis of Sexual Frequency and Concurrent Relationships

(a) Overall Concurrency Measure in 180 Days

(b) Sexual Frequency vs. Concurrent Relationships
Figure 6.8: Analysis of Sexual Frequency

(a) Sexual Frequency Desired vs. Sexual Frequency Measured

(b) Sexual Frequency Desired vs. Cumulative Infections
Table 6.8: Bangkok Simulation Parameters

<table>
<thead>
<tr>
<th>Group</th>
<th>Lower Age</th>
<th>Upper Age</th>
<th># People</th>
<th># HIV</th>
</tr>
</thead>
<tbody>
<tr>
<td># 1</td>
<td>15</td>
<td>22</td>
<td>2840</td>
<td>366</td>
</tr>
<tr>
<td># 2</td>
<td>23</td>
<td>28</td>
<td>3820</td>
<td>669</td>
</tr>
<tr>
<td># 3</td>
<td>29</td>
<td>45</td>
<td>3350</td>
<td>697</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value (Sd)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Operational Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td># Years</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Probabilistic HIV Model</td>
<td>True</td>
<td></td>
</tr>
<tr>
<td>Probabilistic Dating Model</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>Symptomatic Discovery</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>Replace Removed Agents</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>ART Parameters</td>
<td>Not used</td>
<td></td>
</tr>
<tr>
<td>PrEP Parameters</td>
<td>Not used</td>
<td></td>
</tr>
<tr>
<td>Serosorting</td>
<td>Not used</td>
<td></td>
</tr>
<tr>
<td>Mixing</td>
<td>Random</td>
<td></td>
</tr>
<tr>
<td><strong>Sexual Roles</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Insertive</td>
<td>47.8</td>
<td>[78]</td>
</tr>
<tr>
<td>% Receptive</td>
<td>18.0</td>
<td>[78]</td>
</tr>
<tr>
<td>% Versatile</td>
<td>34.2</td>
<td>[78]</td>
</tr>
<tr>
<td><strong>Casual Relationships</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td># / Year</td>
<td>7.173 (10.944)</td>
<td>[88]</td>
</tr>
<tr>
<td><strong>Steady Relationships</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Duration</td>
<td>947.6 (1103)</td>
<td>[95]</td>
</tr>
<tr>
<td>Sex / Year</td>
<td>13.8 (42.948)</td>
<td>Estimate</td>
</tr>
<tr>
<td><strong>Condom Use</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Always Condom - Steady</td>
<td>46</td>
<td>[65]</td>
</tr>
<tr>
<td>% Always Condom - Casual</td>
<td>50.3</td>
<td>[65]</td>
</tr>
<tr>
<td>% Condom Use - Steady</td>
<td>68.8</td>
<td>Estimate</td>
</tr>
<tr>
<td>% Condom Use - Casual</td>
<td>0</td>
<td>Estimate</td>
</tr>
<tr>
<td><strong>Other Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% with Concurrent Partners</td>
<td>15.5</td>
<td>[79]</td>
</tr>
<tr>
<td># Concurrent Partners</td>
<td>1.181 (2.210)</td>
<td>[79]</td>
</tr>
<tr>
<td>% Circumcised</td>
<td>10.3</td>
<td>[165]</td>
</tr>
<tr>
<td>HIV Testing Rate</td>
<td>0.89</td>
<td>[116]</td>
</tr>
<tr>
<td>% Never HIV Testing</td>
<td>56.2</td>
<td>[78]</td>
</tr>
<tr>
<td>% Initial HIV w/PHI</td>
<td>31.79</td>
<td>Estimate</td>
</tr>
</tbody>
</table>
CHAPTER 7
DISCUSSION & EXPERIMENTAL RESULTS

In this chapter we discuss the design considerations in developing the MASHIV simulator as well as the implications of our application of the MASHIV simulator to the MSM epidemic of HIV in Bangkok, Thailand.

7.1 Impact of Concurrency

While concurrency has a theoretical possibility of interacting with PHI stage HIV infection to produce significant HIV spread, the impact of concurrency in actual populations is reduced by numerous factors. HIV saturation in the social networks can limit access to HIV negative partners. High HIV prevalence may additionally contribute to feedback behavior from the population such as increased condom usage, increased HIV testing, and other risk reduction strategies.

