Using Network-based Modeling to Implement Strategies for Reducing HIV Drug Resistance

Juandalyn Leaunda Coffen Burke

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2022

Reading Committee:
Peter Tarczy-Hornoch, Chair
John Mittler
Geoffrey Gottlieb

Program Authorized to Offer Degree:
Biomedical Informatics and Medical Education

University of Washington

Abstract

Using Network-based Modeling to Implement Strategies for Reducing HIV Drug Resistance

Juandalyn Leaunda Coffen Burke

Chair of the Supervisory Committee:
Peter Tarczy-Hornoch
Biomedical Informatics and Medical Education

Mathematical models have evolved to capture epidemiological and infectious disease data and complex relationships between the disease, host, and environment. Infectious disease and epidemic models serve as ethical approaches for evaluating strategies intended to prevent, control, and reduce disease progression, transmission, and exposure, especially when the risk of harm and cost are high. These models also allow researchers, clinicians, and other health professionals to explore strategies and interventions that have yet to be widely explored, either due to limited information or resources. This thesis uses a network-based, stochastic modeling system called *EvoNet to* capture the spread of the human immunodeficiency virus (HIV) and explore strategies that may reduce HIV drug resistance (HIV-DR), within-host and between-host. The first aim focuses on predicting the frequency and emergence of drug-resistant mutations, given the use of first-line antiretroviral treatment, patient-specific information such as drug adherence levels, and the incorporation of pharmacogenomics (PGx) and pharmacokinetic (PK) data. The second aim investigates the effect of host genetic variation and drug resistance mutations in the population, using PGx and PK study data from two sub-Saharan African populations. Lastly, in the third aim, we build a treatment-switching optimization routine and develop a method similar to the grid

search optimization method and use an R package that performs simulated annealing. The objective of using these two optimization methods is to find the best parameter values that reduce the levels of drug resistance in the simulated population and to also inform how health professionals may prioritize individuals for second-line treatment. The results of these aims include the following: (1) the inclusion of host genetic or PGx data influences the frequency of drug resistance mutations, within-host and how rapidly the drug resistance mutations emerge, (2) the presence of individuals with drug-resistant strains in the population at the start of the model simulation yields higher levels of predicted drug resistance and in particular, transmitted drug resistance, and (3) the adapted grid-search optimization approach had a higher computational time burden than simulated annealing but provided a wider range of options for group prioritization for second-line treatment conditions that dramatically reduced HIVDR in the simulated population. In all, these methodologies and results may extend to future investigations of new drugs and treatment regimens.

Acknowledgements

Firstly, all praises be to God (Allah), without whom this accomplishment would not be possible.

I have dreamt of receiving my Ph.D. since I was 8 years old, sitting in a bookfair, reading through all the books about space, the universe and Dr. Mae Jemison. After that day, I used up all of my time and energy to consume any source of knowledge related to science, technology, engineering, and math, and made the plan to pursue a Ph.D. I imagined obtaining a Ph.D. would be a wonderful experience where I could practice the philosophy of paideia or deep learning and education because I loved school. Little did I know, the Ph.D. would serve as a greater challenge than what I assumed. However, in my pursuit of this terminal degree, I have had amazing experiences and met wonderful teachers and mentors whose ability to "see me" and key in on my thirst for learning and research has allowed me to challenge myself and explore areas of interest that I may have never accessed or examined.

I would like to thank all my advisors for their guidance and support throughout my years as a researcher and scientist, and especially my doctoral committee members at the University of Washington. I am honored to have had the opportunity to collaborate with you all. I would like to give a special thanks to my dissertation committee chair: Dr. Peter Tarczy-Hornoch, for being such an encouraging and motivating mentor. You were always willing to explain or provide detailed guidance, especially at the later stages of my dissertation. You have championed me from the beginning of my time in the BIME program and somehow in your busy schedule, you were available to meet with me regarding any aspect of my dissertation. Dr. Ruanne Barnabas, I cannot thank you enough for your all your support. Your ability to be present and demonstrate excellence as a professor, scholar, and researcher, has been an inspiration to me. Dr. Geoff Gottlieb, thank you for being such a positive mentor, always giving me constructive feedback and encouragement as I completed my research. I enjoyed our conservations related to infectious disease and HIV, where I absorbed a vast amount of information related to your experiences and expertise in both clinical and public health settings. Dr. John Mittler, thank you for being such an excellent mentor.

You have been instrumental in my growth in academia, as a research scientist, biomedical data scientist and modeler. Your knowledgebase about a variety of scientific subjects is so vast and incredibly detailed. I also cannot say and express enough, how much each whiteboard and parallel programming session meant to me.

I would like to thank the National Library of Medicine (NLM) Pre-doctoral Fellowship program and the UW Institute of Translational Health Sciences (ITHS) TL1 Pre-doctoral training program for providing funding support during the early stages of my Ph.D. In addition, I am thankful for the Graduate Student Equity and Excellence (GSEE/GOMAP) program, specifically Dr. Carolyn Jackson and Kayla Newcomer, for making sure that I had the adequate funding during the final stages of my Ph.D.

Next, I would like to thank the research and academic groups I have been involved with at UW. Thank you to the EvoNet group and team members – Dr. Steve Goodreau, Dr. Josh Herbeck, Dr. Sarah Stansfield, Dr. Kathryn Peebles, Molly Reid, and James Murphy for your support at every team meeting and presentation. Specifically, I would like to thank, Dr. Neil Abernethy, for your advice over the years and funding support for equipment needed during the middle stages of my research journey.

I would like to extend my thanks to my URP family. Dr. Jennifer Harris, Dr. Jessica Salvador, Dr. Sophie Pierszalowski, Qanani, Daisy, Pavi, Dan, Grecia, Jeamel, Victoria, Trisha, Sam, Binh and other UAA staff, thank you for being my safe space and providing a welcoming environment for me to go to on the UW campus. I would also like to express my gratitude to Dr. Sharona Gordon and Dr. Susan Ferguson. Thank you for believing me. Full stop. You both ensured that I was well-supported as I sought to complete my degree. You are incredible mentors and women in science that I feel thankful to have connected with during this journey.

Next, I would like to thank the students (past and present) and staff of the Biomedical Informatics and Medical Education department at UW. I am thankful for all the BIME students I have had the pleasure of getting to know including Dr. Dae Lee, Dr. Shefali Haldar, Dr. Ryan James, Dr. Lucy Wang, Dr. Yong Choi, Dr. Nikhil Gopal, Dr. Uba Backjona, Dr. Katie O'leary, Dr. Ross Lordon, Dr. Laura Kneale, Dr. Jimmy

Phuong, Will Kearns, Harkirat Sohi and other BIME colleagues and staff. Thank you all for always reaching out to connect and be supportive at each milestone of my Ph.D. I am happy to have met all of you.

I would like to send a special thanks to my friends and colleagues in the Get PhinishD, GSEE Writing group – Dr. Antonio Olivas-Martinez, Dr. Arianna Gomez, Dr. Elizabeth Ramirez Arreola, Dr. Maria Blancas, and Dr. Natalie Gasca. Thank you for providing me with so much guidance and support, including early mornings and late nights working sessions online, providing feedback on my work and messages of positive affirmations (even through a pandemic) to persist on. You all are such awesome people, and I am so grateful that I was able to truly build friendships and a community with all of you while in Seattle and beyond. I would also like to thank my high school, college and graduate school friends: Angel, thank you for being such a highly energized and loving friend; Fiana, thank you for always wanting to go on adventures and talk about everything scholarly and academic; Wutyi, thank you for being another voice of reason for me and sharing lots of laughter, always; Dr. Silpa Taunk, my A-squared roomie, thank you for your ability to not only check-in and connect with me, but also for your kindness, empathy and support, especially during challenging times; to my UMich, A-squared crew (Angela, Ginger, Christine, Melanie, Claudia, and all), thank you all for keeping our friend group going and creating authentic experiences for us to share; and to my Spelman sisters, I am so grateful that we met and continue to uplift one another.

Finally, I would like to thank my family. I am incredibly thankful and grateful to my patient, kind and loving husband, Caleb Burke. You are my best friend, colleague, and mentor. You have helped me at each stage of this process: practice, ask questions, and challenge my knowledge and for that, I am so appreciative. Also thank you for all your IT support and your ability to be a listening ear and soundboard when I need advice the most. You have been an amazing partner! I am also grateful for my mother-in-love, Angela Burke and sister-in-love, Savannah. Thank you both for always sending encouraging words and messages of support. You both rock.

Nadia and Laila, thank you both for being the best daughters with such incredible minds and loving spirits. I look up to the both of you and you inspire me beyond

measure. You share and bring me so much joy and are truly the brightest lights and stars. You give me the best advice! I am so grateful for the two of you and hope that my journey allows you to see that anything is possible, as you have taught me that as well.

Natashia, thank you for being the best sister, for always understanding my point of view, and being such a vibrant and genuine spirit. You always know what to say and do to keep me focused and make me laugh. You have always demonstrated excellence, showing me how to both conduct science and become a researcher, from an early age and into adulthood. Luqman, thank you for always providing a listening ear and visiting me when I needed it the most. You have been a voice of reason for me and for that I am grateful. Your artwork has also brought me joy as well during the difficult parts of my journey. Ismael, thank you for always encouraging me to see the positive side of any situation during my Ph.D. journey and in general. Your words have been so genuine and kind, giving me hope to move forward, no matter the situation.

Lastly, I feel so blessed to have such beautiful, thoughtful, caring, and loving parents. Ma and Daddy, your sacrifices, and prayers both kept me going, and gave me the courage to always reach for excellence and persevere, even in the darkest hours. You both have always advocated for me, Natashia, Luqman, and Ismael. You never missed a game, performance, exhibit/art gallery showing, science fair, or opportunity to help us simply grow our interest. Ma, you are my rock and an undeniable source of strength, patience, and love. You have taught me that faith and kindness will be my greatest assets. Daddy, may you rest in peace and in love - I finished and right before your birthday! Thank you for showing me how to take action, have courage, and be bold. You have given me so many gifts and so much wisdom that I take with me daily and have been instrumental in providing me with the mental strength and health to not only complete my Ph.D. but thrive beyond this milestone in my life. And I am forever grateful. Ma and Daddy, I love you both with all my heart and soul.

I also want to thank my grandparents – Julia Elijah, Arthur Lee Elijah, Jocelyn Lewis, Willie Lewis, Elson Coffen Sr., and great-grandparents – including my great-grandmas, Annie-Claude, and Big Mama Bessie, and beyond. I took all of you with me.

I am my ancestor's wildest dreams, and I am merely walking on the path that they have laid out for me. I am their accomplishment! - Unknown

Dedication

I dedicate my dissertation work to my daughters, my bright lights, and mentors – Nadia and Laila; my loving husband and partner – Caleb; my siblings, and best friends - Natashia, Luqman, and Ismael; and to the best parents I could have ever imagined, Bessie and Elson Coffen.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

The purpose of this chapter is to: (1) introduce mathematical modeling and its application to investigating infectious disease, (2) discuss antimicrobial and HIV drug resistance, (3) review the effect of antiretroviral medication on HIV replication, (4) discuss several resources and tools used to surveil HIV and HIV drug resistance, and (5) provide the main goal of this thesis work and an overview of the remaining chapters that describe this research project. The overarching question of this dissertation is "What strategies and conditions may assist in predicting, reducing, and preventing HIV transmission and HIV drug resistance within-host and between-host (i.e., in a multi-level system)?" explored through mathematical modeling. The remaining chapters of this thesis include: a review of the global impact of HIV drug resistance and epidemic modelling (Chapter 2), the development of a within-host model and a brief introduction to pharmacogenomics (Aim 1, Chapter 3), the extension of the within-host model framework for developing a between-host model and the introduction of types of HIV drug resistance (Aim 2, Chapter 4), the creation of a optimization routine to determine treatment prioritization of specialized groups to optimized a desired HIV epidemic outcome (Aim 3, Chapter 5), and the summary of research findings for thesis aims 1, 2, and 3 (Chapter 6).

1.1 An Introduction to Mathematical Models

Mathematical modeling is a process that utilizes math to represent, analyze, and predict real, observational information or phenomena[1]. In practice, mathematical modeling theorizes and simulates an event to arrive at a decision or actuate a new solution to a problem. For example, in economics, mathematical models may evaluate impact, cost and savings, and the effect of laws and policies on society[2,3] In engineering, mathematical modeling has been used, for instance, to optimize the aerodynamic performance of an aircraft or improve upon airplane size and wingspan[4]. In the natural

and life sciences, particularly environmental science, mathematical modeling has been used to identify solutions for the bioremediation of toxic sites and waterways by providing models that predict exact biochemical measurements needed to create clean water sources[5]. Mathematical modeling is endless and highly applicable to many fields, including epidemiology and infectious disease[6–8].

The following explains the basic process of constructing a mathematical model, similar to carrying out the scientific method[9,10]. In the first stage of building a mathematical model (i.e., the what and why), there is an observation about an object, device, event, or condition. This observation helps form the research question or motivation for building the model. The second stage (i.e., data collection and hypothesis) involves creating a theory and assumptions, given all known information about the observation. During the third stage (i.e., how), mathematical equations are selected to represent the relationship between model variables. These mathematical equations serve as a basis for forecasting or making predictions about the model outcomes. The fourth stage (i.e., make the prediction) involves simulating the model using the mathematical equation(s) and based on model assumptions. The fifth stage (i.e., data analysis and validation) evaluates the accuracy and consistency of the model in comparison to real-world observation(s). Finally, the six stage (i.e., report and conclusion) confirms or draws new findings of the initial hypothesis, given the use of the model, and determines how to best improve on the model.

The advantages of using a mathematical model include optimizing limited resources such as time, money, and safety, using numerical data and information. Particularly, mathematical models may serve as a more accessible approach for including parameters and resources that may not be easily attainable and may assist in the evaluation of strategies for prevention, control, and interventions without causing physical harm and death on groups and populations of individuals[7,11]. The advantages of mathematical modeling are illustrated in the modeling of infectious agents for diseases such as the influenza virus, malaria, Ebola, the 2019 coronavirus (SARS-CoV-2) and the human immunodeficiency virus (HIV)[12].

1.2 Mathematical Models and Infectious Disease

Infectious disease models require substantial empirical data about the identification and evolution of an infectious agent and an understanding of how the infectious agent interacts with the host. These models typically incorporate network-based parameters that simulate the transmission of the virus between hosts[13,14]. I discuss the model framework for the network-based simulation of epidemiologic data and host interactions in more detail in later chapters of this dissertation. By way of context however, some studies have highlighted how network-based simulations have impacted the evolution of epidemiologic modeling, especially with modeling the rates of change between susceptible, infected, and recovered states for individuals. In the past, within-host modeling or simulations that represent the movement and transition states between the infected agent and biological or chemical entities have been explored separately from the field of population-based, epidemiologic modeling[15]. In this thesis, we build on this work to apply a mathematical modeling framework that includes both within-host and population-based parameters that form a dynamic, multiscale modeling system that predicts infectious disease outcomes such as HIV transmission, specifically HIV drug resistance. In the next section, we will briefly describe antimicrobial resistance, the impact of HIV drug resistance, and the use of HIV medications.

1.3 Antimicrobial Resistance

Researchers have documented early civilizations such as the ancient Egyptians (i.e., Sudanese Nubia) may have evidence of one of the earliest accounts of the administration of antimicrobial agents in the human body[16–19]. Through the consumption of drinking a beverage (i.e., a beer-like substance), Nelson et al. 2010 found tetracycline, an antibiotic that fights bacterial infections, in the bones of an ancient population[16–19]. In addition, in 1928, scientist Alexander Fleming discovered antimicrobial medication (i.e., penicillin) to kill, reduce, and stop the growth of micro-organisms such as bacteria, fungi, parasites, and viruses[20,21]. The literal meaning of antimicrobial is derived from Greek terminology with “anti,” meaning against, “mikros,” meaning little, and bios, meaning life[22,23]. Broadly, the impact of antimicrobials has been tremendous and far-reaching.

Health professionals and healthcare agencies are now able to treat, control, and prevent certain infectious diseases and improve the overall survival rate of fatal conditions for human and animal life. Illnesses such as pneumonia and influenza that may be considered fatal for certain vulnerable populations, for instance, can currently be treated using either antibacterial or antiviral medication, respectively. In addition, due to the significant strides in improving the health of the general population, chronic infectious diseases that once spread rapidly through entire communities such as the human immunodeficiency virus (HIV) are currently treated using antiretroviral medication which significantly reduced the spread of the virus to at risk populations.

After the dawn of the antimicrobial era, antimicrobial resistance became a new issue for the field of medicine [24–26] Antimicrobial resistance occurs due to mistakes and changes during the replication of DNA or RNA sequences within microorganisms also called mutations. Resistance can also occur through other processes like recombination (i.e., the rearrangement of a DNA molecule, or the process of restructuring a section of the genome), transposons (i.e., repetitive DNA sequences that relocate to other parts of the genome), and plasmids (i.e., where antibiotic resistance genes can be found). In relation to antimicrobial resistance due to the errors made during replication of DNA or RNA, when the number of mutant strains in the viral population outcompete the non-mutant viral strains, antimicrobial medications may become ineffective, and the mutant strain confers an advantage (i.e., selective advantage) that allows the mutant strain to survive and replicate. In addition, although antimicrobial medication aids in stopping and/or killing the infectious agent, antimicrobials may also kill "beneficial" microorganisms that help the body fight off infection, cause infections that were once treated or cured to return. In the case of antiviral medication, mistakes (i.e., mutations) made during the replication process in the presence of antiviral medication, allow for the survival of the virus and the replication of mutant strains, especially if a person is not taking their medication as prescribed which may lead to non-suppression of the virus over time [23,25,27]. Otherwise, the organism is less fit, and the antimicrobial medications work as prescribed. Health professionals and institutions began to realize that if antimicrobial resistance were not reduced and prevented then effectively treated conditions would return to levels of infectivity prior to when antimicrobial medication

was available and result in higher rates of morbidity and mortality. Though antimicrobial resistance has been seen in bacteria, fungi, parasites, and viruses the focus of this dissertation will be on resistance to antiviral agents, specifically to HIV.

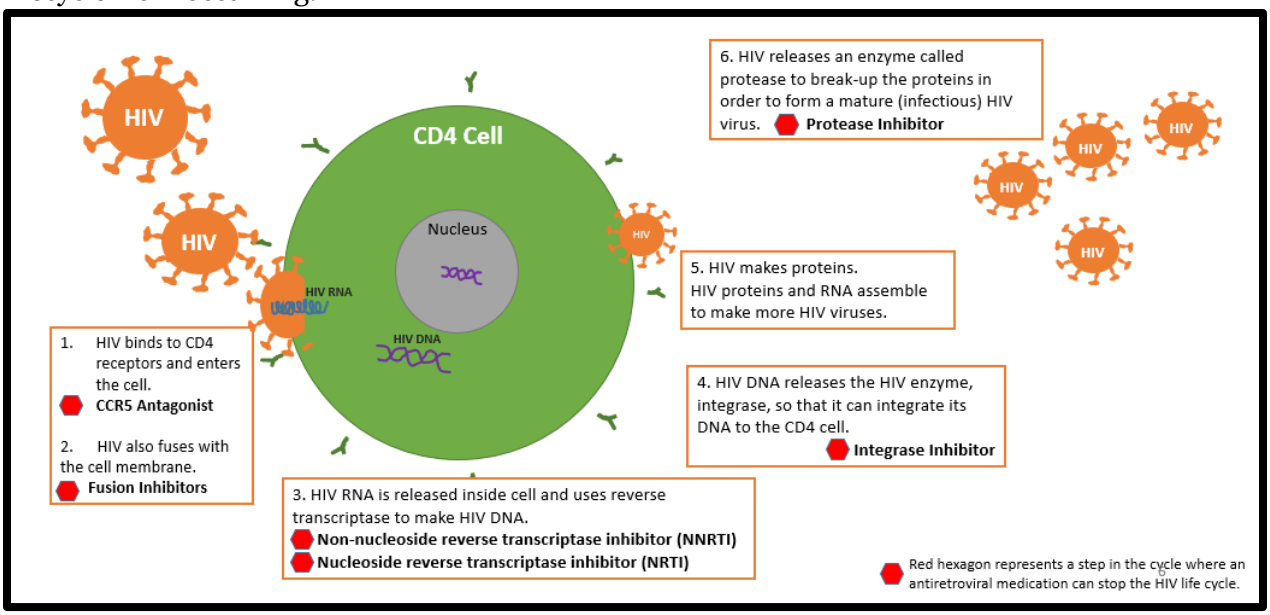
The field of HIV/AIDS has been greatly impacted by the evolution of antiretroviral drug resistance also known as HIV drug resistance (HIV-DR). HIV drug resistance occurs when mutations of the virus occur in the presence of antiretroviral medication. Those strains with mutations that impede the antiviral treatment from working are selected for an increase in frequency in the host. When HIV-DR occurs the current HIV treatment used may no longer be effective in suppressing the reproduction of HIV to non-detectable levels. In 2019, there were 37.7 million people living with HIV and 1.7 million people newly infected with HIV[28]. These numbers reflect the past and present strategies to reduce the spread of HIV which include HIV testing, access to condoms, the availability of sterile syringes for needle exchange for persons who inject drugs (PWID), programs tailored towards HIV education for vulnerable populations and those experiencing substance abuse, testing for sexually transmitted infections, and the increased use of antiretroviral treatment, including pre-exposure prophylaxis (PrEP) and treatment as prevention (TASP) [23–25,29–32]. Each strategy involved extensive research, time for planning and implementation, funding and healthcare resources spanning over four decades. With the rise in HIV drug resistance, the some of the strategies used to reduce and work towards stopping the spread of HIV will be jeopardized such as other forms of health promotion and education that reduce HIV stigma and increase adherence to antiretrovirals including PrEP and TASP campaigns. Understanding the HIV lifecycle and the use of antiretroviral medication provides a foundation for the modeling work described later. In addition, the treatment of HIV has evolved over time. Thus, the work of this dissertation focuses on WHO clinical recommendations for treatment during 2016[33–35]. We will note that changes and evolution of antiretroviral medication through this dissertation.

1.4 The HIV Lifecycle and the Use of Antiretroviral Medication

The HIV Lifecycle. Antiretroviral medication is used to specifically inhibit certain stages of the HIV lifecycle (see Figure 1). In the first of six stages of the HIV lifecycle, the CCR5 antagonist drugs are used to block the reception of HIV to binding to CCR5 co-receptors on the CD4 cell. Attachment inhibitors bind to a protein called gp120 located on the outside exterior of HIV and prevent HIV from binding and entering the CD4 cell. At this stage, post-attachment inhibitors may block HIV from attaching to CD4 receptors. At the second stage, fusion inhibitors (FIs) stop the virus from merging with the cell membrane. Next, non-nucleoside reverse transcriptase inhibitors (NNRTIs) and nucleoside reverse transcriptase inhibitors (NRTIs) drugs stop the process of reverse transcription from HIV RNA into DNA. The integrase inhibitors stop the replication of HIV through blocking HIV DNA from integrating into the host genome. Lastly, protease inhibitors (PIs) block the protease enzyme from continuing the process, necessary for the maturation of the HIV virus to infect another CD4 cell.

Antiretroviral Medication. The most widely used regimens for treating and preventing the progression of HIV are from the following drug classes: non-nucleoside reverse transcriptase inhibitors (NNRTIs), nucleoside reverse transcriptase inhibitors (NRTIs), nucleotide reverse transcriptase inhibitors (NtRTIs), protease inhibitors (PIs) and integrase strand transfer inhibitors (INSTIs) (see Table 1). As of 2016, the WHO recommended regimen given to patients diagnosed with HIV are the first-line regimen drugs: tenofovir (TDF), lamivudine (3TC), and efavirenz (EFV) (see Table 2), particularly in low-middle income countries (LMICS), outside of the United States and Europe. These medications reflect the recommended regimen for clinical use in 2016 and still utilized in 2017 along with strong recommendations prescribe a more effective medication called dolutegravir. Newer antiretrovirals have been approved by the United States Federal Drug Administration (FDA) and only requires two drug regimens, dolutegravir-3TC (Dovato)[36] and dolutegravir-rilpivirine combination (Juluca)[37]. Again, although these new medications exist, this thesis will reflect the HIV medications recommended in the year 2016 and readily accessible in the year 2017.

Figure 1.1: Stages of the HIV lifecycle and HIV Drug classes used to prevent each stage of the lifecycle from occurring.



(*Illustration adapted from aidsinfo.nih.gov)*

Table 1.1: WHO Consolidated Guidelines on the Use of Antiretroviral Drugs for Adults and Adolescents

Generic Name	Dose
Nucleoside Reverse Transcriptase Inhibitors (NRTIs)	
Abacavir (ABC)	300 mg twice daily or 600 mg once daily
Emtricitabine (FTC)	200 mg once daily
Lamivudine (3TC)	150 mg twice daily or 300 mg once daily
Zidovudine (AZT)	250–300 mg twice daily
Tenofovir (TDF)	300 mg once daily
Non-Nucleotide Reverse-Transcriptase Inhibitors (NNRTIs)	
Efavirenz (EFV)	400-600 mg once daily
Etravirine (ETV)	200 mg twice daily
Nevirapine (NVP)	200 mg once daily for 14 days, followed by 200 mg twice daily
Protease Inhibitors (PIs)	
Atazanavir + ritonavir (ATV/r) twice daily) or, SQV/r (SQV 400 mg + RTV 400 mg twice daily), with close monitoring.	300 mg + 100 mg once daily
Darunavir + ritonavir (DRV/r)	800 mg + 100 mg once daily[a] or 600 mg + 100 mg twice daily[b]
Lopinavir/ritonavir (LPV/r)	400 mg/100 mg twice daily

	Considerations for individuals receiving TB therapy In the presence of rifabutin, no dose adjustment required. In the presence of rifampicin, adjusted dose of LPV/r: (LPV 800 mg + RTV 200 mg twice daily or LPV 400 mg + RTV 400 mg
Integrase strand transfer inhibitors (INSTIs)	
Dolutegravir (DTG)	50 mg once daily
Raltegravir (RAL)	400 mg twice daily

a For individuals with no previous use of protease inhibitors.
b For individuals with previous use of protease inhibitors.

Table 1.2: First-line ART regimens, as recommended by the WHO in 2016.

Recommended First-line regimens	**Alternative first-line regimens**
a. TDF + 3TC (or FTC) + EFV, as of 2016 [35,38]	AZT + 3TC + EFV (or NVP) TDF + 3TC (or FTC) + DTG TDF + 3TC (or FTC) + EFV_{400} TDF + 3TC (or FTC) + NVP
b. DTG with a NRTI backbone, as of 2021[39]	TDF + 3TC (or FTC) + EFV_{400}

a As of 2016, the first-line regimens (i.e., EFV-based) and second-line HIV treatment regimens (i.e., previous HIV treatment recommendations).
b As of 2021, the first-line regimens (i.e., DTG-based) and second-line HIV treatment regimens (i.e., current HIV treatment recommendation).

HIV drug effectiveness. In the past, clinical studies have shown that at least three antiretroviral medications from two or more drug classes are needed to suppress HIV (combined into a single pill). HIV treatment has also now evolved to the use of a two-drug regimen (i.e., for dolutegravir-based treatment)[36,37]. The aim of ART is to not only suppress the viral load of HIV-positive individuals and prevent the spread of HIV in the population, but also to prolong the life of HIV-positive individuals. Thus, studies examining life expectancy for individuals diagnosed with HIV aim to understand how interventions such as initiating therapy and when to initiate therapy improve an HIV-positive individual's survival compared to the general, HIV-negative population[40,41]. Mills et al. 2011 demonstrated a significant increase in life expectancy rates for different age ranges in geographical areas where HIV has been most prevalent[42]. The study found

increases by 20 years or more were observed to be the long-term effect of taking antiretroviral medications, regardless of age or gender.

Other studies such as Nsanzimana et al. 2015 also demonstrated the effect of scaling up ART to increase life expectancy[42]. They found that, in Rwanda, before the scale-up of ART (from 1997-2007) the life expectancy for HIV-positive people at age 20 was 20.4 years and after the scale-up of ART (from 2008-2011), the life expectancy was 25.6 years. This increase was also due to the early start of ART, when CD4 count was at least 500 cells per μL[42]. In areas of Europe, the UK, and Canada, these trends in scaling up ART, resulted in increases in life expectancy[43,44]. In addition, observations related to the sex of individuals, age at HIV diagnosis, and diagnosis of other coinfections when starting ART also play a role in the life expectancy. Therefore, the use of ART alone, as an intervention and strategy has improved the long-term care of individuals diagnosed with HIV, moving them closer to the life expectancies of the general, HIV negative population. However, HIV drug resistance poses a possible threat to improvements in the life expectancy of individuals being treated with ART.

Mechanism of HIV Drug Resistance. Mutations form because of the high mutation and replication rates of HIV. HIV has a high mutation rate of approximately 3.4×10^{-5} per base pair per replication cycle[45]. The virus makes mistakes during the replication process of the virus' genetic sequence (i.e., mutations) based on nucleotide substitutions, deletions, and/or insertions. As a result, various viral strains of HIV are formed and increase the genetic diversity of the HIV viral population within a HIV-positive individual.

HIV mutations, generally, have a disadvantage in relation to viral fitness and survival due to replication errors that may prevent necessary proteins from forming or functioning which are necessary in the replication process. However, even though these viral strains do not survive replication, other HIV viral mutant species are able to bypass the pressures of the body's immune system[46–48]. The 'wildtype' viral strain is the HIV viral strain that is not a mutated form of HIV or drug-resistant strain of HIV. Combination antiretroviral medication is very effective at reducing the number of wild-type viruses in the viral population. However, while antiretrovirals continue to stop the replication of

the wildtype viruses, mutant viruses continue to replicate and thus, can dominate the viral population, leading to resistance to the antiretroviral medication and eventually virological failure. Virological failure occurs when viral suppression, or the ability to sustain less than 200 copies per mL, is not achieved. When virological failure occurs, the virus is more likely to be resistant to an entire HIV drug class and prone to cross resistance as the existing mutant population continues to make different populations of HIV. Thus, it is imperative that HIV drug resistant mutations are identified early, and the appropriate alternative treatments are started immediately.

Genetic Testing for HIV Drug Resistant Mutations. Identifying the prevalence of common HIV drug resistant mutants is an effective strategy in reducing the number of drug resistance mutations for HIV-positive individuals and populations. Specifically, genetic testing provides the ability to both identify the variant strains of HIV and their susceptibility to available antiretrovirals, assisting the clinician in finding the best treatment option[49]. The two methods for conducting HIV-DR testing are: genotyping assays and phenotyping assays.

Genotyping assays help to understand the genetic code or instructions for the structure of the virus. HIV genotyping identifies mutations in the pol region of the genome of the virus, where enzymes such as reverse transcriptase, integrase, and protease assist in creating the proteins for the virus and continue the process for making new viruses. When conducting HIV genotyping, a polymerase chain reaction (PCR) test is first used amplify the viral genome and then sequencing is used to determine the order of nucleotides for the DNA of the HIV strain. The sequence is then compared to the wildtype HIV genetic sequence to determine the unique nucleotide changes in mutant strain. Then, the key mutations, or the most common HIV mutations are identified. This process takes about 1-2 weeks and is available, in recent years, at a moderate cost of between US $250-$300 [50–52] Golubchik et al. 2020 demonstrated that conducting HIV genotyping in Zambia could be cost effective when if testing is done during the process of collecting plasma samples for virological failure and in the long-term prevent more drug resistant strains from circulating in the population through the ability to tailor medication based on genotype test results. Thus, in even low-income settings, genotyping may be cost-effective.

Phenotypic assays measure the ability of HIV to replicate in the presence of an antiretroviral medication and thus, can determine the level of drug susceptibility. Particularly, phenotype tests have been stated by some researchers to be more effective than genotype tests for individuals due to the ability of clinicians to adjust the dosing or concentration of the drug for a given patient[53]. However, the process of phenotypic testing tends to take longer to generate than the results of genotype testing. Chang et al. 2015 found that the genotype and phenotype testing to be in 97% concordance for 19 recombinant sequences in response to six antiretroviral drugs tested[53]. In all, both genotype and phenotype tests are relevant to understanding what treatment options best serve the patient in a clinical setting and at the population level.

HIV Mutation Nomenclature. HIV mutations are labeled using a specific nomenclature that relates to the location of the mutation within the HIV-1 RNA strain. HIV nomenclature for mutations is important for researchers and clinicians to use to readily identify mutations that may cause an antiretroviral medication to have reduced susceptibility and be ineffective. Also, having nomenclature for genotyped mutations allows for data standardization and thus, the ability to validate and reproduce results between research studies and clinical trials. The standard structure for HIV mutant nomenclature is: (1) a number detailing the location or position of the mutation along the HIV-1 RNA sequence, (2) an amino acid letter on the left-side of the position number, signifying the amino acid typically found on the wildtype viral strain, and (3) a letter on the right-side of the position number, signifying the amino acid change for the genotyped viral strain, or mutant. Thus, for example, a mutation resistant to protease inhibitors, I47A, means this strain typically includes an amino acid, isoleucine (I), at position/residue 47, but has mutated to have amino acid, alanine (A), at position 47. As such, this naming convention will be used throughout the dissertation to identify major and minor mutations.

Protease Inhibitor Resistant Mutations. The primary function of HIV-1 protease is to make enzymes and structural proteins that allow HIV viruses able to mature towards the end of the viral life cycle[54] Protease inhibitor drugs bind to HIV protease to stop the process of HIV viral maturation. However, when HIV produces errors as the virus reproduces, selective pressure allows for certain mutations that can avoid protease

inhibitor drugs[55]. Protease has 1-99 residues or amino acid positions where major and minor mutations can be found. Historically, Ridky and Leis 1995 found that the most common types of protease inhibitor resistant mutations were I84V, V82I, V82A, I47V, and V32I. These mutations were found to impact inhibitor binding[55]. A more recent study by Thompson et al. 2019 found that patients who failed second-line protease inhibitor therapy in Brazil had similar prominent protease resistant mutations: M46I, I54V, and V82A, with most mutants not changing the viral load of patients significantly in the presence of combinatory antiretroviral medication. Although, the study reported that I47A was associated with an increase in viral load[56]. Some research studies have identified up to 40 protease-inhibitor resistant mutations[57,58].

Nucleoside/Nucleotide Reverse Transcriptase Inhibitors (NRTIs) Resistant Mutations.

The process of replicating the HIV virus begins with the HIV enzyme, reverse transcriptase. In particular, the virus goes through the process of reverse transcription, or the conversion of HIV RNA to HIV DNA. The reverse transcriptase enzyme has two subunits p66 and p51. There are 560 amino acids within the creation of the p66 subunit which facilitates DNA binding and serves as an active site in the process of viral replication. There are 450 amino acids comprising the p51 subunit which is responsible for scaffolding and catalyzing the p66 subunit activity. The mechanism of action for Nucleoside/nucleotide reverse transcriptase inhibitor (NRTI) drugs is to block the DNA chain formation and binding by incorporating itself into the process of DNA binding. However, mutations that form in the presence of an NRTI specifically stop the NRTI binding. Mutations for reverse transcriptase are commonly observed between positions 98-190, such as those associated with NNRTIs and NRTIs, respectively like K103N/S and M184V/I. Although some mutations may occur between positions 225-238 such as P225H, F227L/C/I/V, and thymidine analog mutations (TAMs) like T215Y/F.

Specifically, NRTI mutations such as M184V/I are highly resistant towards antiretroviral medications such as emtricitabine (FTC) and lamivudine (3TC), commonly used first-line antiretroviral medications[59,60]. Another study by Meyer et al. 1999 demonstrates how NRTI mutations assist in the hydrolysis of the inhibitor drug

and thus “unblocks” the DNA replication process, allowing the virus to continue to form the DNA chain and replicate[61]. NRTI mutants have been found to quickly accumulate, within-host[61]. NRTI mutations reduce susceptibility HIV treatment and some NRTI mutations have moderate-to-high levels of resistance to NRTI drugs[62].

Non-Nucleoside/Nucleotide Reverse Transcriptase Inhibitors (NNRTIs) Resistant Mutations. Non-nucleoside reverse transcriptase inhibitors (NNRTIs) also act to bind to the reverse transcriptase enzyme and block the enzyme. The chemical structure of NNRTIs are chemical compounds that are hydrophobic, which allows them to bind to the hydrophobic site of reverse transcriptase. The binding of NNRTIs to RT modifies the structure of residues or amino acid chains required to perform DNA polymerization. NNRTI resistant mutations act to modify both the association (i.e., binding) and dissociation (i.e., the propensity to separate) constants to disable the NNRTI drugs. Many of the older first-line regimens include one NNRTIs and two NRTIs. NNRTI drugs also characteristically have a low resistance barrier and thus, conferring only one resistant mutation can lead to therapy failure to all available NNRTI medications [60]. There are about 22 NNRTI resistance mutations commonly found among the NNRTI medications: efavirenz (EFV), nevirapine (NVP), etravirine (ETR), doravirine (DOR), and rilpivirine (RPV). Previous studies have reported that NNRTI mutants like K103N reduce the drug susceptibility, meaning a higher concentration of the drug is needed to stop the virus[63]. Specifically, some studies such as Marcelin et al. 2006 have shown that for codon, 103, 30% of patients who fail antiretroviral therapy had a K103N mutation [52]. Guha et al. 2012 found that K103N reduces EFV and NVP sensitivity, 25-fold, and 50-fold, respectively[63]. Mutations such a Y181C also are very prevalent in individuals failing therapy that includes NNRTIs, with approximately 5% of patients having the mutation present[63].

Integrase Inhibitor Resistant Mutations.

Integrase inhibitors block the integrase enzyme used to integrate HIV DNA into host DNA of CD4 cells and thus, is one of the steps in stopping viral replication. Mutants that form in the presence of INSTI medications typically catalyze the infection process of allowing HIV DNA integration cell by creating a drug resistant integrase enzyme.

Overall, INSTIs have a higher barrier to resistance compared to other HIV antiretrovirals, meaning more resistance mutations are needed for the drugs to have decreased susceptibility. Raltegravir (RAL) and elvitegravir (EVG) are both antiretroviral drugs that are potent and efficacious in suppressing the viral load of HIV. However, both drugs can lead to drug resistance in the long-term. The drug dolutegravir (DTG) has shown more promise in being both efficacious and having a high barrier to resistance, although resistance has been shown to occur. The mutant, R263K, was shown to have decreased susceptibility to DTG but may have emerged due to the accrual of RAL and EVG resistant mutations[64]. Another study by Bouzidi et al. 2020 reports that the INSTI mutations, L74M/I, lower susceptibility to DTG in the presence of other RAL/ETR resistant mutations[65]. In addition, a few studies have reported DTG-related mutations that were found outside of the integrase region and along the HIV-1 3-polypurine tract (3' PPT) related to the *nef* gene[66]. Patients with mutations found along this region have experienced DTG-treatment failure and the 3'PPT mutations have been identified as highly conserved or unchanged [66].

1.5 Tools and Databases for HIV Drug Resistance

There are several tools created to track, store, and manage information related to antimicrobial and antiretroviral mutations and medications. Some of these databases include: the Stanford HIV Drug Resistance Database (HIVdb) (https://hivdb.stanford.edu/), the Los Alamos National Laboratory (LANL) database (https://www.lanl.gov/collaboration/pathogen-database/index.php), and the EuResist database (https://www.euresist.org/).

HIV and HIV-DR Databases. The Stanford HIVdb was created in 1999 by Dr. Robert Shafer and colleagues. The database was created to allow clinicians, public health researchers and scientists, and educators access to public data on the interaction of drug resistant mutations with antiretrovirals. In addition, the database has evolved into a platform able to analyze HIV sequences to help clinicians determine the best drug regimens for a specific patient using a system called, "Sierra." The Sierra system is a rule-based expert system that uses literature published on both HIV mutations and the susceptibility of antiretrovirals in the presence of HIV mutants. The system specifically

has penalty scores and annotations necessary for interpretation of the system outputs and genotypic results. The rule-based approach to determining antiretroviral susceptibility uses various algorithms that include the ability to reference different sources of data such as research studies and clinical trials regardless of whether raw, empirical datasets are accessible. The algorithms also are verified with expert opinion. The database specifically adds weighted scores to determine resistance. Some of the limitations of the Stanford HIVdb algorithm is that due to the algorithm involving expert opinion, the results are subjective and prone to bias. Additionally, the algorithms used do not provide information about the patient such as viral load information, CD4 count, treatment recommendations, or dosing recommendations.

The Los Alamos National Laboratory HIV database, houses thousands to millions of HIV sequence data and provides tools for analysis and quality control, phylogeny, immunology, alignment and sequence manipulation, data visualization for selected sequences, and general database search tools. Thus, the database provides researchers and clinicians detailed information about individual or a collection of HIV sequences including amino acid changes that indicate genetic differences[67]. Similar to the Stanford HIV drug resistance database, the LANL database only stores information on sequence, genetic and treatment data with accompanying analytical tools that investigate biological and genetic characteristics. However, neither database, has evolved to mathematically model the impact of the biological and genetic characteristics within host or in the population, which could assist both researchers and clinicians in discovering new solutions to preventing, reducing, and stopping HIV drug resistance from occurring.

One platform originating in Italy called *Euresist* attempted to create an HIV drug resistant database and prediction engine[68–70]. The database was an internationally, collaborative project focused on building a modeling system that recommended optimal treatment regimens for people with HIV using clinical and genetic data from patient data. The database and recommender system has significant impact on clinical decision making but is limited to treatment recommendations for individuals and not population-based policies. In addition, *Euresist* does not capture the impact of how each individual patient may contribute to driving the HIV epidemic. Therefore, although

these databases and tools provide a copious amount of information about HIV sequences and the clinical results needed to help with individual HIV treatments and patient care, these tools do not capture the biological, genetic, epidemiological, treatment, and social network data needed understand and analyze the evolution of HIV progression and link within-host data to population-based outcomes.

1.6 Dissertation Outline

In this thesis, I aim to develop tools based on mathematical models to predict and reduce HIV transmission and the levels of HIVDR, including acquired and transmitted drug resistant strains, within-host and in the population, using a stochastic, agent-based HIV epidemic modeling system called *EvoNet*. In this thesis, I answer the main question:

> "What strategies and conditions may assist in predicting, reducing, and preventing HIV transmission and HIV drug resistance, within-host, and between-host (i.e., in a multi-level system)?"

In aim 1 (Chapter 3), the goal is to use a mathematical model to predict the emergence and proportion of acquired drug resistant mutations, within-host, and other outcomes such as viral load, drug concentration, and the other viral mutant strains that makeup single, double, triple, quadruple, or quintuple HIV mutations, within-host. Additionally, in aim 1, we parameterize the model with pharmacogenomic, and pharmacokinetic information related to a first-line HIV drug regimen (i.e., tenofovir, lamivudine, and efavirenz) to predict the emergence of drug resistance, given the presence or absence of a genetic variant, within-host. variation. In Aim 2 (Chapter 4), the goal is to use the mathematical model to investigate the distribution and proportion of the drug resistance, and specifically, transmitted drug resistance and acquired drug resistance, within the population (i.e., between-host). Lastly, in Aim 3 (Chapter 5), the goal is to develop an optimization routine for treatment switching to find model parameter values for the mathematical model that would optimize an HIV outcome such as percent

resistance in the population, and the supply of antiretroviral medication in the population. Next, I will provide more details on the chapters 2 through 6.

Chapter 2 includes substantial information on background information necessary for understanding the contents of this dissertation. Specifically, the aim of this thesis is to find strategies that reduce HIV drug resistance and thus, it is necessary to explore the range of factors contributing to the development HIV drug resistance. Chapter 2 will discuss two types of HIV drug resistance: (1) acquired drug resistance, or when a treatment naïve person undergoes HIV treatment and develops drug resistant mutations, and (2) transmitted drug resistance, or when a treatment naïve person is infected with a mutant or drug resistant strain of HIV. Additionally, Chapter 2 will explore the epidemiologic modeling concepts and methods used to evaluate HIV drug resistance using within-host and between-host models, and the use of multilevel/multimodal modelling in the field of infectious disease. Then, I will address the use of pharmacokinetic and pharmacogenomic data in understanding the emergence of HIV drug resistant mutations.

Chapter 3 (Aim 1) focuses on the development of a model that predicts the emergence of acquired drug resistance within-host. The objective of Aim 1 is to,

> "Model the effect of host genetic variation on the concentration of HIV-DR mutations, within-host, by modeling "patient-specific" pharmacokinetic properties of antiretroviral medication, and drug adherence."

For Aim 1, I explore whether being non-adherent to an HIV treatment leads to greater risks of developing HIVDR than having a higher drug decay rate, and vice versa. In this chapter, I model the effect of host genetic variation on the frequency of HIV-DR mutations, within-host, using pharmacokinetic information for antiretroviral medication, and patient-specific information such as the levels of drug adherence. Given that numerous studies have demonstrated the effect of drug adherence on virological failure, this chapter also aims to determine if certain levels of adherence lead to an additive effect on the emergence of drug resistant mutations emerging.

Chapter 4 (Aim 2) investigates the presence of drug resistance (i.e., both acquired and transmitted or pre-treatment drug resistance) in the population. The objective of Aim 2 is to,

> "Model the effect of host genetics and prevalent drug resistance mutations on HIV-DR levels in two sub-Saharan African populations."

For Aim 2, I identify the populations that may be more likely to develop drug resistance, given the frequency of fast or slow metabolizers in the model, or the presence of individuals with a K103N mutation. Particularly, I model the effect of viral genetics and prevalent drug resistance mutations on HIV-DR levels in two different sub-Saharan African populations. The modelling study will stratify the difference between acquired and pre-treatment drug resistance to assess an overall burden of transmitted drug resistance on the population. Furthermore, this chapter will model the effects of three levels of drug metabolizers in the population.

Chapter 5 (Aim 3) tackles finding the optimal values for the model that will achieve a reduction of drug resistance in the population and the prioritization of individuals for second-line treatment. The overall objective of Aim 3 is to,

> "Create an ART optimization routine for a stochastic, network-based model that will modify drug dosing and/or switch treatment regimens, given the patient's characteristics and the threshold level of a desired HIV outcome."

For Aim 3, I determine which patients to prioritize for second -line therapy, given the availability of resources under specific cost and conditions. This chapter focuses on using an optimization algorithm a rule-based logic similar to the grid search method, hill climbing and simulated annealing. I explore the tradeoff between adjusting one or more parameter values of the model to reduce a cost function, specifically.

Finally, in Chapter 6, I conclude the dissertation with a summary of research findings for aim 1 (Chapter 3), aim 2 (Chapter 4), and aim 3 (Chapter 5). I then address limitations of this study. This chapter also compares the research findings to other sources of empirical data related to HIV and HIV-DR. Lastly, I discuss future research

plans for this thesis work and next steps for using informatics tools and approaches to investigate infectious diseases and specifically, reduce and prevent HIV transmission and drug resistance.