Concurrency may contribute to the spread of HIV in Bangkok, Thailand, but our simulation results suggest that it does not currently contribute substantially to the HIV spread. Figure 7.1 displays the relationship between the percent of the population having concurrent partners and the cumulative number of HIV infections after 15 years. While we did note a correlation between the degree of concurrency in the population and the cumulative number of infections, the overall change in the number of infections between 0% concurrency and 100% concurrency was less than 2% of the total population size. Our result is in agreement with Leung and Kretzschmar who concluded in 2015 that while concurrency could be a driving factor of the HIV epidemic in sub-Saharan Africa, it would not be a significant factor in MSM populations [109].

Our dataset definition used in our simulation represents a highly simplified view of a subset of the MSM population in Bangkok, Thailand. While we have attempted to choose appropriate values to represent the MSM population in Bangkok, consultation with an expert demographer is needed to more adequately define a more representative dataset to drive our simulations. Such expertise would likely be useful in guiding the development of a more accurate depiction of the Bangkok MSM epidemic in a manner that conforms to the requirements of our simulator.

Additionally, further refinement of the MASHIV agent model and user interface is needed to both more faithfully represent the parameters specified in the dataset as well as improve the usability of the interface for a variety of users. Running our dataset for Bangkok on MASHIV with a 17.3% initial HIV Prevalence did not yield the 28% HIV prevalence expected within a 2 year period. Instead, the epidemic plateaued at approximately 18% and declined. Further refinement of the Bangkok dataset and agent model is needed to more faithfully represent the conditions present in Bangkok and the dating behavior. We hope that more detailed data on the dating behavior of MSM in Bangkok will become available to aid us in this pursuit.

Further data collection is likely needed in regard to the composition of steady and casual
relationships for MSM in Bangkok, Thailand. While several surveys of this region regularly ask whether an individual has had a steady or casual relationship in the past several months [78], there is a dearth of information on the number of concurrent partnerships. That is, we may know the percentage of the population to have had a casual relationship in the past 3 months, but we do not know how many partnerships occurred in that period nor do we know how many of those existed concurrently. Attempting to overcome this lack of data, we tested the simulator across a wide range of possible combinations of the number of overall relationships, number of concurrent partnerships, and number of sexual acts.

In regard to the use of multi-agent simulation, we believe our approach demonstrates the flexibility of MAS for modeling individual-level behavior. The MASHIV simulator incorporates a novel combination of a CD4 compartment model to model HIV disease progression, individual-level agents making decisions based upon personal histories of interactions, and normal distribution hazard functions to produce daily event probabilities. The MASHIV simulator is a flexible tool incorporating numerous aspects of the complex transmission factors affecting HIV epidemics. We believe that with further development and expert demographer input, MASHIV could be a useful tool added to the arsenal of HIV modeling.
7.2 Implications of Hazard Models

We use daily probabilities for events based on a hazard function to influence the distribution of events, meaning control is local, not central. The normal hazard function ensures that approximately 50% of events will occur before the mean with an increasing daily probability that will result in the typical normal bell shape past the mean. While this approach yields results that bear similarity to a centralized assignment of events, since we maintain the decision at the agent-level then we can further extend this model to involve agent-specific influences on the probability distribution. That is, we can use the agent’s personal history to change the likelihood of events occurring. For example, if we used a hazard function to model relationship length, we might decrease the hazard of a relationship ending proportionally to relationship duration and increase the probability in the case of infidelity. The constructions of such a function may involve the use of a Weibull hazard function incorporating multiple phases. Weibull hazard functions have been previously applied to generate the typical bathtub-shaped mortality pattern [174]. This model incorporates three distinct contributions to mortality—heightened mortality during infancy, slowly increasing mortality with age, and rapidly increasing mortality near the average lifespan. In regard to modeling HIV, one could build a Weibull function that models sexual activity with age.