Chapter 2: Background

2.1 Goals for Ending the HIV/AIDS Epidemic and Stopping HIV Drug Resistance

To understand the context of my exploring via mathematical modeling my overarching question of "What strategies and conditions may assist in predicting, reducing, and preventing HIV transmission and HIV drug resistance within-host and between-host (i.e., in a multi-level system)?" it is important to understand the goals for ending the HIV/AIDS epidemic and strategies being used to stop HIV drug resistance.

The Joint United Nations Programme on HIV/AIDS (UNAIDS) aims to end the HIV/AIDS epidemic by 2030[71–73]. In 2016, UNAIDS established the 90-90-90 plan, with the goal of the following measures would be achieved by year 2020: 90 percent of people living with HIV (PLWH) would receive HIV testing, 90% of individuals diagnosed with HIV would receive antiretroviral treatment, and 90% of individuals receiving antiretroviral treatment will have suppressed viral loads below 1000 copies per milliliter. Achieving the UNAIDS 90-90-90 plan goals would result at least 73% of PLWH having viral load suppression. However, the goals of the 90-90-90 plan were not achieved by the year 2020[71]. In 2019, the UNAIDS reported that 81% of people living with HIV were tested, 67% were on ART, and 59% had achieved undetectable viral load levels, globally[71]. In response, more action is being taken to determine the impact of various factors that may contribute to the failure of the 90-90-90 plan goals being reached, specifically including the impact of HIV drug resistance (HIV-DR) on the viral load suppression to undetectable levels.

In recent years, HIV-DR (HIV drug resistance) has increased in the population, contributing to challenges in the efforts to prevent, stop the progression of, and end the epidemic of HIV/AIDS. To combat the emergence of drug resistant mutations, within-hosts and in the population, surveillance surveys, databases, and models have been created to capture laboratory, clinical, and public health data.[74,75] The World Health Organization (WHO), along with other public health organizations, also sought to track

early warning indicators (EWIs) of HIV-DR. These indicators include the following: the percentage of ART prescriptions in agreeance with national and international recommendations, clinical retention at 12 months, on-time pill pick-up, completed on-time appointments, drug stock-outs, viral load suppression, and the completion of viral load testing at 12 months. EWI surveys are conducted over several years and gather data from healthcare facilities to report (on a Likert scale) a target performance percentage ranging from 0% to 100%, categorized into excellent, fair, or poor performance for each indicator[76–78]. Additionally, the report highlights thresholds of HIV-DR observed in the population that could have deleterious effects on progress in reducing and preventing HIV transmission, morbidity, and mortality.[79,80]

In this chapter which sets the stage for my aims, I: (1) provide information on the types of HIV-DR and research that identifies acquired versus transmitted drug resistance, (2) discuss different types of epidemiological models, and (3) the structure of modeling infectious diseases at different scales (within-host, between-host, and multi-scale) and the application of these scales to modeling HIV transmission and HIV-DR.

2.2 Types of HIV Drug Resistance

HIV drug resistance (HIV-DR) is classified as either: acquired, transmitted, or pretreatment drug resistance. Each type of drug resistance can lead to virological failure, because of ART not suppressing the viral load in the body. Understanding how drug resistance developed in a person is crucial to determining the best regiment for short- and long-term treatment. Additionally, determining the type of HIV-DR can provide insight as to how the epidemic may be evolving in the population.

Acquired drug resistance (ADR) happens when a person living with HIV has HIV mutations that emerge after starting treatment. Pre-treatment drug resistance (PDR) occurs with individuals that are drug-naïve or individuals that have prior drug exposure and are initiating or re-initiating ART such as with prevention of mother-to-child transmission (PMTCT) regimens where the child has had a prior exposure to the medication and may develop drug resistance[81]. Transmitted drug resistance (TDR) occurs when a treatment-naïve individual has been infected with a virus with existing drug resistant mutations before starting treatment. To properly address strategies that

will reduce and eradicate levels of drug resistance, there must be a consensus whether an individual has obtained ADR, PDR, or TDR.[82]

A study by Yang et al. 2015 investigated the importance of evaluating the difference between acquired versus transmitted drug resistance in a Swiss HIV Cohort study from 1997 to 2011, as new antiretroviral drug classes were introduced. The research team found that using genotypic testing, ADR mutants were found and declined from 85% to 38% within the study population that were treatment experienced and had virological failure (VF), over the course of 12 years. However, TDR mutations in the population fluctuated with the introduction of new antiretroviral medications.[82] Another study by Aldous et al. 2017 found that ADR accounted for most of the drug resistance found in the population of HIV positive patients living in the Washington, DC, where HIV prevalence rates are high in the United States. [83] Aldous et al. 2017 reported an ADR prevalence of 40.9%, (n=309).[83] In addition, in comparison to the transmitted drug resistant mutations, ADR mutations were equivocally the same in proportion.[83] The highest prevalence of drug resistant mutants being the K103N (18.8%) and M184V (17.8%), of the NNRTI and NRTI drug classes, respectively (see Figure 2.1).[83] These findings support the continued need for new drug medications to ensure viral load suppression in a population of either treatment experienced or treatment naïve individuals, as these new antiretrovirals may be effective for an unspecified time period.

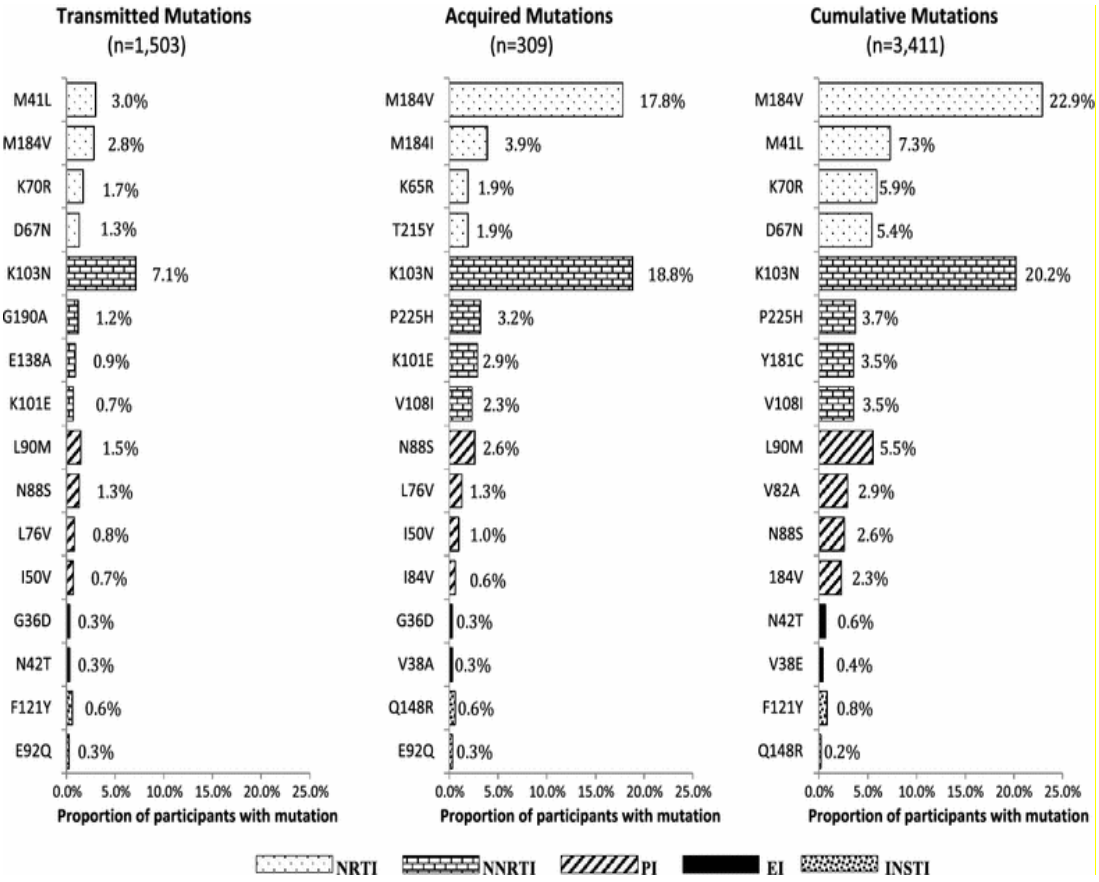


Figure 2.1. Proportion of major and most common drug resistant mutations by drug resistance type (transmitted or acquired) and categorized by drug class. The drug classes include nucleoside reverse transcriptase inhibitors, NRTI; non-Nucleoside Reverse Transcriptase Inhibitor, NNRTI; Protease Inhibitor, PI; Entry Inhibitor, EI; Integrase Strand Transfer Inhibitors, INSTI. Source: Aldous AM, Castel AD, Parenti DM. Prevalence, and trends in transmitted and acquired antiretroviral drug resistance, Washington, DC, 1999-2014. BMC Research Notes. 2017;10(1):474. doi:10.1186/s13104-017-2764-9 [83]

Contributing Factors to the Development of HIV-DR. Several studies, [84–87] have shown that poor drug-adherence, or on-time pill-taking is a contributing factor to the emergence of HIV-DR. Poor adherence to ART causes treatment switching to a second or sometimes third regimen impossible, resulting in virological failure.[88] Thus, HIV adherence education and periodic genotypic is often recommended and practiced when drug adherence is an issue because ADR is more likely to occur in face of poor adherence.[88]

Other studies have quantified the threshold at which viral suppression can be achieved with an inverted U-shaped curve, or non-linear curve describing the relationship between percent adherence and the probability of selecting a mutation.[89]

Nachega et al. 2011 illustrated that a moderately low to moderately high drug adherence levels (30 to 70 percent) may select for a higher likelihood of selecting for a drug resistant mutation (see Figure 2.2 by Nachega et al. 2011)[87]. Oyugi et al. 2007 investigated the threshold of adherence needed for the suppression of HIV among HIV-positive patients in Kampala, Uganda.[90] Researchers found that for patients that had an adherence >=95%, viral suppression ranged from 50% to 80% versus patients with an adherence <95% and had high rates of drug resistant mutants. Although these percentages may be high, there are other studies have reported good adherence to be at levels between 70-90% for certain drug classes like NNRTIs and boosted PIs.[89]

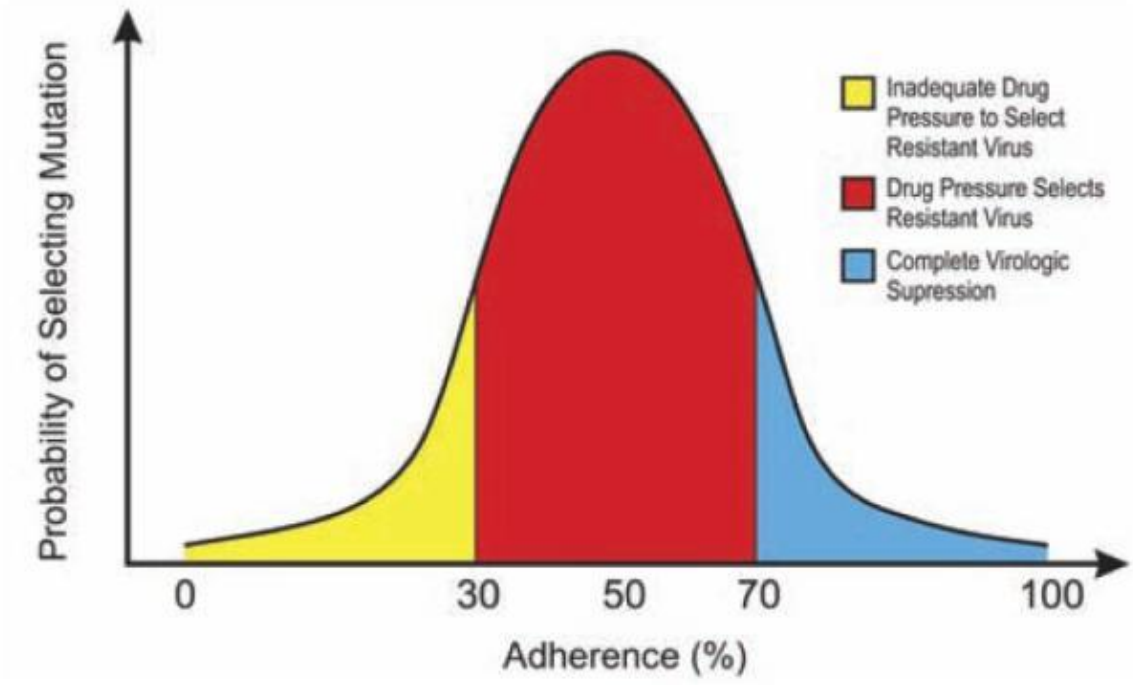


Figure 2.2. A generic example of the relationship between percent adherence and the probability of an individual selecting for a resistant mutation[87,89]. Figure from Nachega JB et al. Marconi VC, van Zyl GU, Gardner EM, Preiser W, Hong SY, Mills EJ, Gross R. *HIV treatment adherence, drug resistance, virologic failure: evolving concepts*. (Infectious Disorders Drug Targets, 2011), 17. Journal article.

Although missed doses (i.e., in relation to pill taking) is associated with poor adherence, the burden of pill-taking and social stigmas affecting an individual's mental health also play a role in development of ADR and virological failure.[91–93] The stigma of pill-taking often leads to cycles of non-adherence (i.e., during holidays and vacations)

and signifies the importance of tracking the emergence of acquired drug resistant mutations. Some studies[93–96] have also found that psychological distress, trauma and/or the diagnoses of a mental health disorder is associated with an individual having distrust in the effectiveness of their ART regimen and thus, do not take medication as prescribed and miss doses. Therefore, utilizing electronic tools and methods that track adherence and on-time pill-taking using electronic tools, and periodic viral load (with reflexive DR genotyping) testing have been recommended[97,98].

Another contributing factor to the development of ADR is the length of time an individual is on antiretroviral medication.[99,100] Villa et al. 2018 reported that although individuals diagnosed with HIV had a self-report of 100% drug adherence and followed the prescribed drug regimen, major resistance mutations were observed at the baseline or start of the study and at 6-months while being in ART.[100] The study found that high levels of adherence that still lead to HIV-DR could be a consequence of a higher drug selective pressure and signifies the need to know when to switch to a second-line regimen.[100] Bangsberg et al. 2003 conducted a study looking at 330 HIV-positive patients over a 12 month period, finding 23% of had adherence levels of 90-100%, and still experienced an accumulation of drug resistant mutations.[99] This finding was surprising considering the HIVDR mutations emerged at levels of approximately 81% adherence and did not improve with increasing levels of adherence to 100%. However, self-reported data may not be as reliable as clinical and laboratory follow-up testing and thus, may not be reliable or reflect actual adherence levels. Not only did the study observe major ADR at high levels of adherence, but researchers also identified that certain drug resistant mutations commonly occurred at high levels of adherence.[99] This study demonstrated that the drug class was indicative of the threshold of adherence needed to fully suppress the virus and reduce drug resistant mutations. For instance, participants in the study were administered combination therapy, but had a had an increasing risk of resistance to protease inhibitors, peaking between 70%-80% adherence; ritonavir-boosted protease inhibitors peak for risk of resistance between 45-55% and have a significant decrease in risk of resistance between 60-100% adherence, within a small change in risk of resistance between 90-100%. However, NNRTI-based drugs, have a significantly decreasing risk of resistance as adherence levels increase. Thus, reduction in HIV-DR, within-host, may involve determining the level of

adherence needed according to findings from medication labels and published works from studies like Bangsberg et al. 2003 which detail the levels of adherence where drug resistance mutations are likely to replicate and emerge.

Another explanation for high levels of drug resistance may be due to the genetic barrier of certain antiretroviral medications[101,102]. For example, antiretrovirals such as NNRTIs like efavirenz have low barriers to resistance and thus, a single point mutation is needed to confer resistance and cross-resistance to all other NNRTIs[101]. In addition, NRTIs like lamivudine and emtricitabine have low barriers to resistance[101]. Other NRTIs like thymidine analogues, and tenofovir, have a moderate genetic barrier to resistance and require a few resistance mutations to emergence for drug resistance to be diagnosed[101]. Whereas INSTIs, like dolutegravir, have a high genetic barrier and thus, will require the accumulation of many mutants for drug resistance to occur[101]. The accumulation of resistant mutations does not usually occur all at once but may be sequential. This accumulation of mutations over time may also occur when the wildtype sequence of a person diagnosed with HIV only has 1-to-2 nucleotide changes needed for an amino acid change to occur.

A major mutation is a mutation that is first to appear in the viral population and by themselves, reduce the susceptibility of an inhibitor[62]. A minor mutation is a mutation that later occurs in the viral population and does not significantly affect inhibitor susceptibility[62]. The accumulation of minor mutations has also been shown to improve the replicative fitness of major resistant mutations. Thus, understanding the genetic barrier to resistance and the presence of minor mutants can assist clinicians with prescribing the optimal treatment for individuals diagnosed with HIV, a method we see when describing genotype and phenotypic testing. HIV-DR and poor adherence could also be a result of a cascading effect of antiretroviral side effects and the accumulation of minor drug resistant populations.[103] For example, if an HIV-positive patient's body responds to an antiretroviral medication with intolerable side effects, that patient may stop or become inconsistent with taking the drug leading to poor drug adherence.[103] In turn, this has created an ideal environment for the, once, low accumulation of minor mutations to reproduce and subsequently virological failure to occur.[104] Thus, using mathematical modeling to capture the factors such as replicative

fitness that contribute to the transitions between wildtype and mutant viruses also is essential to predicting HIV-DR.

2.3 Previous Mathematical Modelling Approaches to Predicting Drug Resistance

Mathematical Modelling and HIV. Mathematical modeling has historically been used to predict and analyze outcomes in various disciplines such as the social sciences (political science, psychology, sociology, economics), computer science, engineering, linguistics, the life sciences (biology, chemistry, earth, and space science), medicine and specifically, the in public health settings. Mathematical modeling in the fields of public health and epidemiology involves using individual characteristics data (i.e., biological, social, and behavioral data) to predict population-health level characteristics [105]. The purpose of mathematical models in the epidemiology of infectious disease is to predict, assess and control potential outbreaks. These outbreaks can occur via indirect transmission (i.e., airborne, or vector-borne) and/or direct transmission (i.e., droplet spread, skin-ot-skin, or sexual intercourse).

After the 2014 Ebola outbreak in Africa that claimed thousands of lives in Central and West Africa, modeling studies investigated the potential contributions that created rapid spread of the disease from direct contact (human-to-human transmission [106]). Berge et al. 2017 used a mathematical model that determined whether the rapid spread of Ebola was a result of endemic conditions (i.e., provisionary processes like eating bush meat/fruit from rainforests, the orchestration of funeral proceedings for those that died from contracting the Ebola virus, or sanitary practices in the area) or global contact.[106] The study found through modeling that disease transmission and severity increases with the inclusion of provisional practices such as eating bush meat and decreases in the absence of those practices [106]. However, these findings involved factors that may have not contributed to the outbreak of Ebola in West Africa, which again, researchers believe was due to person-to-person transmission (i.e., from direct contact with bodily fluids). As such, infectious disease models describing the Ebola transmission relationship, for instance, may include sets of parameters that evolve over time when more information is found or in this case, when the existing information does not support the rapid

increases in transmission. Thus, in the case of Ebola and other instances of infectious disease outbreaks, health policies and practices that otherwise would have taken involved more resources and time may be simulated, using mathematical modeling.

In relation to HIV, mathematical models have been used to understand the cost-effectiveness of certain treatments and/or interventions and ultimately used to promote and implement healthcare and public health policies and laws. Bohdan et al. 2019 found that using a cost-effectiveness model, policies should tailor interventions and health promotion practices such as adherence to ART that increase quality-adjusted life-years (QALYs) instead of interventions developed to prevent new infections.[107] Other studies have investigated the cost-effectiveness of certain ART regimens. Nosyk et al. 2015 assessed ART scale-up in Columbia, Canada, as patients have access to universal healthcare, specifically for individuals diagnosed with HIV/AIDS[108]. The mathematical model simulated the HIV/AIDS epidemic from 2007-2010 and estimated epidemiological outcomes for HIV like morbidity, mortality, costs on healthcare expenditures and QALYs[108]. The results indicated a linear relationship between access to ART and the number of averted infections. A longer simulation of the cost-effectiveness model spanning 38 years and the comparison between the observed scale-up of ART ("actual practice") and modeling scenarios of access to 75% and 50% of the observed ART scale-up, indicated a total savings of 25.1 million dollars and 65.5 million dollars when comparing model predictions for using the actual practice versus 75% and 50% of the actual practice, respectively[108]. Thus, mathematical models can also serve in creating scenarios related to funding and costs prior to the scale-up of any intervention and can ultimately assist in determining the feasibility of the study.

Clinically, HIV mathematical models have been used to predict and understand clinical outcomes like virological failure, within-host related to viral dynamics. Kiweewa et al. 2019 studied the factors contributing to an increased risk of viremia, persistent viremia, and virological failure using a generalized linear model to assess relative risk.[109] The researchers found that using a statistical analysis such as the multivariate analysis (generalized linear model, GLM), allowed them to explain the relationship between the covariates (i.e., medical history, physical exam, demographic and behavioral questionnaire characteristics, ART treatment history, and laboratory tests to identify CD4 count, viral load, and the binary variable categorizing virological failure) and the

response variables (i.e., virologic failure: plasma HIV-RNA ≥ 1000 copies/mL at a recent clinic visit, viraemia: plasma HIV-RNA ≥ 50 copies/mL at a recent clinic visit, and persistent viraemia: plasma HIV-RNA ≥ 50 copies/mL at two consecutive visits) in a linear, additive approach. Thus, the findings included an increased risk of virological failure if attending a clinical care site other than the Ugandan clinic site and being on a 2nd line regimen, taking other antiretroviral combinations that were not first or second-line regimens, missing adherence, a low CD4 count and having a history of fever within a week. In contrast, an occurrence of virological failure was less likely if the CD4≥ 500 cells/mm3. Thus, in this case, statistical modeling helped to identify predictors of virological failure providing a targeted approach to address issues surrounding ART regimens and adherence, as well as the need to further investigate specific care clinics to address unique predictors of virological failure at the specified site.[109]

Some models have also been utilized to understand other biological characteristics such as the multi-phasic viral load decay[110]. These decay models help researchers understand the occurrence of blips in the viral load, and the emergence of latent (long-lived) cells at the early stage of infection and while on treatment. These mathematical models often lead to testing scenarios for HIV eradication to occur, how to initiate treatment, and the ability to design new strategies for social, behavioral, and pharmacologic therapies. One study investigated viral latency that threatens the suppression of HIV using mathematical models.[111,112] Chun et al. 1997 investigated whether latently infected cells decay after antiretroviral are present in the host and how quickly the cells may decay, using a model of exponential decay. The exponential decay model used in this study was based on prior models that used single exponential decay that applied predictions to decay beginning immediately after antiretroviral treatment is administered and estimated a timeframe of 10 years before the latently infected cells died out[110]. Thus, HIV biological models serve as tools that approximate the transitions from one disease state to another on a molecular and cellular level and provide additional information that contribute to the development of epidemic models used to understand disease progression at the population-based level.

The next sections discuss the types of epidemic and infectious disease models used for disease tracking/surveillance. Then, I detail the structure and parameters

needed to build and simulate within-host and between host models for predicting HIV-DR.

2.4 Types of Epidemic and Infectious Disease Models

The basic framework of models includes elements, states, and transitions. Elements are the actors in a model.[110] Examples of elements may include human host, animals, pathogens, and vectors. States are characteristics of the elements such as an infection status, life cycle, and demographic status.[110] Transitions are movements or the rate of movement between states. These transitions are traditionally represented through two types of models: deterministic or stochastic.[110] A deterministic model is used to predict what occurs in a population, on average. The input values for deterministic models are fixed and include information like the rate of disease onset and/or the rate of recovery. As such, the outcomes of the model are known because they can be pre-calculated based on the set of initial conditions, set of equations used, and input values. Deterministic models are typically useful with more applicable large populations. A stochastic model involves input parameters that move from one compartment to another based on probability. Thus, the outcome variables have a predicted value that range for each simulation due to the probability that a specific outcome will occur, given the probability of the input value occurring. Next, we present the basic setup of each model type, how the general predictions are determined, and how these models apply to HIV dynamics within-host and between-host.

Deterministic Models. The structure of a deterministic model includes compartments or sub-groups that the population is divided into based on, for example, their infection status in the population (e.g., susceptible, infectious, or recovered).[113] Differential equations or difference equations are used to establish the formula needed to define the transitions from one compartment to the other. A difference equation is a formula that computes the n, number of individuals in each compartment of the model using time steps (e.g., expressed in days) for various points in time such as t+1 (tomorrow) using the numbers computed from the previous day, t.[113] The following

equations explain how to compute the values using difference equations for a Susceptible-Infected-Recovered model (SIR).

1. Calculate the number of susceptible people.

Number of susceptible people at time, t+1 =

(Number of susceptible people at time t)

–

(Number of susceptible people newly infected at time t and t+1)

- The number of susceptible people newly infected at time t and t+1 is referred to the multiple of the risk that a susceptible person is infected and denoted by, λ_t and the number (count) of susceptible people at time t, denoted by, S_t.
- Thus, the difference equation for the number of susceptible people at time, t+1 is:

$$S_{t+1} = S_t - \lambda_t S_t$$

2. Calculate the number of pre-Infectious people. Also, pre-infection happens when the host is infected by the virus and is at a state where the host can transmit to a susceptible person during a specified duration of time.

Number of pre-infectious people at time, t+1 =

(Number of pre-Infectious people at time, t)

+

(Number of newly infected people between time, t, and t+1)

-

(Number of infected people that are infectious between time, t, and t+1)

- The addition sign (+) represents the adding of people to the infected population as those people are leaving the susceptible compartment.

- The product of the proportion of the people who are pre-Infectious between time t and t+1 (denoted by the f), and the number of pre-infectious people at time, t (denoted by E_t), can be written as, $f E_t$.
- Thus, the difference equation for the pre-infectious compartment is,

$$E_{t+1} = E_t + S_t - f E_t$$

3. Calculate the number of infected people.

Number of infectious people at time, t+1 =

(number of infectious people at time, t)

+

(Number of people that transition to being infectious between time, t, and t+1)

-

(Number of infected people that are no longer infectious between time, t, and t+1)

- The product of the proportion of the people who stop being infectious between time t and t+1 (denoted by the r), and the number of infectious people at time, t (denoted by I_t), can be written as, rI_t.
- Thus, the difference equation for the pre-infectious compartment is,

$$I_{t+1} = I_t + fE_t - rI_t$$

4. Calculate the number of recovered people.

Number of immune people at time, t+1 =

(Number of immune people at time, t)

+

(Number of people that became immune between time, t, and t+1)

- The number of immune, or recovered people at time t is denoted by R_t. The product of the proportion of the people who stop being infectious between time t and t+1 (denoted by the r), and the number of infectious people at time, t (denoted by I_t), can be written as, rI_t.

- Thus, the difference equation for the recovered compartment is,

$$R_{t+1} = R_t - rI_t$$

The deterministic model works by beginning with initial values for the number of people that are susceptible, pre-infectious, infectious, and recovered, at time, 0, or the first day of the observed outbreak.[113] The equations in steps 1-4 detail the formulas and calculations needed to predict the next day of the outbreak (day 1) and in turn, the values from day 1 are used to calculate day 2 and so on. Deterministic models like these can be observed in numerous studies for understanding infectious disease and in particular HIV.[113] Hoare et al. 2009 demonstrated how a deterministic model investigated the impact of not having a second-line regimen available in Southeast Asian countries (such as Thailand)[114]. The model was structured to compartmentalize the infected population into two groups: (1) infected with drug sensitive virus and (2) infected with the drug resistant virus [114]. The 13-disease progression and treatment compartments all include average rates of transition and thus, the outcomes of the model reflect a set of fixed outcomes including whether a majority wildtype, major drug resistant mutant, or minor drug resistant mutation viral population[114]. The study found that approximately 24% of new infections would be from a drug resistant virus. The advantage of using a deterministic model in this case was that the model helped researchers validate the assumption that having more accessibility to second-line regimens will reduce potential drug resistant levels[114]. However, like most deterministic models for infectious disease, incorporating time-variance of infectivity and the ability to incorporate multi-strain resistance also remains a challenge. Thus, stochastic models provide a level of randomness and probability of transitional states that sometimes more accurately approximate model outcomes.

Stochastic Models. Stochastic models estimate the probability of different outcomes given the variability of one or more parameters.[113] Specifically stochastic models are used when there is variability in the transmission, birth, death, infection, or recovery that predict a variability of outcomes.[113] Ordinary differential equations are used to calculate model predictions. The use of stochastic models also depends largely on what type of question is being asked and/or the type of data available or collected.

The setup for a stochastic model includes using ordinary differential equations. In the simplest form, the stochastic form of a susceptible-infected-recovery model is consistent with the setup of a deterministic model. However, the parameters for the susceptible and infected compartments are random variables and depend on infection and recovery probabilities. The following assumptions are the basic structure of a stochastic Susceptible-Infected-Recovery (SIR) model as described by Ming et al. 2016.[115] The number of susceptible individuals (S), along with the number of infected individuals (I) and the number of recovered individuals (R) are tracked using the following assumptions [115]:

1. The population of infected people, $I_i(t)$ is small in comparison to the sub-population, $S_i(t)$.
2. $S_i(t)$ is a constant.

Thus, the following stochastic differential equations are representative of a stochastic SIR model with the assumptions above[115]:

$$dI_i(t) = (\alpha + \delta_i I_i(t)dt + \sigma_i d\beta_i(t),$$

where α is the auto recovery rate for the illness and δ_i determines the ability of the disease to transmit in different sub-populations. $\beta_i(t)$ is the standard Brownian motion that determines the probability of moving from one state to another, and σ_i is the measure of spread of disease from neighbors.

Advantages and Disadvantages of Deterministic and Stochastic Models

Although there are different classifications of models, for the purposes of simplicity, this section aims to compare the advantages and disadvantages of deterministic and stochastic models. Deterministic models, as previously described, have a fixed set of inputs that yield a set of outputs that are pre-determined or because the fixed input values will tend cause the direction of the model output. Thus, when applied to a real scenario, deterministic models can describe what happens, on average. Using averages is suitable when answering questions about large populations or groups of individuals. However, due to the lack of random variation in the model, measures of uncertainty

related to individual interactions in real scenarios are not accounted for in deterministic models. These measures of uncertainty are typically needed when modeling outcomes inform decision-making. Another disadvantage of a deterministic model is in the numerical values related to individuals in the population, where an individual(s) can be described as a fraction, the model will still predict that transmission can occur, and the infection will persist[113].

In contrast to deterministic models, stochastic models do not predict that transmission will occur when in cases where an individual is counted as a fraction, as this scenario is not representative of a real-world account of transmission between two individuals[113]. For this reason, stochastic models usually are appropriate when simulating outbreaks in small populations, as the stochastic models work with addressing individuals with numerical integers. Thus, once the population, n, decreases to less than 1, transmission will no longer occur. In healthcare populations like clinics, this may be beneficial as some empirical and observational studies have demonstrated that once a small number of people are infectious, transmissions will no longer occur.[113] In addition, these models have a discrete time continuum that is tracked over time, making it necessary to have a different set of mathematical equations to be used to capture models that incorporate time being continuous. Stochastic models are closer to the real, empirical observations and the system being modeled. Differential equations calculate the changes in time for each of the stochastic parameters whose value may be in the form of a distribution (i.e., normal distribution, Poison, etc.). To this end, the results of the repeated simulations in a stochastic model can differ even with slight changes in the parameter values or when the values remain the same. Stochastic models, although more complex than deterministic models, due to the introduction of uncertainty and chance in the model, also may take longer to parameterize and perform model fitting based on available data. Thus, the objective of stochastic and deterministic models differs in the building blocks of how model predictions occur, either as a function of time or time and probability, respectively.

Agent-based models (ABMs). The purpose of an agent-based model is to simulate the relationship and interactions between individuals (i.e., hosts, organizations, groups). To this end, agent-based models help researchers understand behavior and other outcomes

[116]. In the 1970s, models related to social and/or economic phenomena and were the first set of agent-based models. [117,118] Schelling infamously utilized agent-based modeling to describe and theorize racial segregation.[118] The ABM model detailed the agents in terms of preference to live near a neighbor that looked like them: if two or more of the new neighboring houses had individuals that looked like them, then they are happy, else if one or less then the house is unhappy. This logic theorized that racial segregation is built from individual preference to be in proximity to a homogeneous group of people but could not detail a direct cause to the model's simulation of segregated neighborhood formation.[118]The result of ABMs in the field of social science resulted in understanding residential patterns and trends within neighborhoods. In the same way that patterns and trends were identified to address racial segregation using Schelling's ABM, infectious disease can also utilize the individual characteristics related to a propensity to transition from one state to another state.

ABMs also aid in decision-making needs, specifically to help formulate interventions and policies. The use of these models has been particularly helpful in the areas of ecology, biological systems, health, and policy, where an ABM may also be referred to as individual based models (IBMs) [116]. Using stepwise logic or (i.e., rules) and more complex non-linear mapping to execute the decision-making, other tools like neural networks and genetic algorithms have also been incorporated into the ABM framework.[117] In this way, using rules and algorithms are beneficial in the process of understanding host-pathogen relationships, where a pathogen is an infectious agent that can infect or cause illness to the host.[117] Pathogens can include the 2019 coronavirus virus, Mycobacterium tuberculosis, the influenza virus, and HIV. In the same way that the immune system works to learn the mechanism of action of the pathogen and store memory, computer systems can also store information about pathogen-host interactions and the evolution of the pathogen in the host using rules or algorithms. This is the basis of agent-based modeling (ABM). The benefit of computer systems tracking the changes of the pathogen (i.e., the agent) within the host is that we can capture a dynamic process with limited costs to the ethical measures of the process such as needing to infect a susceptible agent along with lessening the financial costs that may be associated with buying materials carryout the experiment. The implications of using computer-based

systems to conduct disease modeling also benefits both science and the community/individual effected, especially when tracking bigger populations. Importantly, the stage of the host-pathogen relationship can also dictate what type of modeling tools are used.

Advantages and Disadvantages of ABMs. Specifically, ABMs differ from purely deterministic and stochastic models, in that ABMs can capture more complex and dynamic relationship based on the attributes of individuals.[116] The structure of the model is usually built as a stochastic model, capturing the probabilistic and chance interactions that occur in the natural world and includes rules or algorithms signifying the way agents interact with other agents and their environment[117]. Regression-based methodology usually captures and tries to predict the cause and effect in the field of epidemiology and other health-related disciplines. However, regression-based methods are not always able to capture the dynamic, interactive, and contact-based systems, detailing how each individual action influences another individual's disease state or condition and the time-space continuum that may also affect the next action of everyone in the model. ABMs also allow the researcher, clinician, or healthcare professional to investigate observational/empirical data, experimental, and theory-based approaches that may otherwise not have occurred or may be unethical to investigate. Although ABMs assist in testing, understanding, and predicting otherwise complicated systems, the models may be difficult to build as modeling relies on valid and accessible data on a particular phenomenon. Thus, if there is little evidence of a particular occurrence or the information is not publicly/readily accessible, the model may not accurately capture enough information to be informative or reliable. To this end, ensuring that data validation through comparing model outcomes to actual, real-world, and empirical data is conducted via published literature and clinical trials. Overall, no models can capture every dynamic needed to be completely accurate, and as such, ABMs remain a useful tool in disciplines such as infectious disease.[117,119]

ABMs and HIV transmission/HIV-DR. In the field of infectious disease, ABMs have assisted in detailing how HIV viral strains evolve and reproduce in a single host and have historically captured intracellular interactions and the spatial conditions of the cells and virus. Zarrabi et al. 2010 used an ABM to capture the act of HIV replication

with one replication cycle being observed per simulation.[120] The ABM captured parameters based on the process that the virus goes through beginning with cell entry to the integration of viral cDNA into the cell DNA and divides this process into six stages (a varied stages, including an uninfected cell, early infected, latently infected, actively infected, productively infected, and dead).[120] Thus, the parameters include intracellular states and genomic information impacting cell changes in the form of rates and frequency/counts of cellular agents. The model calculated the count of viral cDNA and mRNA at the end of each replication.[120] The use of an ABM model to capture the steps/events necessary for HIV replication to occur in a single cell also demonstrates how ABMs can influence and find ways to eliminate the virus and stop its replication, for example, by incorporating drug therapy into the models. Another study by Castiglione et al. 2007 captured the effect of ART and the use of scheduled treatment interruptions to offset adverse side effects of ART for individuals diagnosed with HIV. Their ABM was used to capture the events occurring in the immune system and a genetic algorithm was used to interrupt the adverse events happening in the immune system. This, in turn, identified times needed to stop and restart therapy to lessen poor immune system function (i.e., such as the emergence of opportunistic infections) and does not progress the disease to AIDS or death.[121]

ABMs and population-based HIV transmission. The use of ABMs to capture contact networks and the transmission of HIV in the population has been useful in understanding social behavior and the impact on societal. Adams et al. 2018 utilized an ABM to investigate the impact of the mass incarceration of African American males on HIV acquisition among African American women in Philadelphia and the factors that may modify HIV acquisition in African American women.[122] The model described the movement of African American men being admitted, released and/or re-admitted into the prison or jail.[122] Each timestep captured the movement of the men and out of prison/jail and the sexual relationship they encountered with African women [122]. Epidemiological outcomes included HIV transmission, HIV disease progression, HIV treatment administered, and HIV testing as well. The ABM model specifically included an algorithm of rules where each agent has a probability of interacting with an agent and the environment such as a prison or jail. Rules were programmed as algorithms such as

the partnering algorithms, sexual contact algorithms with the probability of transmission occurring. Depending on these algorithms, the state and agent attributes for risky behavior, HIV testing eligibility, HIV status, and survival are also updated as men enter and leave the environment at each timestep. The data used to reflect probability and prevalence of parameters such as gender makeup, sexual orientation, HIV serostatus, diagnosis of HIV, adherence to ART, HIV disease state, and incarceration status were gathered from published literature and accessible data from the Philadelphia Commission on Sentencing.[122] The study found that the increase in partnerships among African American males incarcerated was the main contributing factor to HIV acquisition among African American women and not an increase in risky behavior amongst women or other factors like the duration of risky behavior among men lasting years. In this way, ABMs have may assist in understanding social behaviors and the impact on disease acquisition among select groups in a specialized setting.

ABMs can also help to determine which interventions may be the most effective, singularly and in combination, in stopping and reducing HIV morbidity and mortality. Marshall et al. 2012 demonstrated that among injecting and non-injecting drug users, combinatory interventions were needed to reduce and stop HIV acquisition and transmission. Particularly, the ABM simulated the probability of events occurring such as syringe exchange programs, treatment facilities, and being on ART using statistical outcomes based on previous published studies detailing observational studies and intervention programs. Brookmeyer et al. 2014 introduced the notion of HIV prevention packages or combinatory intervention that are believed to increase the effect of only one intervention being used.[123] In particular, the use of an ABM assisted in understanding the effect size (i.e., such as, levels of coverage, acceptance, and adherence) needed to benefit from combinatory intervention such as HIV prevention packages such as the use of contraceptive methods specifically preventing unprotected anal intercourse and pre-exposure prophylaxis (PrEP).[123]

Summary of Deterministic, Stochastic, and Agent-Based Models.

In summary, deterministic (compartmental) models have fixed transition rates for input parameters representing average rates of transmission or disease onset, for example. While stochastic models have probabilistic transition rates for each input parameter

where a probability distribution may be used to capture age range, rates of transmission, etc. Agent-based modeling is useful for tracking individual characteristics, while demonstrating how interactions and other modes of contact influence epidemiological outcomes. Thus, using the information gathered on the types of models used to predict infectious disease outcomes, stochastic, agent-based models will be used to help answer the main question of this thesis work (described in Chapter 1), which states,
"What strategies and conditions may assist in predicting, reducing, and preventing HIV transmission and HIV drug resistance, within-host, and between-host?"

In addition, each of the model types discussed previously have characteristics that are essential to building within-host and between-host models for predicting HIV transmission and HIV-DR. Next, I explore the basic modeling framework and model interpretation of within-host and between host modeling in reference to HIV-DR.

2.5 Within-Host versus Between Host Models and HIV

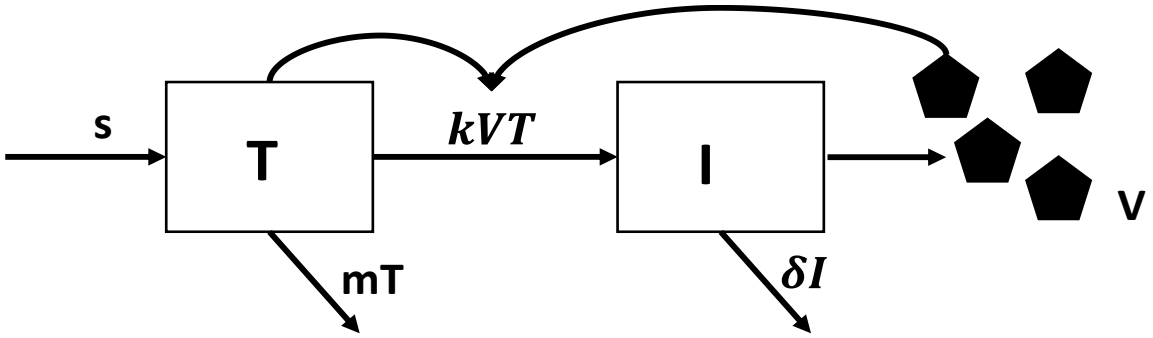
Within-Host Modelling. Within-host models are essential to understanding disease progression and the evolution of HIV viral strains [110,124]. The earliest models of HIV transmission included few parameters that described within-host HIV mechanisms. These within-host (immunological) models had three major input parameters/elements: uninfected target cells or CD4+ T cells (T), infected cells (I) and free virus (V)[124]. These elements outlined the basic, within-host, TIV model. The relationship between these parameters were typically described in the form of differential equations where cells were created (or birthed), infected, and died in the model at certain transition rates (see Table 2; Equation 1). The advantage of the within-host TIV model is that the model makes adding new variables, related to observed and pharmaceutical advancement possible. As combination ART became readily available in 1997[125], within-host HIV models evolved and included parameters such as drug susceptibility, drug dosing, other drug properties, adherence, and cell types such as latently infected cells.[110] Thus, the advantage of having accessible laboratory tests and results such as viral load measurements and immune responses is that scientists and researchers can enhance and validate the within-host models. In addition, many studies have been able to establish outcomes such as basic reproduction numbers for HIV, as well as identify

other tricellular and intercellular events that may, for example, lead to co-infections or other undesirable events such as HIV-DR. However, within-host models are limited in scope as these models only simulate a single individual's condition and thus, have solutions of higher specificity for a single individual and may not provide the necessary interventions needed to stop HIV transmission from happening within a population of HIV positive individuals with a different set of host and pathogen attributes. To understand the movement (i.e., contact network) and the transmission of HIV in a population over time, or the epidemiology of disease, I will briefly explain between-host models and their impact on the field of infectious disease and the HIV epidemic.

Table 2.1: Within-host HIV model parameters.

Parameter Symbol	Description
S	the rate at which the T-cells are being produced
T	Target cells
I	Infected cells
V	Free virus
K	Infection rate constant
c	Clearance rate
M	Target cell death rate
mT	Death of target cells
δ	Death rate
δI	Death of infected cell
pI	Number of virions produced per day; p is the number per infected cells per day

Figure 2.3: Basic Structure of the within-host model.



Equation 2.1: Differential equations for a simple within-host HIV model

$$\text{Number of Target Cells} = \frac{dT}{dt} = s - mT - kVT$$

$$\text{Number of Infected Cells} = \frac{dI}{dt} = kVT - \delta I$$

$$\text{Number of Virons} = \frac{dI}{dt} = pI - cV$$

Between-Host Modelling. Between-host (epidemiologic) modeling is a very common method used in infectious disease to describe population-based changes for a disease outbreak. Between-host HIV models include individuals in a represented population are assigned into infection states such as susceptible or infected states. These infection states help to provide the basis of the susceptible-infected (SI) model, commonly used in modeling HIV outbreaks. In a between-host SI model, individuals are birthed into a model (classified as either susceptible or infected), and if infected, the individual will remain infected unless affected by an intervention (i.e., treatment, change in behavior, and/or death), which are assumptions carried out via deterministic or stochastic modeling, respectively. That is, most between host models have been deterministic historically. However, in recent years, between-host, stochastic modeling has been advantageous in being able to understand the ties and link between individuals and individuals from sub-groups (e.g., intravenous drug users, men who have sex with men, people of African descent, etc.), where the HIV epidemic is most prevalent.[113] Again, SI models are typically described using differential equations (see Table 2.2; Equation 2.2). The major parameters/elements for a SI model include time (t), susceptible individuals $S(t)$ at time t, infected individuals $I(t)$ at time t, contact rate per unit time (c), transmission probability (ρ), total number of persons in the population at time t,

N(t)=S(t)+I(t), and death (μ).[113] In addition, with HIV model includes other parameters such as the effects of treatment on transmission probability, in this case.

Table 2.3: Between-host SI model parameters.

Parameter Symbol	Description
T	Time
b	Number of births
S(t)	Susceptible individuals at time t
I (t)	Infected individuals at time t
c	contact rate per unit time
ρ	Transmission probability
N(t)	total number of persons in the population at time t, or S(t)+I(t)
μ	Death rate
μS	Number of deaths of susceptible individuals per unit time
μI	Number of deaths of infected individuals per unit time
λ(t)	Force of infection or the rate a susceptible person becomes infected per unit time t

Equation 2.2: Differential equations for a simple, between-host SI model.