7.3 Normal Hazard Computation Load

The normal hazard function, defined in Equation 4.12, requires the calculation of an integral. Integration may be approximated by the use of adaptive integration, for example, or by other means. Still, integration is somewhat computationally intensive and thus reducing unnecessary integration is desirable. For our CD4 compartment model, which uses a normal hazard function to calculate daily probability of transition, the values from the hazard function are reused for every HIV infection. To avoid recalculating the same probabilities repeatedly, we pre-calculate and store the vast majority of daily probabilities we anticipate needing. As discussed previously, this pre-calculation results in a 73.9% reduction in calculation time.

However, it is also possible to eliminate the integration altogether by considering the use of an alternative distribution for our hazard function. A Weibull distribution can be used to approximate a normal distribution and has the added benefit of a much simpler hazard function. The hazard function, defined in Equation 7.1, requires we define a location parameter ($\delta$), shape parameter ($\beta$), and a scale parameter ($\eta$) [147]. The location parameter, $\delta$, represents an x-offset. As for the shape parameter, Rinne concluded that a shape parameter of $\beta = 3.60235$ produces a satisfactory approximation of normal probabilities on the lower tail of a normal distribution [141]. As the majority of daily probabilities we expect to use are on the lower tail, this is a desirable quality.

To match a normal distribution with standard deviation $\sigma_N$ and mean $\mu_N$, we must adjust the Weibull scale parameter $\eta$ and x-offset $\delta$ respectively. Note, however, that changes in the
scale parameter alter the mean of the Weibull distribution. Hence, we suggest correcting for the standard deviation first, and then the mean. First, using Equation 7.2 [147] we can estimate \( \eta \) by assuming our normal distribution sigma \( \sigma_N \) will equal \( \sigma_W \) and calculating \( \eta \) through basic algebraic manipulation.

Once we have established the appropriate standard deviation, we can correct for the difference between the Normal distribution mean (\( \mu_N \)) and the Weibull distribution mean (\( \mu_W \)). For the normal distribution, the mean, median, and mode are all equal. However, in our approximation using the Weibull distribution that relationship does not hold true. Experimentally, we have found that producing a \( \delta \) from the difference of the medians or modes of the two distributions produces satisfactory results. Hence, we calculate the median (\( \tilde{\mu}_W \)) of the Weibull function using Equation 7.3 [141] and subtract it from the normal median \( \tilde{\mu}_N \) to calculate \( \delta \) in Equation 7.4.

\[
h(t) = \left( \frac{\beta}{\eta} \right) \left( \frac{t - \delta}{\eta} \right)^{(\beta - 1)}
\]

\[
\sigma_W = \eta \sqrt{\Gamma \left( 1 + \frac{2}{\beta} \right) - \Gamma \left( 1 + \frac{1}{\beta} \right)^2}
\approx 0.27787\eta \text{ for } \beta = 3.60235
\]

\[
\tilde{\mu}_W = \eta \Gamma \left( \ln(2) \right)^{1/\beta}
\]

\[
\delta = \tilde{\mu}_N - \tilde{\mu}_W
\]

To test the Weibull approximation, we approximate a normal hazard distribution with \( \mu = 1 \) and \( \sigma = 0.6532 \). In Figure 7.2a, we can observe the percent of the trials that have experienced the event at different times. The two approaches show nearly identical transition rates with one noteworthy difference. In the Weibull hazard approach, the entire population transitions slightly sooner than the normal hazard approach. This can be explained by looking at the daily probabilities in Figure 7.2b. As Figure 7.2b demonstrates, the two approaches show great similarity on the lower tail below the mean of 1 year, but slowly diverge past the mean.