Rate of change of the number of susceptibles at time t,

$$\frac{dS(t)}{dt} = b - \mu S(t) - S(t) * c * \frac{I(t)}{N(t)} * \rho$$

Rate of change for the number of infected at time t,

$$\frac{dI(t)}{dt} = S(t) * c * \frac{I(t)}{N(t)} * \rho - \mu I(t)$$

Prevalence = $\frac{I(t)}{N(t)}$

Incidence rate (force of infection) = λ(t) = $\frac{c\rho * I(t)}{N(t)}$

New Infections = $S(t) * c * \frac{I(t)}{N(t)} * \rho$ = λ*S(t)

Multi-scale Modeling. Although within-host and between host modeling have contributed greatly to the understanding of host-pathogen interactions – including biological, and social and behavioral dynamics, respectively, the singular effect of each model may not be enough to gain a full picture needed to end the HIV epidemic and other infectious disease outbreaks. Thus, the use of multiscale modeling has become increasingly popular to gain insight on how intracellular and intercellular characteristics of individuals in a population may affect population dynamics, and vice versa. In principle, multiscale modeling refers to using different scales including time and space to describe a system. Garira 2020 states that multiscale modeling serves to understand complex systems such as in infectious disease.[126] However, many multiscale models in the field of infectious disease may not have standard approaches due to the inclusivity of other forms of science that have impact on the organization of the model. In this way, multiscale models like any model may not be all-inclusive and exhaustive of the real-world scenario or condition. However, the goal of multiscale models is to form a more holistic approach for the condition, scenario, or event that the model represents. Garira 2020 describes seven level of organization that an infectious disease system may include in a multiscale model.[126] These levels of organization include: (1) cellular, (2) tissue, (3) organs, (4) the microecosystem, (5) the host/organism, (6) the community level, and (7) the macro-ecosystem.[126]

Although, organizing a multiscale model with seven level of organization is achievable, having all seven levels of organization in a model may not be necessary, depending on the disease/illness of interest. In the case of creating a multiscale model of HIV, the within-host model dynamics can be captured using cellular, and tissue level attributes of the host and pathogen, and the between-host characteristics may include information for the host/organism and at community level to provide sufficient information to predict infectious disease outcomes. In this way, the cellular and tissue levels represent cells such as the CD4+ T cells, macrophages, and lymphocytes, and are compartmentalized to demonstrate how the rates of birth, deaths and change may occur moving inside the tissues. Branching into the host-pathogen relationship, these intracellular events are linked to the function of the HIV pathogen and the viral dynamics associated with host attributes such as ART adherence and the use of contraceptives. Finally, an HIV multiscale model further gathers information of each

individual to summarize the average counts, probabilities, and rates necessary for making epidemiological calculations of prevalence, incidence, risk, and mortality.

A study by Martcheva and Li 2013 investigated the effect of HIV superinfections on epidemiological outcomes using partial differential equations.[127] The outcomes of interest included the rates of transmission and mortality of individuals using the viral load levels within host to categorize individuals into differing stages of HIV progression.[127] Thus, the interesting finding that was not captured by most single within-host or between host models was the non-oscillation of HIV prevalence in the population for superinfections which was not a finding previously seen in other studies. Sun et al. 2016 found that through using a multiscale model, early initiation of ART may cause HIV transmission to decrease in the population and increase the presence of drug resistance, and thus may still lead to fast increase in HIV prevalence overall.[128] These findings by Sun et al. 2016 were extracted from using the viral load of each partner to calculate the transmission rate.[128] Another example of multiscale modeling is in the case of understanding the evolution of set point viral load. Cuadros and Garcia Ramos 2012 demonstrated that the number of co-infections in the population may increase as the set point viral load increases.[129] The co-infection has an impact on the replicative capacity of HIV using ordinary differential equations to describe the transmission rate between host coupled with the tracking of viral load of each host.[129] Thus, without the use of multiscale models, findings and interventions needed to understand linked phenomena may not be possible.

2.6 Summary

Modeling HIV transmission and HIV-DR is a complex issue due to the nature in which an individual (host) becomes infected. The use of mathematical and epidemic modeling of HIV and HIV-DR requires an understanding of the model type needed (i.e., deterministic, stochastic, agent-based) to answer the question(s) of interest, as well as a basic understanding of the characteristics at the level of organization (i.e., within-host, between-host, or multiscale) needed to predict model outcomes. Each model type and scale of modeling has advantages and disadvantages as described in various studies investigating HIV transmission and HIV-DR. However, the multiscale modeling

approach has been shown to be particularly beneficial in understanding gaps in knowledge that may be dependent, feedback loops of host-pathogen interaction and transmission. Next, I will discuss how Chapter 2 (Background) relates to Chapter 1 (Introduction) and the subsequent chapters of this thesis work.

2.7 Relationship of Dissertation Research Questions to Modeling

In Chapter 1, I introduce infectious disease modeling, discuss the global threat of antimicrobial resistance specifically HIV-DR, the antiretroviral medical used to combat HIVDR, and several tools such as the Stanford HIV drug resistance database that have been developed to track and identify HIV-DR. Although mathematical modeling for infectious disease is introduced in Chapter 1, I provide more detail of the type of infectious disease models in Chapter 2 such as deterministic, stochastic, and agent-based models and determine that a stochastic, agent-based model may provide the most accurate predictions for my thesis work examining within- and between host dynamics for HIV. I also discuss the threat of HIV-DR in preventing the goals of the UNAIDS 90-90-90 plan of end the HIV/AIDS epidemic and the types of HIV-DR (i.e., transmitted and acquired drug resistance) that contribute to developing HIV-DR. Together, this information from Chapters 1 and 2, help to provide part of the framework needed to answer the main question of this thesis which is, "What strategies and conditions may assist in predicting, reducing, and preventing HIV transmission and HIV drug resistance, within-host and between-host (i.e., in a multi-level system)?". Thus, the questions for each thesis aim include:

<u>Chapter 3 (Aim 1 – investigating acquired drug resistance with using within-host dynamics)</u>

As described in Chapter 1, Chapter 3 (Aim 1) focuses on the development of a model that predicts the emergence of acquired drug resistance within-host. The objective of Aim 1 is to,

> "Model the effect of host genetic variation on the concentration of HIV-DR mutations, within-host, by modeling "patient-specific" pharmacokinetic properties of antiretroviral medication, and drug adherence."

For Aim 1, I specifically explore whether being non-adherent to an HIV treatment leads to greater risks of developing HIVDR than having a higher drug decay rate, and vice versa. In addition, given a specified drug adherence level and host genetic variant using mathematical modeling:

- When do HIV drug resistance mutations emerge?
- What proportion of acquired drug resistance (ADR) accumulates within-host?
- How does the drug concentration change over-time?
- What specific HIV mutations emerge in the viral population, within-host?

Chapter 4 (Aim 2 – investigating acquired versus transmitted drug resistance using between-host dynamics)

As described in Chapter 1, Chapter 4 (Aim 2) investigates the presence of drug resistance (i.e., both acquired and transmitted or pre-treatment drug resistance) in the population. The objective of Aim 2 is to,

> "Model the effect of host genetics and prevalent drug resistance mutations on HIV-DR levels in two sub-Saharan African populations."

For Aim 2, I identify the populations that may be more likely to develop drug resistance, given the frequency of fast or slow metabolizers in the model, or the presence of individuals with a K103N mutation. In addition, given that we have frequency data for a drug resistant mutation and pharmacogenomic data related to types of metabolizers (i.e., fast, intermediate, and slow) from two distinct study populations in sub-Saharan Africa can we use mathematical modeling to answer the following:

- What are the differences in the proportion of drug resistance in the population when model conditions include or do not include DRM or PGx frequency data?
- Which model conditions yields the highest levels of drug resistance?

- What proportions or percentages of transmitted or acquired HIV-DR contribute to the proportion of drug resistance in the population?
- Are there other factors that contribute to the proportion of HIV-DR in the population?

Chapter 5 (Aim 3 – determining the best method and conditions for optimizing HIV-DR levels)

Chapter 5 (Aim 3) tackles finding the optimal values for the model that will achieve a reduction of drug resistance in the population and the prioritization of individuals for second-line treatment. The overall objective of Aim 3 is to,

> "Create an ART optimization routine for a stochastic, network-based model that will modify drug dosing and/or switch treatment regimens, given the patient's characteristics and the threshold level of a desired HIV outcome."

For Aim 3, I determine which patients to prioritize for second -line therapy, given the availability of resources under specific cost and conditions. In addition, HIV-DR levels in specific populations have risen above the WHO recommended level of 10% (as described in Chapters 2 and 4) and could stop efforts (i.e., like switching to second-line drug treatment) made to reduce HIV transmission and HIV-DR. Thus, I used mathematical modeling to determine:

- Which optimization algorithm for treatment switching is best to use for finding policies to prioritize infected individuals, in the population, for second-line drug treatment?
- What parameter conditions of set point viral load (SPVL) and adherence levels yield the lowest levels of HIV-DR in the population when second-line treatment is either expensive or inexpensive?

As described in Chapter 1, for my thesis work, I will include the use of the WHO clinical recommendations for treating HIV in 2016. For each aim of my thesis, utilize the modeling approaches described in Chapter 2, a stochastic, agent-based HIV epidemic modeling system called *EvoNet*. Thus, in the next chapter, Chapter 3 (aim 1), the goal is to predict the emergence and proportion of acquired drug resistant mutations, within-

host, and other outcomes such as viral load, drug concentration, and the other viral mutant strains that makeup single, double, triple, quadruple, or quintuple HIV mutations. In aim 1, I also incorporate pharmacogenomic data (i.e., the presence of host genetic variants – fast, intermediate, and slow metabolizers) and pharmacokinetic data (i.e., drug decay), for varying drug adherence levels, within-host. In all, the next chapter will involve introducing the field of pharmacogenomics and the investigation of within-host dynamics related to acquired drug resistance. This begins to answer the question of "What strategies and conditions may assist in predicting, reducing, and preventing HIV transmission and HIV drug resistance within-host and between-host (i.e., in a multi-level system)?"

Chapter 3: Modeling Within-Host HIV Drug Resistance using Pharmacogenomic Data

3.1 Introduction

Antiretroviral therapy (ART) has proven to be significantly effective in reducing and preventing HIV transmission, morbidity, and mortality, especially in low-to-middle income countries and resource-limited settings[40,130–133]. Common antiretroviral regimens in resource-limited and low-to-middle income countries have largely included prescribing medications from the non-nucleoside reverse transcriptase (NNRTI) drug class. Certain medications within the NNRTI drug class have been known to have deleterious side effects for patients and can lead to the emergence of drug resistance mutations due to the low barrier to drug resistance [134]. Despite new clinical guidelines that recommend taking newer INSTI-based antiretroviral medications that have a high barrier to drug resistance, in many low to middle income countries (LMICs), NNRTI-based drug regimens are still being used. However, NNRTI-based drug regimens (that as of 2021 clinical guidelines are currently recommended in" special circumstances") have been known to have the following side effects: neuropsychiatric effects (like insomnia, changes in mood, and vivid dreams), suicidal ideation, weight gain (i.e., obesity), and the non-tolerance of medications during pregnancy[135,136].

When a patient experiences virological failure while on the NNRTI-based regimen, it is also common to switch the patient to a second-line treatment plan. Efforts by researchers and clinicians to understand the contributions to virological failure and/or drug resistant mutations that form when patients have been prescribed NNRTI-based regimens have included investigating host genetic properties, or the pharmacogenetic profile. The field of pharmacogenomics has helped clinicians and pharmacologists better understand how effective a medication may be for the patient in the presence of other cellular and biological properties. However, few studies have

simulated the effect of the presence of known pharmacogenomic data on the drug resistance developed at the individual level or within-host. Therefore, the aim of this chapter is to predict the emergence and proportion of acquired drug resistant mutations, within-host, and other outcomes such as viral load, drug concentration, and the emergence of specific viral mutants, given the presence of host genetic variants and varying drug decay levels.

In this chapter, I first provide an overview of the fields of pharmacogenomics and pharmacokinetics. Then, I outline the research aims, hypothesis, approach, methodology and simulation results that incorporate pharmacogenomic data. The focus of this chapter is Aim 1 which is to model the effect of host genetic variation on the concentration of HIV-DR mutations, within-host, by modeling "patient-specific" pharmacokinetic properties of antiretroviral medication, and drug adherence within the context of the broader question of "What strategies and conditions may assist in predicting, reducing, and preventing HIV transmission and HIV drug resistance within-host and between-host (i.e., in a multi-level system)?" Thus, for Aim 1, I explore whether being non-adherent to an HIV treatment leads to greater risks of developing HIVDR than having a higher drug decay rate.

3.2 Background

An Introduction to Pharmacogenomics. The field of pharmacogenomics help

can improve, prevent, and sometimes cure adverse health conditions that allopathic or modern medicine has sought to accomplish[137]. Simplistically, some approaches to allopathic medicine include, but are not limited to, using drugs, radiation, and/or surgery to treat patients. However, this field of medicine no longer uses the one-size fits all approach, where patients that exhibit a clinical phenotypic such as observable attributes of a disease may have the same treatment based on the similar clinical phenotype and symptoms[137]. Researchers have found that the differences between individuals, even with similar diagnosed diseases, happen on a bio-cellular level and can directly influence the way that the body responds to a drug treatment, radiation, and surgery. Often patients have side effects or adverse health events from taking a

medication that alerts the healthcare provider and result in the discontinuation of the medication or switching treatment altogether. For emergency conditions, the traditional allopathic medical approach help to alleviate a health crisis and is the basis for routine health care and treatment. However, in situations where a diagnosis in known and verified among several patients, but a prescribed treatment works for one person and not the other, a more personalized approach may be needed. This personalized approach may help prevent an adverse health response, or a patient succumbing to the "trial and error" of treatment regimens and can function as a logical path based on their personal genetic history? This is what the sub-field of pharmacogenomics provides: the ability to map and predict the genotypic-to-phenotypic response associated with a drug treatment. Pharmacogenomics (PGx) is the study of how host genes affect drug response[138,139]. Specifically, PGx details the impact of host genetic variants on drug pharmacokinetics, or the way the drug moves throughout the body[138].

Specific medicines have been found to have divergent effects. For instance, the medicine, warfarin, is most notably used as an anticoagulant (blood thinner) and antithrombotic that reduces and can prevent blood clots and other factors leading to fatal cardiovascular events. However, warfarin has been found to have high interindividual variability, leading to patients falling outside of the therapeutic range. This is primarily due to the host genotypes of the following cytochrome P 450 genes, CYP2C9 and VKORC1[140,141]. The CYP2C9 gene is responsible for the process of drug metabolization. VKORC1 gene encodes instructions for making the VKORC1 enzyme that helps to activate proteins that help form blood clots and as such, warfarin is used to bind to the enzyme to inhibit the blood clotting formation pathway[140,141]. When warfarin resistance occurs, for example, due to a VKORC1 genotype in a patient, warfarin has reduced binding to the VORC1 enzyme, and more drug is needed to prevent blood clots and hemorrhaging. In summary, certain host genetic variants or genotypes can impair enzyme (either by induction or inhibition) leading to the drug concentration increasing or decreasing in the body, resulting in having a supra-therapeutic or sub-therapeutic effect, respectively[142]. Consequently, similar pharmacogenomic effects can be found with HIV medications prescribed to individuals that have varying rates of drug metabolism, leading to treatment failure and drug resistance[49,143–145].

The cytochrome p450 (CYP*) enzymes are responsible for ART metabolism and excretion. Several studies have correlated antiretroviral drug resistance and virolgic failure with the presence of host genetic variation related to CYP* alleles [146–150]. One of the first pharmacogenomic tests approved for clinical use was for testing for the presence or absence of the HLA-B*5701 allele. The HLA-B*5701 allele is associated hypersensivity in individuals who take the antiretroviral, abacavir[151]. Since then, other genotype tests are availalble for testing drug resistance and adverse reactions to antiretrovirals[152,153]. However, most laboratory tests include viral genotypes rather than host genotypes such as the CYP* alleles. Laboratories commonly have tests available for detecting genetic mutations on the reverse transcriptase, protease, and integrase[152,154]. Specifically, these genotype test are usually ordered when and if HIV RNA levels reach levels above 500 copies/mL[152,154]. In the United States, laboratories have guidelines related to when these tests should used related to have detectable levels of HIV, if experiencing virological failure while on first- or second line treatment [152].

I reviewed literature from the field of pharmacogenomics and HIV/AIDS. The studies I found have been documented in Table 1. Specifically, the review answered the following questions: (1) how many studies have included both pharmacogenomic information related to antiretrovirals? (2) what is the frequency of host alleles associated with HIV and antiretroviral medication? and (3) what interventions exist in preventing adverse health effects of host alleles presenting sub- or supra-therapeutic drug effects? Using the databases PubMed/Medline, I searched for the following MESH terms during the years, 2000-2019: ((pharmacogenomics) AND (hiv)) AND (genes). The exclusion criteria included titles and abstracts that did not include information on first-line drug regimens including efavirenz, tenofovir, and/or lamivudine. Journal articles, conference papers, clinical trials, meta-analyses, random controlled trials, reviews, and systematic reviews were included in the search and other sources of literature were excluded. The initial search generated 184 articles from the MEDLINE/PubMed search.

Figure 3.1: The count of articles and other documentation (i.e., conferences, clinical trials, etc.) related to both pharmacogenomics and HIV/AIDS research from 2000-2019

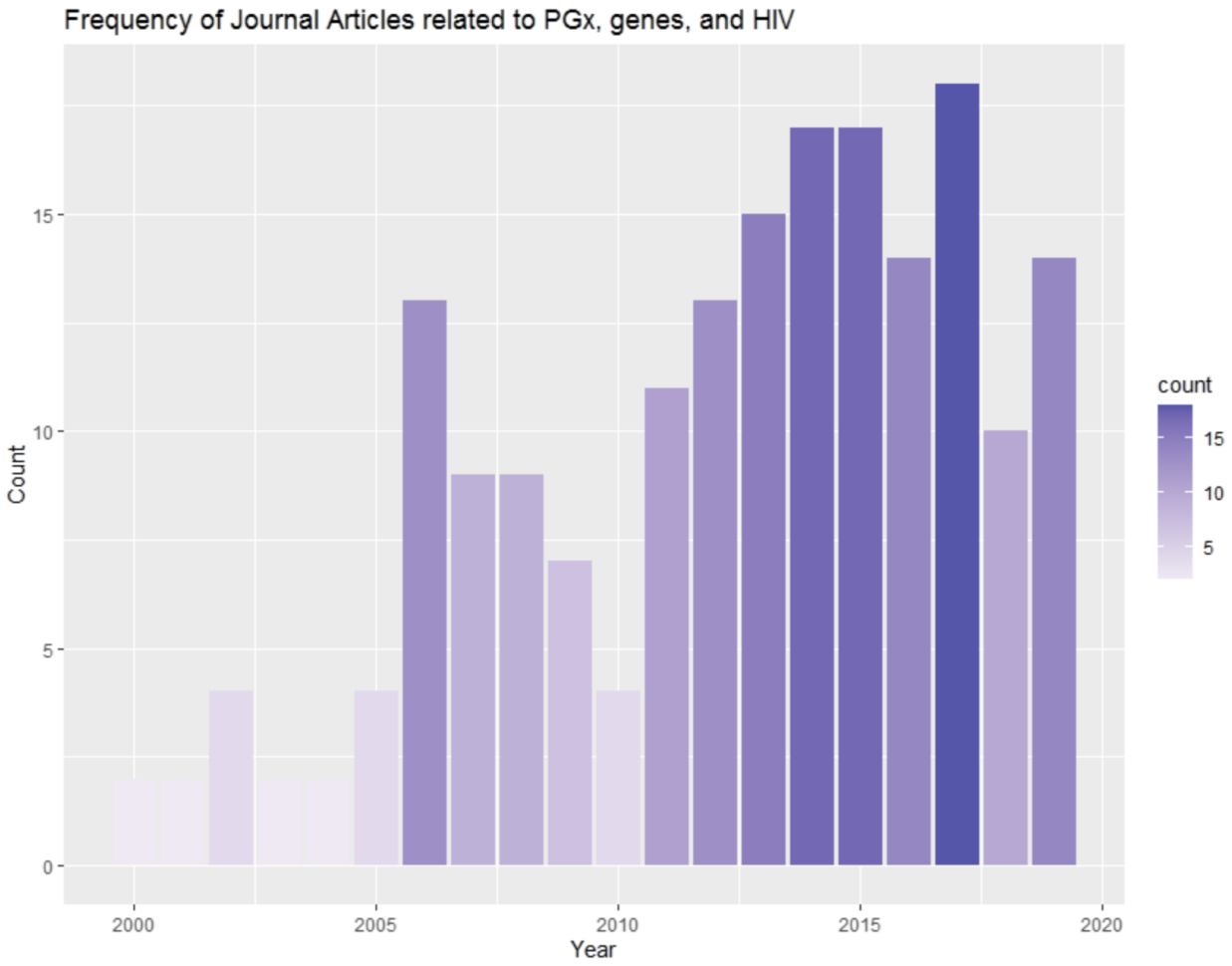


Table 3.1: Pharmacogenomics studies related to pharmacokinetic and HIV-DR outcomes.

Author	Target Population	Associated HIV Class	Genetic Variant(s) studied	Study Aim(s)	Study Outcomes
Borghetti et al. 2019[146]	HIV positive individuals experiencing psychologic symptoms	Integrase Inhibitors (i.e., dolutegravir)	SLC22A2 variants	Investigate the effect of dolutegravir trough concentrations and/or variants of the SLC22A2 gene on experiencing neuropsychiatric symptoms (NP).	OCT-encoding gene variants, along with the trough concentrations, or the concentrations of a drug reached before the administration of the next dose, was associated with neuropsychiatric symptoms with individuals taking dolutegravir-based regimens.
Lakhman et al. 2009[147]	HIV infected individuals with the CYP3A variants	PIs, NNRTIs	CYP3A variants	To summarize the clinical and pharmacologic relationship between CYP3A and antiretroviral drugs.	Function of CYP3A; pharmacokinetic outcomes (i.e., drug clearance, toxicity)
Ngaimisi et al. 2013[155]	ART naïve and HIV positive patients in Ethiopia and Tanzania	NRTIs, NNRTIs	CYP2B6, CYP3A5, ABCB1, UGT2B7 and SLCO1B1 variants	Determine plasma and intracellular efavirenz concentrations and allele frequencies among the study population.	Viral load and CD4 count at 12, 24, and 48 weeks; Drug plasma concentrations; Both genotype and allele frequencies were obtained signifying the differences between the populations of Ethiopia and Tanzania.
Nyakutira et al. 2008[150]	HIV/AIDS patients in Zimbabwe	NNRTIs	CYP2B6*6	To use pharmacokinetic (PK) modeling to understand the relationship between efavirenz and the CYP2B6 516G-->T(*6) genotype in the population.	Efavirenz plasma concentrations; Dose reduction estimates for achieving therapeutic range.

Table 3.2: HIV medication pharmacogenomic information found on the pharmacogenomic (PharmGKB) database, https://www.pharmgkb.org/.

Drug Name	Abbr.	1st-line Regimen	2nd -line Regimen	PGx Info.	Gene
Efavirenz	EFV	1	2	Yes	CYP2B6, CYP2A6
Nevirapine	NVP	1	2	Yes	CYP2B6, HLA-B, ABCB1
Lamivudine	3TC	1	2	Yes	SLC22A2, ABCB1
Emtricitabine	FTC	1	2	No	
Tenofovir	TDF	1	2	Yes	ABCC2, ABCC4, ABCC10, OCRL
Zidovudine	AZT	1	2	Yes	ABCB1
Lopinavir/ritonavir	LPV/r	---	2	Yes	ABCB1
Didanosine	ddI	---	2	Yes	NT5C2
Zalcitabine	ddC	---		No	
Atazanavir	ATV	---	2	Yes	UGT1A1, CYP3A5
Ritonavir	RTV	---	2	Yes	UGT1A1, CYP3A5, ABCB1

Pharmacokinetic Properties. Pharmacokinetics (PK) is the study of how a drug is absorbed, distributed, metabolized, and excreted (ADME). Drug metabolism or the breakdown of a drug plays a major role in determining the optimal drug concentrations for a phenotypic response. One of the most important parameters for determining optimal dosing of the drug is the inhibitory concentration (IC). The inhibition concentration helps researcher determine how susceptible the virus is to the drug administered. However, in the presence of a mutant virus the inhibition concentration there may be modification to the level of susceptibility. Measuring the inhibitory concentration 50 (IC50) identifies the minimum concentration need to inhibit 50% of a biological event. Thus, fluctuations in the value of the IC50 parameter, may detail how much dosing is needed to effectively treat a patient. The IC50 is calculated using a ratio of final to the initial value of the amount of drug needed to inhibit the biological process by half. Thus, in the presence of a mutant virus, understanding how the IC50 values change will greatly affect the clinical effectiveness of the drug. When HIV mutates,

knowing the IC50 fold changes impacts whether a drug will inhibit viral replication at an achievable concentration in vivo, and can forecast or model treatment failure and/or the propensity of more mutations forming overtime. Table 3 details the IC50 fold changes for the first-line or common antiretroviral medication and associated mutations.

Other pharmacokinetic parameters that have been used to explain the elimination and excretion pharmacokinetic stages include the area under the curve, half-life, and clearance rate[156]. The area under the curve or the area under the plasma concentration curve is a measure of the total systemic exposure of the drug in the body and is inversely proportion to drug clearance[156]. That is the higher the AUC, the lower the clearance and the higher the concentration of the drug in the systemic circulation and vice versa. The AUC parameter helps to create a plasma-concentration-time profile that also shows the concentration maximum, and the half-life. The maximum concentration (C_{max}) reached after the drug is administered[156]. The half-life ($t_{1/2}$) of the drug determines the time required for the concentration of the drug to reduce by half[156]. Another pharmacokinetic parameter that contributes to measuring the amount of drug in the body is the volume of distribution and propensity of the drug to migrate to other tissue compartments. The volume of distribution (V_D) is measured as a proportionality constant of the amount of drug in the body to the plasma concentration at a specific timepoint[156,157]. Thus, if the V_D is high, the drug is distributed in other tissues and less in the plasma, and a higher dosage is needed. However, this is dependent of which tissues the drug is targeting. These parameters help to form pharmacokinetic models where the primary goals are to: predict drug concentration levels in the body before administering the drug. In addition, some of the other outcomes of pharmacokinetic modeling include new drug development, risk assessments, and targeted therapy/dosing. In relation to antiretroviral medication, being able to connect adverse HIV clinical outcomes to pharmacokinetic and pharmacogenomic data may aid in a more detailed approach to preventing transmission and creating a more personalized medical approach to treating individuals living with HIV. The next sections describe the approach and methodologies used to incorporate pharmacokinetic, pharmacogenomic, epidemic and within-host modeling to predict the evolution of HIV and HIVDR mutations, within-host.

Table 3.3: Common HIV-DR mutations for first-line ART and corresponding fold-changes to the antiretroviral IC50..[158]

First-line antiretrovirals	Common Mutations	Inhibitory Concentration 50 (IC50) Fold-Change
Tenofovir (TDF)		
	K65R	Reduces susceptibility by 2-fold.
Lamivudine (3TC)		
	M184V/I	Reduces susceptibility by >100-fold.
	K65R	Reduces susceptibility by 5-to-10 fold.
Emtricitabine (FTC)		
	M184V/I	Reduces susceptibility by >100-fold.
	K65R	Reduces susceptibility by 5-to-10 fold.
Efavirenz (EFV)		
	L100I	Reduces IC50 by 10-fold, alone and >50 fold when interacting with the K103N mutation.
	K101E	Reduces IC50 by 1-to-5 fold
	K101P	Reduces IC50 by >50 fold.
	K103N	Reduces IC50 by 20-fold.
	V106A	Reduces IC50 by 5-fold.
	V106M	Reduces IC50 by >30 fold.
	Y181C	Reduces IC50 by 2-fold.
	Y188L	Reduces IC50 by >50 fold.
	Y188C	Reduces IC50 by 20-fold.
	Y188H	Reduces IC50 by 5-fold.
	G190A	Reduces IC50 by 5-to-10 fold.
	G190S	Reduces IC50 by >50 fold.
Nevirapine (NVP)		
	L100I	Reduces IC50 by 5-fold, alone and >50 fold when interacting with the K103N mutation.

	K101E	Reduces IC50 by 3-to-10 fold
	K101P	Reduces IC50 by >50 fold.
	K103N	Reduces IC50 by 50-fold.
	V106A	Reduces IC50 by 50-fold.
	V106M	Reduces IC50 by >30 fold.
	Y181C	Reduces IC50 by 50-fold.
	Y181I/V	Reduces IC50 by >50 fold.
	Y188L/C	Reduces IC50 by >50 fold.
	Y188H	Reduces IC50 by 10-fold.
	G190A/S	Reduces IC50 by >50 fold.
Dolutegravir (DTG)		
	G118R	Reduces the IC50, or susceptibility five-fold
	R263K	Reduces the IC50, or susceptibility two-fold
	V151I/L/A	Reduces the IC50, or susceptibility two- to three-fold
	S153Y/F	Reduces the IC50, or susceptibility two- to three-fold
	S230R	Reduces the IC50, or susceptibility three-fold
	T66K	Reduces the IC50, or susceptibility two- to three-fold
	E92Q	Reduced the IC50, or susceptibility 1.5-fold.
	E138K/A/T	Reduces the IC50, or susceptibility ~ 10-fold when paired with raltegravir (RAL).
	G140S	Reduces the IC50, or susceptibility > 10-fold when interacting with the mutation, G148H/R/K.
	Q148H/R/K	Reduces the IC50, or susceptibility >10 fold when interacting with the G148H/R/K and G140S mutation.

3.3 Approach

Stochastic, agent-based modeling system: EVONET.

The epidemic modeling platform used to simulate our project aims is called *EvoNetHIV* (*https://github.com/EvoNetHIV*). EvoNet development was funded by the National

Institute of Health (NIH) as an R01 grant (R01AI108490, MPI: Herbeck, Goodreau, Mittler). EvoNet utilizes the RStudio package suite, statnet and the epidemic modeling package, EpiModel, to perform computer simulations of an HIV epidemic . The platform was developed using R and C programming languages, and the code auditing software Github found in RStudio. EvoNet has 18 modules of code that simulate: (1) the initiation of the epidemic (i.e., births), (2) contact/sexual network, (3) transmission, and (4) disease progression. Approximately 175 model parameters have been included in the modeling platform based on empirical data representing the socio-behavioral (between-host), and viral dynamic (i.e., within-host) characteristics. The collection of pharmacokinetic data is in progress and currently includes 48 default parameter values from approximately 28 studies.

Within-Host Model: We utilized the basic model structure of a within-host model (see Chapter 2, section 2.5) to simulate the viral reproduction and growth HIV short, intermediate, and long-lived cells[159]. The relationship between uninfected and infected cells is most commonly seen as compartments with differential equations explaining the increase or decrease of cells over time. The EvoNet within-host model builds upon this basic structure to incorporate treatment changes, drug resistance mutations, pharmacokinetic properties, pharmacogenomic dynamics. The next section explains how each parameter has been added to the simple-within host model to show HIV viral dynamics of an infected person with pharmacogenomic changes.

Equation 1: Differential equations for a simple, within-host HIV model

Target Cells

(1) $\frac{dT}{dt} = s - mT - kVT$

(Host) Infected Cells

(2) $\frac{dI}{dt} = kVT - \delta I$

Virions (infected form of the virus, outside the host cell)

(3) $\frac{dI}{dt} = pI - cV$

3.4 Methods

Data Collection. We utilized several information sources to gather data on pharmacogenomic, pharmacokinetic, and clinical information required to parameterize the within-host model. We specifically extracted pharmacokinetic parameter values for the first-line ART regimen: tenofovir (TDF), lamivudine (3TC), and efavirenz (EFV). The information sources included the food and drug administration (FDA) antiretroviral drug labels [160], the pharmacogenomic database, PharmGKB[144], the Stanford HIV Drug Resistance database[3,158] and various published literature[132,161–163]. The information gathered from the FDA drug labels were associated with pharmacokinetic properties (see Table 4). The model primarily utilized parameters including, IC50, drug dose, drug decay, elimination rate constants, and fold changes due to the presence of specific mutations. Specifically, Rosenbloom et al. 2012 provided the data needed to determine the drug decay values for each of the first-line medications (see Table 5). The pharmacogenomic data for each of the first-line medications were detailed in Table 6.

Table 3.4. The pharmacokinetic parameters needed to capture the absorption, and distribution phases of a drug life cycle.

Drug	Other Names	Molecular weight (g/mol)	IC50 (µM or µg/mL)	IC50 minimum value (µM or µg/mL)	IC50 maximum value (µM or µg/mL)	Cmax (µmol)	Dose in milligrams (mg)	minimum AUC (µg·hr/mL)	maximum AUC (µg·hr/mL)	minimum Volume of Distribution (L/kg)	maximum Volume of Distribution (L/kg)
TDF	Tenofovir; Viread; Tenofovir DF	287.2	0.0561 µmol [3]	0.04 µM[1]; 0.5µmol/L[2]	8.5 µM[1]; 2.2µmol/L[2]	1.1 µmol	300	1.6	2.98	0.7	1.9
3TC (once daily)	Lamivudine; Epivir	229.3	0.0298 µmol[3]	0.002 µM	15 µM	15.3 µmol	300	5.53	5.53	0.9	1.7
3TC (twice daily)	Lamivudine; Epivir	229.3	0.0298 µmol[3]	0.002 µM	15 µM	15.3 µmol	300	5.53	5.53	0.9	1.7
FTC	emtricitabine; Emtriva	247.2	0.0079 µmol[3]	0.0013 µM	0.64 µM	7.3 µmol	200	6.9	13.1	1.1	1.7
EFV	efavirenz; Sustiva	315.7	0.0035 µmol[3]	.0034 µM	5.5 µM	12.9 µmol	600	35.04	81.13	1.2	1.2

Table 3.5. The half-life (drug decay), elimination rate constant, and clearance rates were captured from FDA medication labels and published studies (Rosenbloom et al. 2012) for the within-host model.

Drug	Half-life (t1/2) in hours	Elimination rate constant, k(e)	Minimum Clearance rate (mL/min)	Maximum Clearance rate (mL/min)
TDF	17 hours; 60 hours	0.041	928	1158
3TC (once daily)	5-7 hours	0.139-0.099	329.4	467.6
3TC (twice daily)	5-7 hours; 10 hours	0.139-0.099	329.4	467.6
FTC	10 hours	0.0113	124	302
EFV	52-76 hours	0.0133-0.009	30	155

Table 3.6. Drug Resistance Mutations and their corresponding fold change values for the model.

Drug	**Major Drug Resistance Mutations on Stanford HIVDR database**	**Fold Change (effect on the IC50 value)**
TDF	**K65R**, K70E, Y115F, M41L, L210W, T215Y, T215F, T69Ins, Q151M	K65R: 2-4-fold decrease in susceptibility
3TC (once daily)	**M184V, M184 I,** K65R	M184VI: 85 - 299-fold decrease in susceptibility (Trial EPV20001); 32- to 53-fold decrease in susceptibility (Trial EPV40001)
3TC (twice daily)	**M184V, M184 I**, K65R	M184VI: 29- 159-fold decrease in susceptibility to lamivudine (Trial EPV20001); 45-fold decrease in susceptibility (Trial EPV40001)
EFV	L100I, K101E, K101P**, K103N,** K103S, V106A, V106M, **Y181C**, Y181I, Y181V, Y188L, Y188C, Y188H, G190A, G190S, G190E, M230L	K103N: 20-fold decrease insusceptibility to efavirenz. Y181C: 2-fold decrease insusceptibility to efavirenz.

* The bolded and underlined mutations are the most frequently observed mutations for the corresponding antiretroviral drug used.

Within-Host Model: Drug Resistant Mutations. We incorporated drug resistance mutations into the model by adding locus positions to representing the presence or absence of a mutation that corresponds to a specified drug. Equation 2 explains how the simple, within-host model has been expanded to incorporate and calculate the number of drug resistance mutations in the model.

(1) Infected cells with drug resistant mutations,

$$\frac{dI_{ijklm}}{dt} = (1 - 5\mu)(1 - f)r_{ijklm}(t)I_{ijklm} - \delta I_{ijklm}$$

$$+ \mu(1 - f) \sum_{1\,Mut} r_{i'j'k'l'm'}(t)I_{i'j'k'l'm'}$$

Where i, j, k, l, m are the five locus positions 1-5. Each loci are either assigned a zero to represent an absent mutation at the locus position or 1, representing a mutation present at the locus position, such that I_{10001} has mutations at locus positions 1 and 5; I_{ijklm} represents the number of infected cells with mutations at any of the five locus positions; μ is the mutation rate; f is the total newly infected cells, r_{ijklm} is the growth rate of the infected cells with mutations at any mutations at locus positions 1-5, at time t; δI is the death rate of the infected cells. The 1Mut summation represents the genetic variants that are only 1 step away from another variant. Thus, the summation for I_{10001} would involve infected cells with mutations at the following locus positions, I_{00011}, I_{11000}, I_{00001}, I_{10000}, I_{10011}, and I_{11001}. The 1 Mut summation assumes that forward and backward mutation rates are the same; otherwise, another summation with a backward mutation rate multiplied by the infected cells. This assumption is true, regardless of any modification related to drug pressure.

Within-Host Model: IC50 parameters. The within-host model also included IC50 parameters. Equation 5 that describe the effect of mutations on IC50 drug concentrations.

(2) Effect of drug resistant mutations on IC50 values of the drug

$$IC50_{d,\,ijklm} = IC50_{d,\,ooooo}\,(1+FC_{d,1}) \times (1+jFC_{d,2}) \times (1+kFC_{d,2}) \times (1+lFC_{d,4}) \times (1+mFC_{d,5})$$

Where drug, *d* is drug 1, 2, or 3; $IC50_d$ is a specific drug's baseline IC50 value assuming for the wildtype virus; $FC_{d,h}$ is the fold-change value for drug d that mutation at a loci position, 1,…,5 on the IC50 value. Other assumptions made in our within-host model include that ART decreases the infected cells growth rate. In addition, fitness costs are incorporated into the model to have a multiplicative relationship with each mutation, effecting the growth rate of the viral strain.

Within-Host Model: Drug Decay. The half-life of the drug (i.e., drug decay) was utilized in the model to describe the relationship between drug concentration and elimination, within-host. Drug decay was primarily utilized due to the extensive research studies conducted on the Efavirenz-based ART regimens. In addition, I extracted and created a table to display associations between pharmacogenomic and pharmacokinetic parameters.

Table 3.7: Pharmacogenomic-Pharmacokinetic Parameter Associations

Drug Name	Abbr.	1st-line Regimen	2nd -line Regimen	PGx Info.	Gene	Variant	Pharmacokinetic property affected
Efavirenz	EFV	1	2	Yes	CYP2B6, CYP2A6	CYP2B6*6 and CYP2B6*18	half-life ($t_{1/2}$), AUC, Cmax
Nevirapine	NVP	1	2	Yes	HLA-B, CYP2B6, ABCB1	HLA-B*5701	AUC, Cmax, Bioavailability (%)
Stavudine	d4T	1?		Yes	TNF	TNF rs1799964	AUC, Cmax
Didanosine	ddI		2	Yes	NT5C2	NT5C2 rs11191561	Clearance rate (Cl/F)
Lamivudine	3TC	1	2	Yes	MRP4, SLC22A2, ABCB1,	MRP4 T4131G	AUC, Cmax
Tenofovir	TDF	1	2	Yes	ABCC4, ABCC2, ABCC10, OCRL	ABCC4 T4131G	AUC, Cmax
Atazanavir	ATV		2	Yes	UGT1A1, CYP3A5	UGT1A1 rs887829	Ctrough, AUC, Cmax
Zidovudine	AZT	1?	2	Yes	ABCC4, ABCB1	ABCC4 rs11568695	AUC, Cmax
Ritonavir	RTV		2	Yes	UGT1A1, CYP3A5, ABCB1	UGT1A1 rs10929303 and UGT1A1 rs1042640	AUC, Cmax

Equation 6 describes a multiplicative relationship between drug concentration and drug decay.

(3) Effect of drug decay on drug concentration in plasma

$$D_{1,\ldots,4} = -h * t_{1/2,\,d} * D_{agent,d}(t),$$

where D is the drug concentration for drugs 1,…,4 in the plasma; h is the step size; $t_{1/2,\,d}$ is the half-life value of the drug d for the agent; and $D_{agent,d}$, is the drug concentration at time t.

Within-Host Model: Genotype-Treatment Logic/Flowchart. Next, we detail the condition and relationship between pharmacogenomic data (i.e., host genotypes) and changes to pharmacokinetic parameters for the within-host model. We utilized Microsoft Visio to create a flow diagram to explain the logic and relationship between the host genotypes, and pharmacokinetic parameters. Within the diagram, the ovals represent beginning or end of the model, the decisions are represented by the diamond-shape, the model condition is represented by a parallelogram, and the proposed outcome is represented by a square or rectangle (see Figure 3.2).

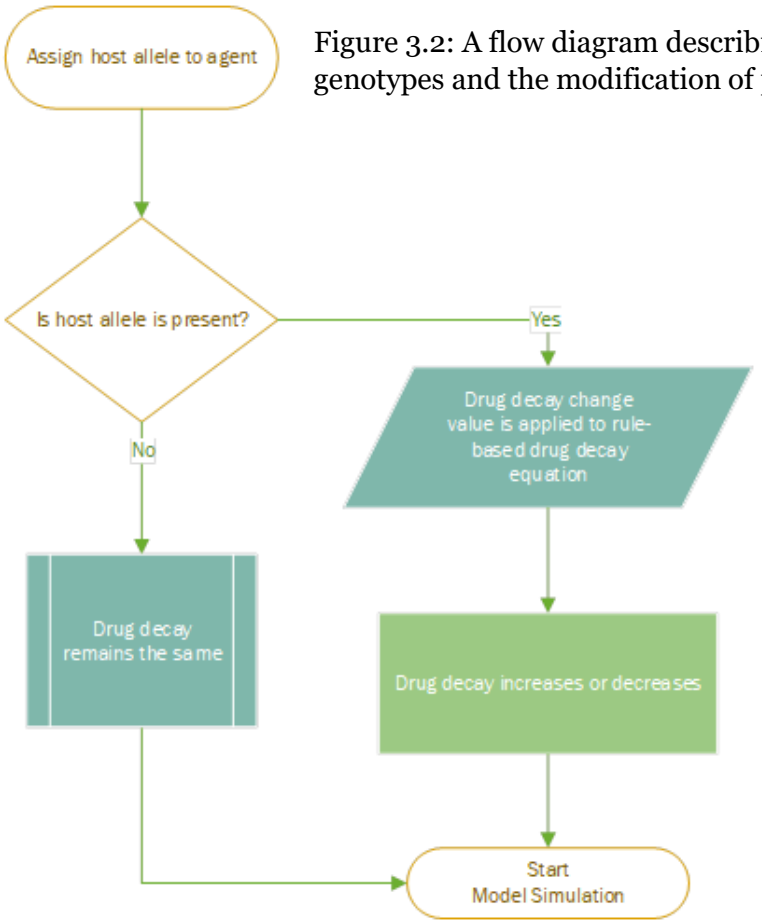


Figure 3.2: A flow diagram describing the relationship between host genotypes and the modification of pharmacokinetic properties.

Within-Host Model: Genotype-Treatment Rule-Based Algorithm in R programming language

A series of rule-based, if-then statements associating host genotype status with pharmacokinetic changes such as drug decay levels using the R programming language. We utilized the R package, *stat* and the rbinom() function to randomly assign agents with binary values, where 0 is no host-genotype and 1 is host-genotype is present. Within the module for drug modifications, the host genotype rule-based statement calculates the following,

```
## Host-Genotype Algorithm
    if(host_genotype==1){
      drug_decay=drug_decay*drug_decay_change
    } else {
      drug_decay=drug_decay
    }
```

Within-Host Model Parameters

All agents were assigned viral dynamic, demographic, and social attributes including age, sex, CD4 count, viral load, and disease status parameters. The input parameters also include network-based, pharmacokinetic and pharmacogenomic data. The following table includes all the parameters by metabolizer type (See Table 8). Each model run assigned each person in the population as fast, intermediate, or slow metabolizers. Thus, each individual in each model run may serve as a representative of a possible outcome. Drug decay/half-life (t1/2) was the main pharmacokinetic parameter that was directly model by the status of the of an individual's metabolizer type. Drug adherence levels were also modified for each run at levels ranging from 0-100%. The percent adherence level corresponded to the daily likelihood, or probability of an individual taking their assigned treatment regimen. Several calculations and decision-based rules were implemented using R and C script/code to estimate within-host outcomes viral load, daily drug concentrations, and quantity of short-, medium-, and long-lived cells. In addition, HIV mutation frequencies were estimated.

Table 3.8: Within-host model input parameters for fast, intermediate, and slow metabolizer using the EvonetHIV modeling system.

Input Parameters	Definitions	Default Values
Model Parameters		
nsims	Total number of simulations (for each metabolizer type – fast, intermediate, and slow)	1
Ncores	Total number of cores	10
Epidemic Parameters		
n_steps	Total duration of the model	300 days
initial_pop	Total population size	2000 agents
initial_infected	Total number of infected agents	1985 agents
poisson_birth_lambda		0.0137*(initial_pop/100)
target_stats	Number of partners	0.7*initial_pop/2
mean_sex_acts_day	Number of sex acts per day	0.0
Host Genetics		
prob_genetic_variant	Proportion of agents with a genetic variant out of the total population [host genetics]	1.00
Treatment parameters		
min_adherence	Minimum level of adherence for Drug1, Drug2, Drug3	0.00
max_adherence	Maximum level of adherence for Drug1, Drug2, Drug3	1.00
prob_eligible_ART	Probability of receiving ART treatment	1.00
start_treatment_campaign	Day to start treatment after infection	Day 100
treatment_threshold	Threshold of viral load copies per mL for treating agents	10
Viral dynamic parameters		
Mu	The forward mutation rate	3e-5
Pharmacokinetic parameters		
BaseIC50Drug1	IC50 for Drug 1 [tenofovir] based on Rosenbloom et al. 2012 and Kuritzkes et al. 1996	120
BaseIC50Drug2	IC50 for Drug 2 [lamivudine] based on Rosenbloom et al. 2012	437
BaseIC50Drug3	IC50 for Drug 3 [efavirenz] based on Rosenbloom et al. 2012	200
drug_decay1	Drug 1 decay rate [for TDF]	0.28
drug_decay2	Drug 2 decay rate [for 3TC]	2.37
drug_decay3	Drug 3 decay rate [for 3TC]	0.46
DrugDose1	The drug dose for drug 1, calculated using the formula:	1100 * (1- exp(-evoparams$drug_decay1))

	Cmax * (1 - exp(decay))	
DrugDose2	The drug dose for drug 2, calculated using the formula: Cmax * (1 - exp(decay))/2	15300 * (1- exp(-evoparams$drug_decay2))/2
DrugDose3	The drug dose for drug 3, calculated using the formula: Cmax * (1 - exp(decay))	12900 * (1- exp(-evoparams$drug_decay3))
Mutation-related parameters		
cost1	Fitness costs for mutation 1 [K65R]	0.1804
cost2	Fitness costs for mutation 2 [M184V]	0.1540
cost3	Fitness costs for mutation 3 [K103N]	0.0440
cost4	Fitness costs for mutation 2 [Generic secondary mutation for drug 3, EFV]	0.0132
cost5	Fitness costs for mutation 2 [Generic secondary mutation for drug 4, TDF]	0.0132
FC_D1_Mut1	The fold change value for susceptibility to drug1 when mutation 1 is present	4.0
FC_D1_Mut2	The fold change value for susceptibility to drug1 when mutation 2 is present	0.5
FC_D1_Mut3	The fold change value for susceptibility to drug1 when mutation 3 is present	1.0
FC_D1_Mut4	The fold change value for susceptibility to drug1 when mutation 4 is present	1.6
FC_D1_Mut5	The fold change value for susceptibility to drug1 when mutation 5 is present	2.5
FC_D2_Mut1	The fold change value for susceptibility to drug2 when mutation 1 is present	61.0
FC_D2_Mut2	The fold change value for susceptibility to drug2 when mutation 2 is present	500.0
FC_D2_Mut3	The fold change value for susceptibility to drug2 when mutation 3 is present	1.0

FC_D2_Mut4	The fold change value for susceptibility to drug2 when mutation 4 is present	1.0
FC_D2_Mut5	The fold change value for susceptibility to drug2 when mutation 5 is present	1.0
FC_D3_Mut1	The fold change value for susceptibility to drug3 when mutation 1 is present	0.7
FC_D3_Mut2	The fold change value for susceptibility to drug3 when mutation 2 is present	0.8
FC_D3_Mut3	The fold change value for susceptibility to drug3 when mutation 3 is present	65.0
FC_D3_Mut4	The fold change value for susceptibility to drug3 when mutation 4 is present	40.0
FC_D3_Mut5	The fold change value for susceptibility to drug3 when mutation 5 is present	1.0

Hypothesis. We hypothesize that increasing or decreasing drug decay based on pharmacogenomic-pharmacokinetic information associated with the first-line ART regimen will reduce levels of HIV drug resistance within-host, and in the population using a stochastic, network-based model.