While our approach demonstrates how to approximate normal distributions with Weibull distributions, there is clearly room for further refinement. Still, our example demonstrates the viability of the approach. This Weibull approximation would simplify the hazard function, eliminating the need for integration. Additionally, while the normal hazard function becomes undefined when the CDF approaches 1, the Weibull hazard function avoids this problem.
Figure 7.2: Comparison of Normal-Hazard and Weibull-Hazard Approaches
7.4 Normal Hazard Standard Deviations

For our normal hazard function that we use to define daily probabilities of CD4 compartment transition, we chose to use a standard deviation that is a constant multiple of the mean compartment duration. Through experimentation, we selected a constant of 0.6532. However, it is possible that we may obtain a closer match to the Spectrum CD4 distribution pattern via other methods of standard deviation estimation. For example, we could have used a separate sigma multiplier for each compartment, or alternatively, a standard deviation that does not depend on the mean. Additionally, if we research the source literature for the transition durations specified in the Spectrum software, which we use to initialize our MASHIV model of HIV, it is possible we may find the standard deviations or confidence intervals for these mean durations in the original studies.

7.5 Object-oriented Issues

Although object-oriented design allows for a very natural way of organizing concepts such as persons or infections into data objects, the use of such complex objects in lieu of primitives adds additional computational burden. For example, in JAVA, one can create an array of “person” objects, where each person will be comprised of numerous variables. However, iterating through these person objects is nowhere near as efficient as iterating through a primitive array of integers or doubles. Primitive data types such as integers or doubles are organized in close proximity during storage as to allow for efficient access. During a profiling experiment, we discovered that our custom object arrays, such as a person object array, were actually stored in hashmap data structures. While this allowed for array-like access, a hashmap is much less efficient to iterate over than an array of primitive data types.

As such, our model would run considerably faster if we converted the data storage from arrays of complex objects to arrays of primitives. However, such an adaptation would likely result in very large arrays of primitive data types representing all the variables we would have previously stored in custom objects. Operating on such large arrays would require careful array indexing to ensure we are accessing and updating the correct values. Additionally, these monolithic arrays would be more challenging to design, maintain, and extend to accommodate additional variables. Thus, such an adaptation would likely sacrifice ease of use for an increase in computational efficiency. For an established model that is unlikely to undergo significant changes in the variable locations, such a conversion may be well worth the effort. However, models with more dynamic development may encounter challenges in keeping array indexing correct.

Barrett et al. developed a high-performance computing framework, EpiSimdemics, capable of scaling to 100 million node social networks [21] This framework was designed to explore pharmaceutical and non-pharmaceutical interventions on a infectious disease spreading through realistic social networks comprised of large numbers of individuals. As HIV is not spread in the same man-
ner as highly-infectious contact-driven diseases, Barret et al.'s model is not directly applicable to modeling HIV dynamics. However, the authors have demonstrated that it is possible to simulate individual-level population dynamics on exceptionally large social networks.

7.6 Probability Calculations

Since each HIV infection experiences the same daily probability of CD4 compartment transitions and mortality, at program initiation, we pre-calculate the bulk of the daily probabilities we expect to be used and store them for future use. This saves considerable calculation time that would have been used for recalculating the same daily probabilities. However, additional calculation time could be further saved by predetermining CD4 compartment transition times instead of performing daily random samples to determine whether or not a transition occurs.

7.7 Fitting Prevalence with Agents

There are numerous solutions that attempt to fit parameter-based epidemic models to prevalence values. However, there are very few multi-agent models that attempt such a feat. One major reason for the dearth of multi-agent fitters is that these models tend to be much slower in comparison to parametric models. While we had initially aspired to build such a fitting engine, when we observed the average calculation times of our model, we abandoned this endeavor. Currently, for a population of size 20,000 with 20% HIV prevalence, our model takes approximately 2 minutes to run a 15-year simulation. A fitting process, such as IMIS, requires many thousands of such runs to establish the parameters which result in the best fit to the data. Although we implemented a multi-threaded mode, which assigns a thread to each run, the long simulation time would result in a fitting time on the order of days or weeks. Assuming no increase in fit time with prolonged use, running 50,000 fits across 8 threads with an average simulation time of 2 minutes would result in a fit time of 208 hours or roughly 9 days. As such, unless we can substantially reduce the simulation time, such as possibly through conversion of the data structures from objects to primitive arrays, the use of such a fitting engine would be a painfully slow process.