Model Limitations

One of the limitations of the model is the variance in certain pharmacokinetic properties such as IC50. The source of each pharmacokinetic parameter varied leading to differing value ranges (see Table 9). For instance, different IC50 ranges were observed for 3TC between for the FDA drug labels, the Rosenbloom et al. 2012 study[132], and the Parkin et al. 2004 study. Additionally, drug resistance mutation data was obtained using the genotype-treatment information on the Stanford HIV Drug Resistance database (HIVdb) and other published works [164–166]. We resolved the discrepancy in the values used for IC50 by simulating the within-host model for each set of IC50 values and choosing the values that were found to result in the HIV outcomes that were like outcomes from empirical HIV within-host, clinical and published peer-reviewed studies.

A sensitivity analysis was conducted to demonstrate how the model outcomes may vary due to the uncertainty of model parameter values and sources of information.

Table 3.9: Variance with IC50 values for first-line regimen antiretroviral medication.

Drug	Other Names	Molecular weight (g/mol)	Rosenbloom et al. IC50 (µMol)	FDA Labels IC50 min (uMol)	FDA Labels IC50 max (uMol)	Parkin et al. IC50 mean (uMol)	Parkin et al. IC50 median (uMol)	Rosenbloom et al. C_{max} (µmol)
TDF	Tenofovir	287.213	0.0561	0.04	8.5	1.1	0.7	1.1
3TC	Lamivudine Epivir	229.3	0.0298	0.002	15	4.2	2.4	15.3
FTC	Emtricitabine Emtriva	247.24	0.0079	0.0013	0.064	N/A	N/A	7.3
EFV	Efavirenz Sustiva	315.675	0.0035	N/A	N/A	4.5	1.6	12.9

3.5 Results

The simulated within-host model output demonstrated significant differences between the three (fast, intermediate, and slow) metabolizer types. The agent-specific simulated runs include viral load, drug concentration, and the different mutation strains (i.e., single, double, triple, quadruple, and quintuple mutants). This set of simulation runs reflect the modification of pharmacokinetic /drug information for the medication, EFV, and drug adherence levels at 45% and 90%. The second set of model predictions demonstrate the trade-off and variation of outcome data for predicting the percent resistance, the number of infectious agents, and the time to drug resistance (i.e., time to the emergence of the triple mutant). Each of these results demonstrate the possible trends and patterns for certain individuals fitting the profile of a fast, intermediate, or slow metabolizer, and their probability of drug resistance given that they are prescribed a first-line regimen that includes tenofovir (TDF), lamivudine (3TC), and efavirenz (EFV).

Agent Specific Plots: Drug Concentration

Drug concentrations were calculated for a duration of 300 days. The drug concentrations for tenofovir (TDF), lamivudine (3TC) and efavirenz (EFV) were included in each of the model outputs for the specified agent. The efavirenz drug was modified for each run to reflect changes in metabolizer type (i.e., fast, intermediate, and slow). The drug concentrations also varied based on the probability of adhering to the drug and drug decay values.

At 50% adherence, the drug concentrations remained low for the fast metabolizer, specifically for the drug efavirenz. The drug concentration for efavirenz for the fast metabolizer peaking to ~0.4 nMols. However, the drugs, tenofovir and lamivudine showed drug concentrations varying between 84 to 761 nMols and .00003 to 715 nMols, respectively, for the fast metabolizer. The intermediate metabolizer, demonstrated varied levels for the drug EFV, reaching peaks of 3151 nMols, while TDF and 3TC concentrations remained between the ranges of 105 to 758 nMols and 3.2E-6-

713 nMols, respectively. The slow metabolizer had drug concentrations with a wide range in value, between 0.1 to 7500 nMols. However, the drug concentrations for TDF and 3TC also remained varied and between the range of 0.1 to 150 nMols.

At 90% adherence, the fast metabolizer had a lowest drug concentration of 2.1 E-5 nMols for EFV. The tenofovir and lamivudine drug concentrations varying between 203-828 nMols and 58- 717 nMols, respectively. The intermediate metabolizer, demonstrated varied levels for the drug EFV, reaching peaks of 3159 nMols, while TDF and 3TC concentrations remained between the ranges of 203 to 830 nMols and 6-719 nMols, respectively. However, the slow metabolizer maintained a high level of efavirenz drug concentration, ranging between 3915- 8000 nMols, after the first 10 days of treatment.

Agent Specific Plots: Viral Load

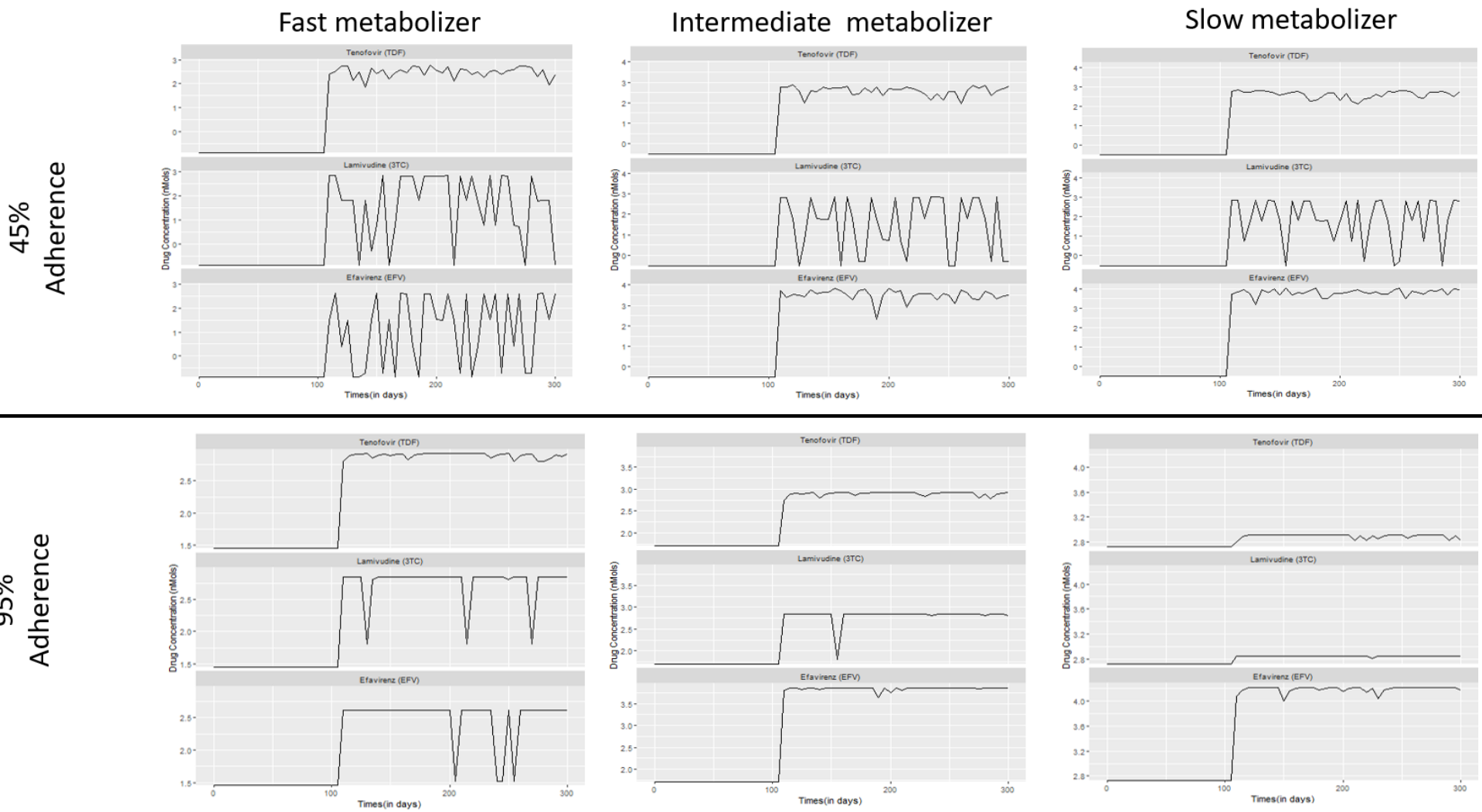
The viral load for each metabolizer corresponded to the initial changes in drug concentration experienced by each metabolizer type. At 45% drug adherence, there was a decrease between 20-30 days after initiating treatment for the fast, intermediate, and slow metabolizers. However, for the fast metabolizer, after the first decrease in drug concentration, due to the fast metabolizer having a higher drug decay rate and low drug adherence levels. Thus, the viral load rebounds at approximately 150 days. This could also be due to the set point viral load. The set point viral load was higher (above a log SPVL of 5) for the fast metabolizer and as such, may be more likely to rebound when not on medication or in this case, when the person has low drug concentrations in their system due to an increase in drug decay. At 95% adherence, the fast metabolizer also rebounded. However, there was a higher level of antiretroviral drugs (i.e., tenofovir, lamivudine, and efavirenz) present, so the levels of efavirenz were sustained until the efavirenz and lamivudine concentrations decreased to close to sub-therapeutic levels. In contrast, the viral load of the intermediate and slow metabolizers at 45% and 95% adherence demonstrated declines 20-30 after starting treatment and sustained undetectable levels by 300 days or the end of the simulation. This result may be due to no changes or lower levels of the drug decay rate, for the intermediate and slow metabolizers, respectively. As such, regardless of adherence levels, these agents can

sustain higher levels of drug concentration levels. However, the slow metabolizers may also reach toxic levels (i.e., above 4000 nMols), as stated in the prior section that discussed drug concentration levels.

Agent Specific Plots: Different Mutant Strains

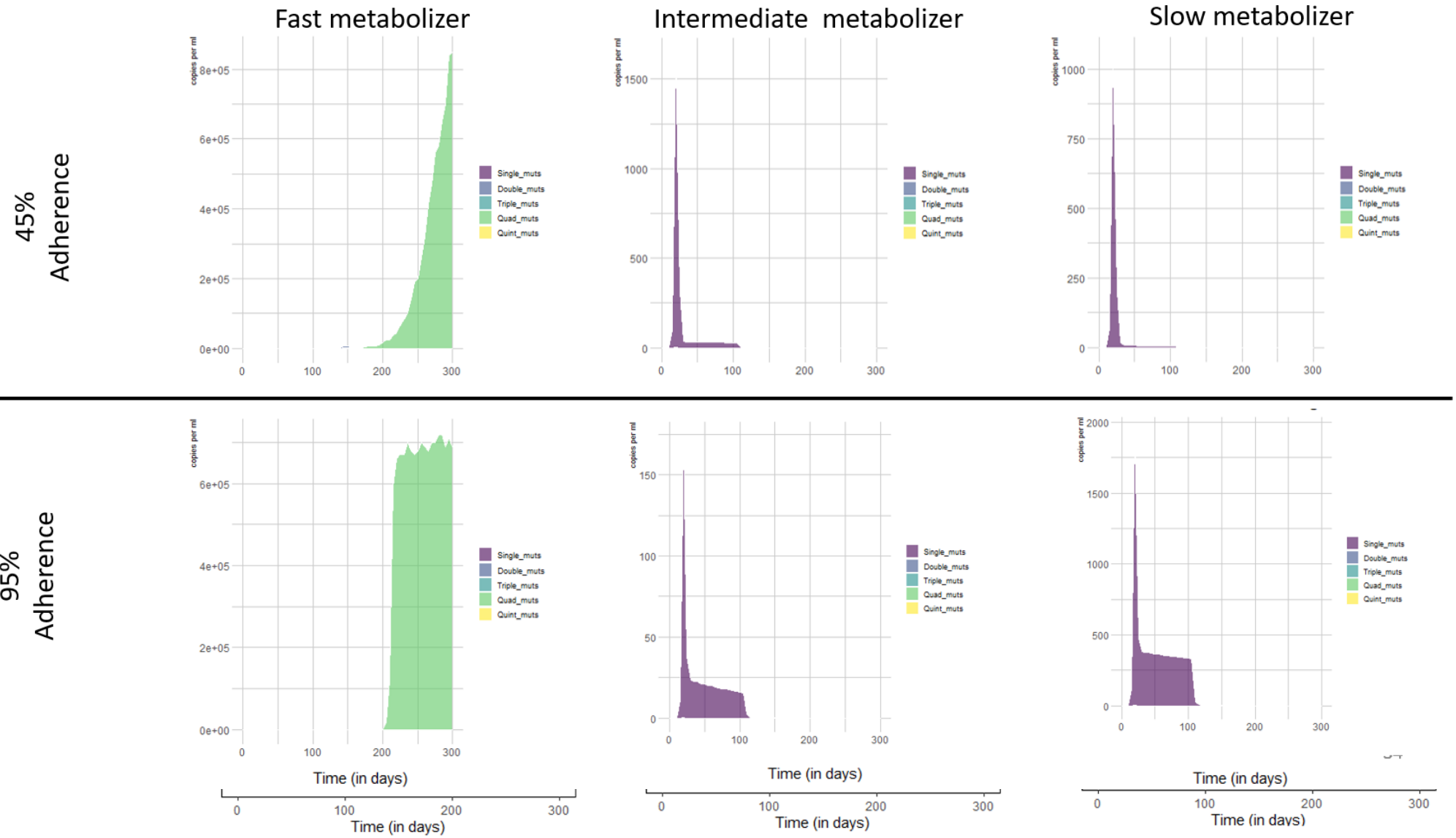
The agent specific mutant growth plots demonstrated the accumulation of single, double, triple, and quadruple mutants in the viral strain populations. The fast metabolizer demonstrated an accumulation of the quadruple mutant population that was dominant among all the viral strains in the viral population. The mutant strain was found to include the following mutants: K65R+K103N+GenEFV+GenTDF. The single, double, and triple mutants were found at lower concentrations and related to the K65R, M184V, and K103N. However, these mutant populations decreased once the quadruple mutant began increases around 200 days. The intermediate and slow metabolizers show accumulation of single mutations of K65R, M184V, and K103N, but rapidly decreased, shortly after starting treatment.

Figure 3.2: Viral Load of fast, intermediate, and slow metabolizers, randomly selected, from a within-host model at 45% and 95% drug



adherence.

Figure 3.3: The accumulation of 1-, 2-, 3-, 4-, and 5-mutant populations, within-host for fast, intermediate, and slow metabolizers



Number of Infectious Agents

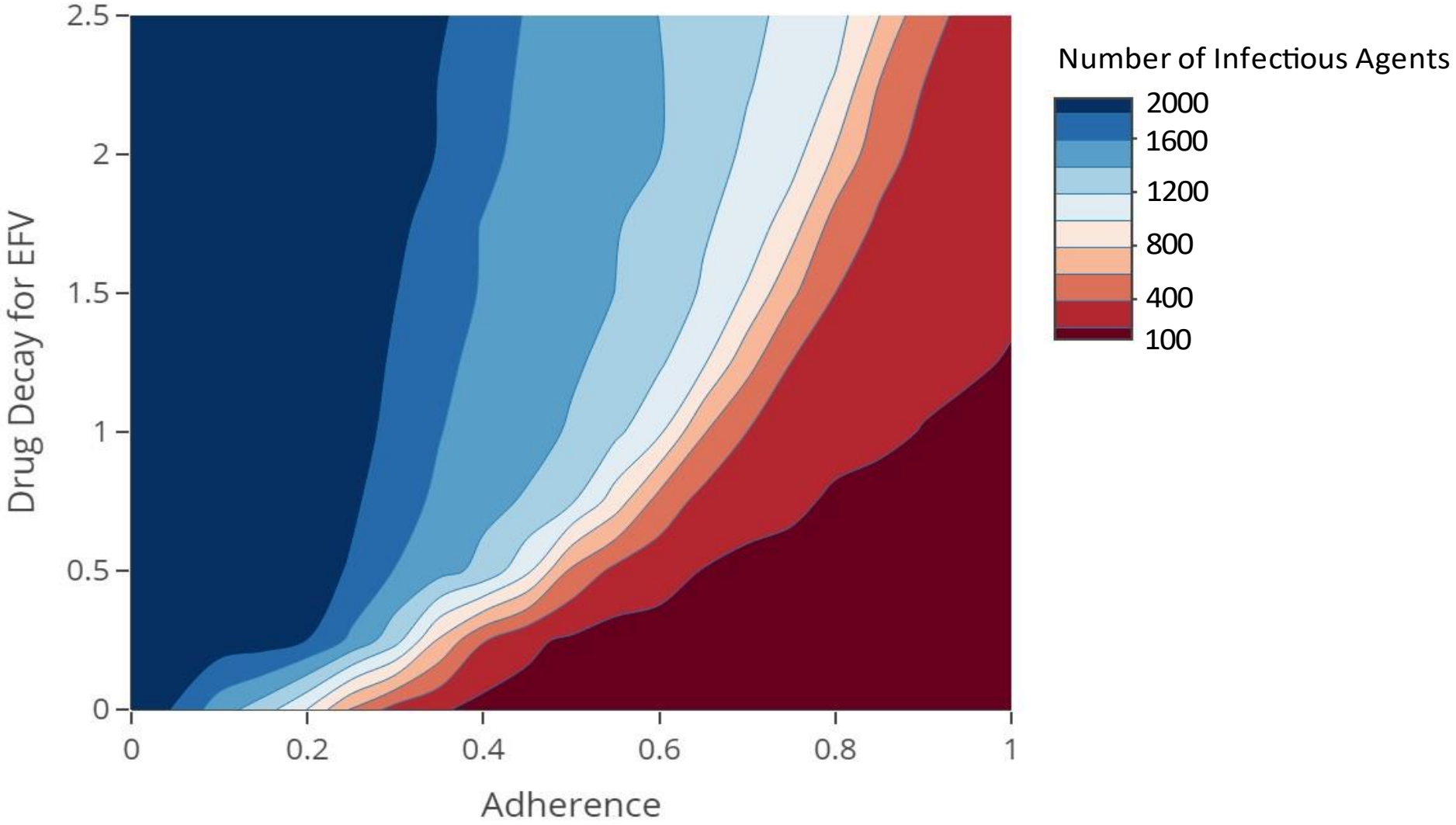
The number of agents that are at a stage of infection where they are more like to transmit the virus to an uninfected person if exposed to the virus. The number of infectious agents were predicted while varying adherence levels and drug decay for EFV. The number of infectious agents was highest among the agents between adherence levels of 0-35% from 550-580 days. For the slow metabolizers, the number of infectious agents was ~575 with adherence levels between 0-20% and a drug decay of 0.25. Whereas the fastest metabolizers had over 570 infectious agents between 0-35% adherence and a drug decay of 2.5. The lowest number of infectious agents occurred between adherence levels between 50-100% and for drug decay rates between 0.25-1.25 (i.e., ~2.5 days to ~12 hours).

For the slow and intermediate metabolizers, respectively, there were 22-50 infectious agents with adherence levels between 50-100% and drug decay levels between 0.25 and 1.25. Slow and intermediate metabolizers must be adherent to the EFV medication at 50% or greater to reduce the number of infectious agents. However, any adherence below 40% yields 100 or more infectious agents, given that the agents are slow or intermediate metabolizers with drug decay levels between 0.25 and 0.5. Fast metabolizer or agents with a drug decay higher than 1.25, have a higher threshold for acquiring the lowest number of infectious agents at 50 when adherence is very high. For example, for a drug decay of 1.25, the adherence has to be at a least 100% for the lowest number of infectious agents and for a drug decay of 2, the lowest number of infectious agents is 100, which is achieved by having an adherence of at least 95%. Fast, intermediate and slow metabolizers are have varying ranges of infectious agents from 22-580 agents. However, fast metabolizers were not able to reduce the number of infectious agents to 50, regardless of adherence levels and slow and intermediate metabolizers have a minimum adherence level of 50% needed to reduce the number of infectious agents to 50 agents or less.

Table 3.8: The lowest, median, and highest values for the number of infectious agents in the population with correspinding mean drug adherence, drug decay value and half life for the each metabolizer type

	Metabolizer Type	**Number of Infectious (count)**	**Mean Drug Adherence**	**Drug Decay 3 value**	**Half-Life**
Lowest Number of Infectious Agents	Slow	20	100%	0.25	2 days
Median Number of Infectious Agents	Slow	45	50%	0.25	2 days
Highest Number of Infectious Agents	Slow	577	10%	0.25	2 days
Lowest Number of Infectious Agents	Intermediate	22	100%	0.50	1.5 days
Median Number of Infectious Agents	Intermediate	194	50%	0.50	1.5 days
Highest Number of Infectious Agents	Intermediate	576	15%	0.50	1.5 days
Lowest Number of Infectious Agents	Fast	89	100%	2	5 hours
Median Number of Infectious Agents	Fast	463	50%	2	5 hours
Highest Number of Infectious Agents	Fast	580	5%	2	5 hours

Figure 3.5: The predicted number of infectious agents among all agents, by varying drug decay values for the HIV medication, Efavirenz (y-axis) and adherence levels (x-axis). A drug decay of 0.5 is equivalent to taking ~1.5 days to reach half the drug concentration and 2 days is equivalent to taking ~5-7 hours to reach half the drug concentration.



Percent Drug Resistance (among all agents in the population)

A sensitivity analysis was performed to investigate the changes in percent drug resistance, given a range of input values for drug decay of EFV and agent adherence levels. Figure 3 shows the variation in percent drug resistance among agents in the population. The contour plot displays levels of percent resistance, ranging from 0 to 100% (i.e., a proportion of 0.0 to 1.0) and the drug decay for EFV between 0.25 (i.e., a half-life of ~ 48 hours) and 2.5 (i.e., a half-life of ~3-5 hours). The plot shows that prior to administration of the EFV drug when adherence is zero, the percent resistance is low or zero. However, the tight contours in Figure 3, show that between an adherence greater than 0.0 and less than 0.05, an increase to 50% drug resistance amongst all agents is possible. As such, the cut-off for having majority of agents with drug resistance seems to occur if the adherence is between 0.05 and 0.20 for slow metabolizers or those with a drug decay of below 0.5. As drug decay increases to values of 0.5 (i.e., the value for an intermediate metabolizer) or to 2.5 (i.e., the value for a fast metabolizer), the contour band shifts to the right indicating a positive correlation between adherence and the drug decay parameters, within the range of 5% to 35% for drug adherence. This finding suggests that when adherence levels are below 35%, there is likely more drug resistance occurring with the infected population.