7.8 PHI & Contact Rates

The impact that PHI can have on an epidemic is strongly dependent on the sexual contact rate of the population. The contact rate of a population is a product of the relationship turnover and the sexual rates within relationships. Thus, the impact of PHI is not solely dependent on how much sex is occurring, but also upon how many different partners this sex involves. A person in a long-term monogamous relationship who contracts HIV from his partner is unlikely to transmit HIV during the PHI period due to the relatively short duration of PHI regardless of the sexual rate. Similarly,
a person with many concurrent partners but a very low personal sexual contact rate would have a low likelihood of transmitted HIV during the PHI period, again due to the short duration of PHI. However, if a person has a high relationship turnover and a high sexual rate, then his unique sexual contact rate is high and he would be more likely to transmit HIV during the PHI period.
The MASHIV simulator has great potential for becoming a useful HIV modeling tool for epidemiologists guiding policy makers. However, the data requirements, agent model, and user interface must be further refined to meet the needs and expectations of epidemiologists modeling HIV. Further consultation is needed with HIV modelers to ensure the MASHIV software is user-friendly, dependable, and provides the analytical tools necessary to both validate and analyze the model results. In this chapter we highlight several avenues of future work we noticed which may be promising, but are beyond the scope of our current work.

8.1 Agent Model Refinement

Our examination of the Bangkok dataset highlighted several areas for future refinement of the agent model. For example, allowing the user to specify a distribution for the initial HIV infections stages would allow for a more diverse and realistic HIV risk exposure. Also, the preferred number of sexual contacts did not match the measure number of sexual contacts suggesting that we need to compensate for the effects of dating which can reduce the actual number of sexual acts. Further analysis is needed to highlight any other limitations in our current agent model. Subsequently, we must continue to refine the agent model to ensure it more faithfully represents the user’s dataset.

8.2 Complex Agent Expansion

While our agents are primarily reactive agents, responding in a fairly rote manner to stimuli, the simulator can be extended to allow for more complex decision making concepts. For example, one could build a cognitive model which adapts the agent’s choices for relationship formation, relationship dissolution, dynamic partner mixing preferences, sexual act timing, reactions to infidelity, risk-assumption with partners, and risk aversion behavior. However, as fascinating as a cognitive model may be, the inclusion of such a model increases the burden of validation. A cognitive model increases the number of possible agent state configurations which in turn requires more extensive stochastic testing to ensure the behavior of the system is well understood.

However, one step we can take towards more complex behavior is to introduce time-variant preferences for the agents’ sexual and dating behavior. Our model currently treats sexual and dating preferences as being statically defined, which is unrealistic. Sexual behavior and dating preferences are dynamic and vary through different phases of a person’s life. As such, it is important to introduce such age-variant behavior into our model in the future.
8.3 PrEP in Thailand

In Chiang Mai City, in Northern Thailand, a trial of oral chemoprophylaxis administered to 551 MSM and transgender (TG) women yielded a 44% reduction in HIV incidence and increased interest in PrEP as a HIV mitigation strategy in Thailand [44]. However, Van Griensven et al. expressed concern that daily PrEP dosing may not be compatible with the sexual practices of MSM in Bangkok, Thailand [77]. The authors argue that intermittent PrEP dosing may be sufficient given the long intracellular half-life of TDF. Two-thirds of participants in the author’s study had a window of opportunity to allow an effective dose of PrEP prior to sexual activity. However, Van Griensven et al. noted that those who did not have such a window period were also those at higher behavioral risk of HIV infection. Other arguments stated for preference toward an intermittent PrEP dosing regimen are reduced costs, decreased pill burden, reduced toxicity, and diminished side-effects.

In 2005, responding to concern regarding the 30 to 50% HIV prevalence in Bangkok’s IDU population, the US CDC sponsored the Bangkok Tenofovir Study to examine the effect of tenofovir-based PrEP on HIV transmission among 2413 IDUs [119]. In 2013, Choopanya et al. reported that the trial had demonstrated a 48.9% reduction in HIV incidence among participants [45]. Restricted to those with detectable concentrations of tenofovir in the study, the PrEP demonstrated a 74% efficacy. The authors observed an 83.8% adherence to PrEP during the study.

It would be an interesting experiment to apply our MASHIV simulator to the problem of PrEP dosing and efficacy in the Bangkok region of Thailand. Although we have already implemented support for daily PrEP dosing, we do not currently support strategic dosing as described by Van Griensven et al. above.

8.4 Partnership Duration

Contradicting Kretzschmar’s model which demonstrated a decreased role of PHI as the duration of partnerships increased, Kim observed in his individual-based model that while there was an initial reduction, the role of PHI increased again as partnership duration further increased [94]. It would be interesting to determine whether our model produces a pattern of PHI effect that is more similar to Kretzschmar’s or Kim’s model. To accomplish this, we would need to study the effects of increasing the mean duration of steady partnerships for each simulation. We would observe the annual proportion of PHI caused infections for each simulation, and calculate the correlation between mean duration and proportion of PHI across all simulations. However, the proportion of HIV infections due to PHI may not be the best metric for analyzing the contribution of PHI as ART can reduce the number of asymptomatic infections resulting in a higher PHI proportion relative to the other stages of infection.
CHAPTER 9
SUMMARY

The purpose of this dissertation was to develop a HIV modeling software that would be useful in illuminating the role that concurrent sexual relationships play in the transmission of HIV. Concurrent relationships, occurring when a person has multiple active overlapping sexual relationships, have been attributed as contributing to the spread, severity, and pervasiveness of HIV epidemics [101].

Previous efforts by researchers to examine HIV epidemics have often treated both HIV disease progression and transmission in highly simplified manners that consider populations in aggregate. While researchers have developed individual-level simulations of HIV transmission to examine concurrency [140, 10, 102], few have applied this type of modeling to MSM populations [121, 193, 26]. MSM populations do not have the inherent sexual role segregation of heterosexual populations shown to limit HIV spread [26]. Additionally, MSM are subject to higher rates of HIV transmission due to the higher rates of transmission during UAI. Current research into simulating the effects of concurrent relationships often treat relationship formation and dissolution by simplified means such as regular durations or fixed daily probabilities. Similarly, HIV disease progression is often treated as a fixed duration toward mortality which negates variability in survival.

We have attempted to fill this gap in the research by the development of an individual-level MAS for HIV epidemics of MSM, dubbed MASHIV. MASHIV is a individual-level population simulator incorporating complex HIV transmission and disease progression. In a simulation, agents interact to form sexual relationships according to either duration-based or time-variant probabilities. During acts of sex, HIV may be transmitted between the agents according to the risk of transmission for the HIV positive partner and risk of reception for the HIV negative partner. The probability of transmission is variable based upon the infected agent’s stage of HIV infection, circumcision status, and ART use. Additional mitigating factors include the use of condoms and the HIV negative partner’s use of PrEP. HIV progression toward mortality is governed by a CD4 compartment model, adapted for use in a individual-level daily timestep MAS, that was first introduced in the Spectrum HIV modeling suite [62] for modeling aggregate populations.

We applied MASHIV to a simplified representation of the MSM epidemic in Bangkok, Thailand to investigate whether concurrency may be a driving factor of the HIV epidemic of the Bangkok MSM community. Our test results suggested that, while concurrency does have a theoretical possibility of contributing significantly to HIV spread under certain conditions, concurrency does not currently represent a substantial contribution to the spread of HIV in Bangkok despite the positive correlation measured. However, given that our model under-represents the expected growth of the HIV epidemic in Bangkok compared to observed values in literature, further refinement of the MASHIV model and Bangkok dataset is needed to further validate this result.
BIBLIOGRAPHY


[69] Steven M. Goodreau, L. Pedro Goicochea, and Jorge Sanchez. “Sexual role and transmission of HIV Type 1 among men who have sex with men, in Peru”. In: Journal of Infectious Diseases 191 (Supplement 1 2005), pp. 147–158.


Maria J. Wawer et al. “Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda”. In: Journal of Infectious Diseases 191.9 (2005), pp. 1403–1409.


David P. Wilson et al. “Serosorting may increase the risk of HIV acquisition among men who have sex with men”. In: Sexually Transmitted Diseases 37.1 (2010), pp. 13–17.