In contrast, when drug decay is below 0.5 (i.e., when the person is a slow metabolizer of EFV), drug resistance is less likely to occur at adherence levels starting at 50% adherence when the percent drug resistance drops from ~14% drug resistance at 45% adherence to ~7% drug resistance at 50% adherence. The lowest level of percent drug resistance has at 0.0 percent adherence, regardless of the drug decay level and between 50-100% in combination with a drug decay of no higher than 1.25 (i.e., a half-life of ~12 hours). After the threshold of a drug decay of 1.25, the percent drug resistance begins to increase again as adherence decreases, regardless of drug decay levels. Thus, these findings indicate that drug resistance will likely occur for agents who are faster metabolizers of EFV (i.e., a drug decay of >=1.25) in comparison to the intermediate (i.e., a drug decay of ~0.5) and slow metabolizers (i.e., a drug decay of <0.5), regardless of adherence level. Whereas the best options for achieving little to no drug resistance

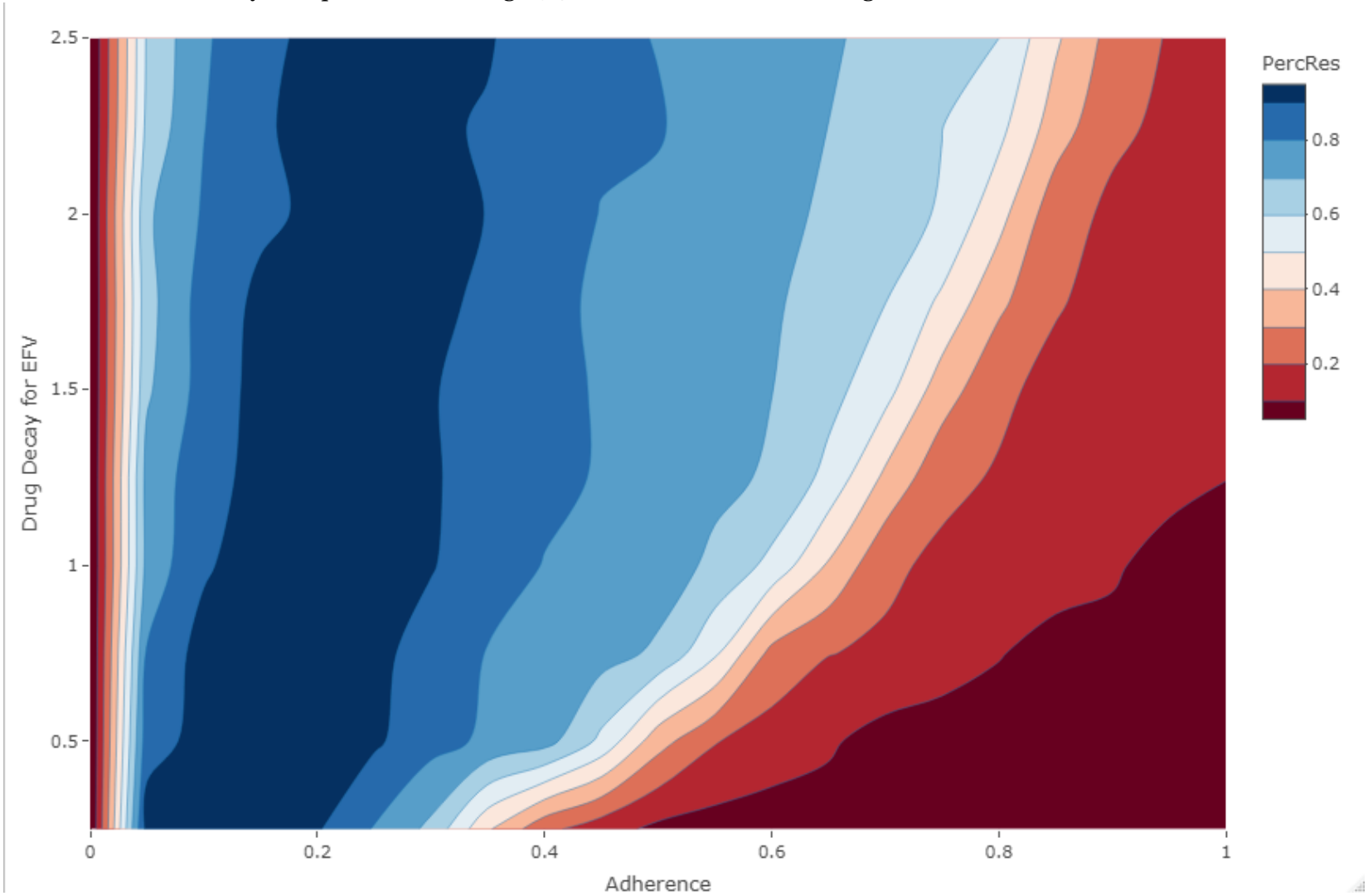
occurs with intermediate and slow metabolizers when adherence levels are at least 65% and 50%, respectively.

Table 3.9: The lowest, median, and highest values for the persistent drug resistance in the population with correspinding mean drug adherence, drug decay value and half life for the each metabolizer type

	Metabolizer Type	**Drug Resistance**	**Mean Drug Adherence**	**Drug Decay 3 value**	**Half-Life**
Lowest Drug Resistance	Slow	3%	100%	0.25	2 days
Median Drug Resistance	Slow	6%	65%	0.25	2 days
Highest Drug Resistance	Slow	98%	15%	0.25	2 days
Lowest Drug Resistance	Intermediate	3%	100%	0.50	1.5 days
Median Drug Resistance	Intermediate	21%	55%	0.50	1.5 days
Highest Drug Resistance	Intermediate	98%	15%	0.50	1.5 days
Lowest Drug Resistance	Fast	15%	100%	2	5 hours
Median Drug Resistance	Fast	6%	60%	2	5 hours
Highest Drug Resistance	Fast	93%	25%	2	5 hours

DR = Drug resistance

Figure 3.6: The predicted percent drug resistance (z-axis) among all agents, by varying drug decay values for the HIV medication, Efavirenz (y-axis) and adherence levels (x-axis). A drug decay of 0.5 is equivalent to taking ~1.5 days to reach half the drug concentration and 2 days is equivalent to taking ~5-7 hours to reach half the drug concentration.



Time to Drug Resistance

The time to drug resistance and the triple mutation outcompeting the wildtype virus, was simulated, and averaged among all agents. Several peaks and groups were identified with higher and lower ranges of time (in days) to the dominance of the triple mutant in the viral population, among all agents (Figure 3.7). The lowest and fastest times occurred when the drug adherence was between 00 and 50% adherence for each metabolizer, regardless of the value of drug decay. In addition, for lower drug decay levels (i.e., classified as either slow or intermediate metabolizers, the adherence levels included levels between 50-90%. However, the slowest time to a triple mutation for the slow and intermediate metabolizers happened when drug adherence was between 0-10% for drug decay levels between 0.6-2.0 (i.e., classified as fast metabolizers. This indicated that given the slow rate of decay for slow and intermediate metabolizers, moderate to high levels of drug adherence could help prolong the emergence of a triple mutant or drug resistance. Whereas fast metabolizers would have to have a high level of drug adherence.

Table 3.10: The fastest and slowest times (in days) for the time to drug resistance in the population with correspinding mean drug adherence, and drug decay values for the each metabolizer type

	Metabolizer Types	Drug Adherence	Drug Decay
Fastest Time	**Slow, Intermediate, and fast metabolizer** **Slow and Intermediate metabolizers**	**>0% and <50%** **50% and 90%**	**0.1 – 2.0**
Slowest Time	**Slow and Intermediate metabolizer** **Fast metabolizer**	**95-100%** **0-10%**	**0.1. – 0.2** **0.6 - 2.0**

In addition, Figure 3.8 shows that with a fast metabolizer with a drug decay of 1.3, when adherence is low at 0.6 or 60%, the partially resistant strains can grow if the resistant strains appear early after the individual has been infected with the virus. As adherence continues to decrease to 0.05 or 5%, the partially resistant strains are able to grow at a faster rate and lead to many individuals eventually developing drug resistance to at least three mutations (i.e., triple mutant drug resistance). However, as adherence increases to 0.7 or 70%, the partially resistant strains grow at a slower rate and

eventually, at 1.0 or 100% the drug resistance is only among 0.1 or 10% of individuals in the population.

Figure 3.7: The predicted time to drug resistance, or the triple mutant outcompeting the wildtype virus (z-axis) among all agents, by varying drug decay values for the HIV medication, Efavirenz (y-axis) and adherence levels (x-axis). A drug decay of 0.5 is equivalent to taking ~1.5 days to reach half the drug concentration and 2 days is equivalent to taking ~5-7 hours to reach half the drug concentration.

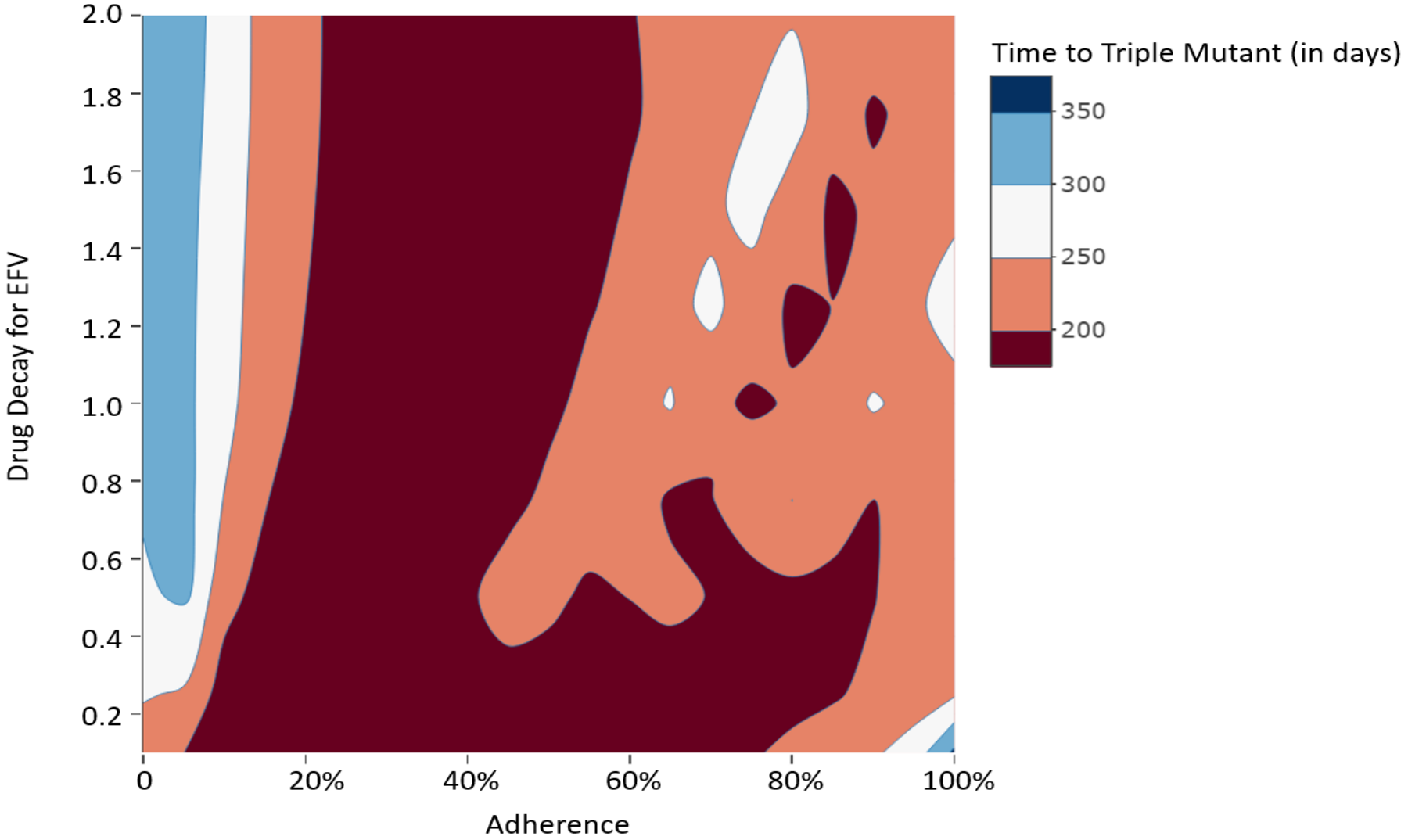
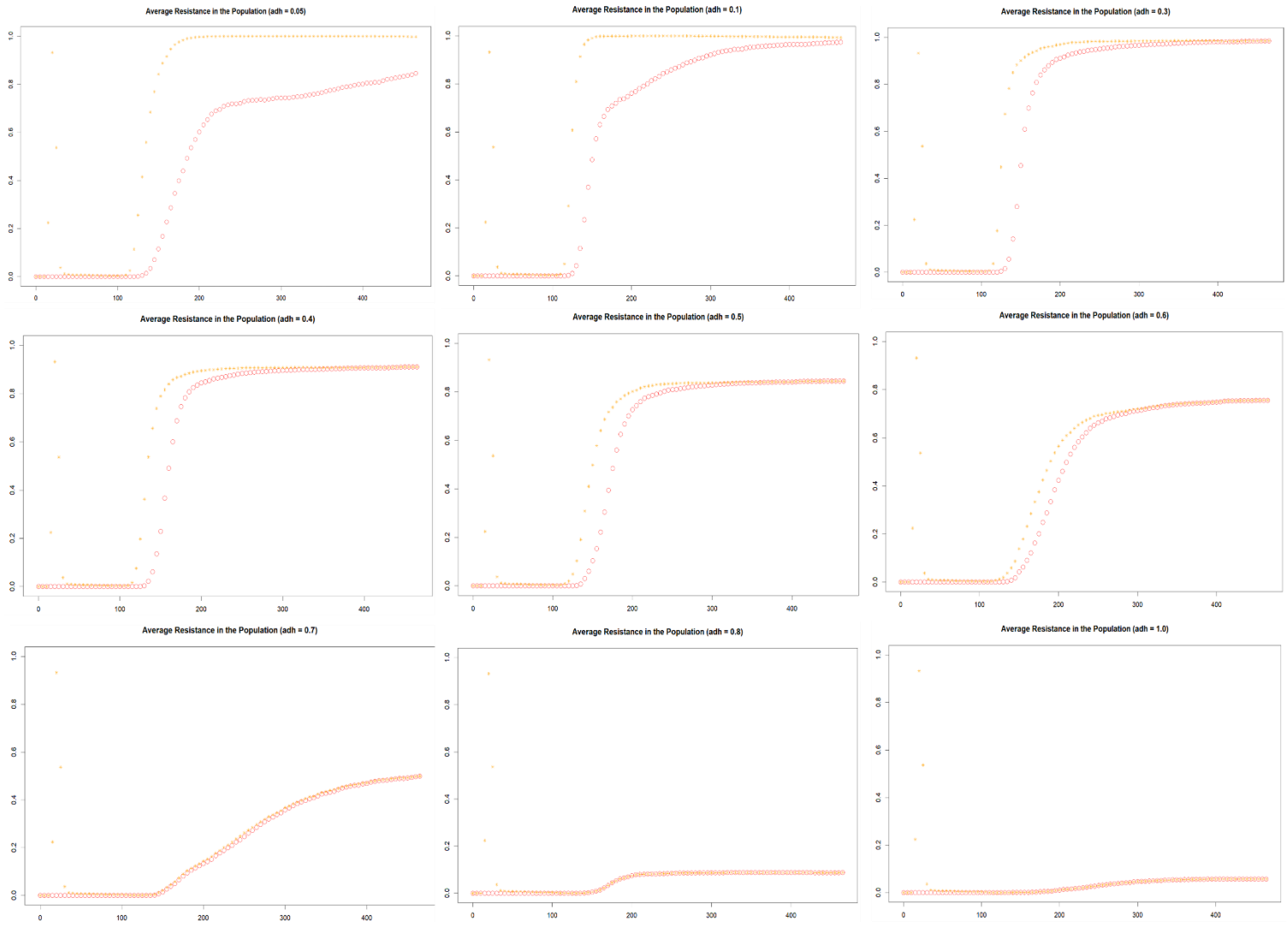


Figure 3.8: The average percentage of drug resistance in the population as adherence increases for a duration of 465 days. The adherence level ranges from 0 .05 to 1.0 (i.e., 0-100%).



* Partial Drug Resistance (2-mutants)

O Drug Resistance (greater than 3 mutants)

3.6 Discussion

Time to drug resistance

Patients that are non-adherent to an HIV treatment are likely to experience HIVDR. This is apparent when individuals who are fast metabolizers are also non-adherent to HIV medication. The Aim 3 simulations indicated that more individuals experienced drug resistance with either 2- or 3-mutant viral strains if adherence was below 60% for fast metabolizers or those with a higher decay rate. In terms of the time to drug resistance, the emergence of drug resistance mutations occurred after starting treatment for the agent specific graphs if drug adherence was lower than 50% of metabolizer type. However, the occurrence of drug resistance with at least three or more mutants (i.e., the triple mutant population) demonstrated unique time frames specific to the agent's metabolizer type.

The intermediate metabolizers, subsequently, having a drug decay that was unaffected due to the nature of intermediate metabolizers not having a genetic variant that decreases or increases drug decay for an individual, demonstrated drug resistance emergence for the single mutation, but if resistance occurs, these metabolizers will have a slow time to drug resistance (i.e., after about 200 days or 100 days after treatment. This result suggests that switching treatment to a more effective regimen within a 100-day period of starting treatment may effectively prevent drug resistance for intermediate and slow metabolizers. The emerging trend demonstrated that fast metabolizers more rapidly acquired drug resistance of at least the triple mutant population beginning approximately 90 days after starting treatment, where the triple mutant population outcompeted the HIV wildtype population. This, in large, may be due to the rapid decline in drug concentration and the need for a higher drug dose for fast metabolizers of EFV. Overall, the constant interruption of pill-taking and low-to-moderate levels of drug, within-host, helped to quicken the emergence of drug resistance (i.e., to 180days). In addition, triple and quadruple mutant populations were most prevalent in those that were fast metabolizers. These triple and quadruple mutations consisted of the following locus positions and specific mutants: position 1 (K65R), position 3 (K103N), position 4

(resistance to a generic TDF-associated mutant), position 5 (resistance to a generic EFV-associated mutant).

Surveilling the emergence of specific mutations

The emergence of certain drug resistance mutations also reflects the fitness costs associated with the mutant and the interaction between mutations. For example, the K03N and generic EFV mutations all have low fitness costs. Thus, in these simulations, K103N and generic EFV mutations are more likely to persist in the viral population over time and thus, are major contributors to the triple and quadruple mutations. The K103N mutation specifically has been found to persist from 1-3 years in PLHIV[167], which our simulations also reflect. In the case of a resistance mutation not emerging as frequently in the viral population due to fitness cost, the M184V mutant serves as an example. M184V has been observed to have a high fitness cost during primary infection [39]. The M184V mutation may be present at low levels. Also, in vivo studies M184V in the presence of K103N-mutants, for instance, was less fit, and thus, less frequent in the population[167]. This phenomenon reflects the simulated results of this aim, as M184V mutations were at low levels of copies per mL among the viral population.

Single-step drug resistance versus multi-step drug resistance

The emergence of the single, followed by double, triple, and quadruple mutations were a consistent pattern in each of the simulated agents regardless of metabolizer type or probability of drug adherence. This pattern reflects the propensity of a single mutation to give rise to another mutation in each replication cycle. The pattern of single step versus simultaneous drug resistance is also demonstrated in the simulated results of this aim and is built-in in the EvoNet framework. Specifically, the within-host model simulations reflect a stepwise or single-step resistant mutation accumulation[168,169]. Another possibility is that the accumulation of these drug resistant mutations is due to hitchhiking. Pennings et al. 2014 describes this phenomenon (i.e., the hitchhiking effect) as an occurrence where a single mutation appears and increases in frequency, allowing other mutations that are at closely linked loci to increase within the viral population as well[168]. Subsequently, less favorable viruses decreased in frequency. Specifically, this

phenomenon was present when observing the evolution of the single, double, triple, and quadruple mutations in the simulated within-host results. The single and double mutations often did not have high frequencies within the viral population after the start of treatment. However, the triple and quadruple mutations would often outcompete these single and double viral populations. Due to the randomness in the model, although having a single mutation makes the chance of having a double mutation more likely. Hitchhiking may not truly demonstrate the relationship between mutations well, as each drug is unique and may acquire single mutations at the same time, due to selective pressure when starting the triple drug therapy.

3.7 Summary

The time to drug resistance introduced a clear relationship between how rapidly a drug, in this case, EFV can be metabolized and excreted from a simulated agent in the context of the agent designated as being a fast, intermediate, or slow metabolizer. Consequently, adjusting drug dose may be beneficial to offset or prevent the emergence of drug resistant mutations. In particular, the World Health Organization recommends lowering the EFV dose to 400mg for individuals that are slow metabolizers versus the previous 600mg for intermediate metabolizers[75,170]. Several patterns and trends also emerged for the individuals classified as fast, intermediate, and slow metabolizers in the within-host model with respect to the drug concentration, viral load, specific mutations, number of infectious agents, developing drug resistance, and the time to drug resistance. Additionally, drug adherence levels played a role in the accumulation of certain mutations, possibly due to an interaction between X and Y. The contributions of this aim to the larger context of using pharmacogenomic data to understand and better dose certain drug treatment plans lays in the details of timing to the onset of drug resistance and the emergence of certain drug resistant mutations. Characterizing each metabolizer type to understand how and when drug resistance mutations appear maybe helpful for clinicians, researchers, and the patient in managing daily treatment options. Clinically, there are HIV-DR genotypes tests available in the United States using resources such as the University of Washington (UW), the Mayo Clinic or clinical laboratory networks and

companies like LabCorp that host many laboratory and genetic tests[152,153]. These laboratories commonly use Standard Genotypic Resistance testing (SGRT) that involves using Sanger sequencing of the reverse transcriptase, protease, and integrase to detect DRM[171]. These tests although thorough cost between $400-$550[171]. However, in low-to-middle income countries, these costs and equipment may not be affordable or feasible in certain settings, especially those outside of well-resourced cities. As such, other technologies have been developed like point-of-care assays that help to identify key DRM for under $5, for each locus[171]. Thus, it may be possible in a single visit to test patients, who have experience virological failure, for DRM prior to administering medication and thus prevent either an adverse health event like treatment failure, or toxicity, and ultimately, drug resistance, within-host.

In the context of "What strategies and conditions may assist in predicting, reducing, and preventing HIV transmission and HIV drug resistance within-host and between-host (i.e., in a multi-level system)?" and of Aim 1 that "predicted the effect of host genetic variation on the concentration of HIV-DR mutations, within-host, by modifying 'patient-specific' pharmacokinetic properties of antiretroviral medication, and drug adherence", the key conclusions were that:

(1) low drug adherence leads to the highest values for the number of infectious agents, percent resistance and the fastest time to a triple mutation,
(2) individuals that are fast metabolizers of the drug efavirenz would need to have high drug adherence level to achieve the lowest percent resistance and not become infectious. If not, these individuals will become infectious,
(3) individuals that are slow or intermediate metabolizers with an adherence level between 10-30%, are likely to become infectious,
(4) individuals that are slow and intermediate metabolizer with a drug adherence at or above 38% adherence or are fast metabolizers having a drug decay no greater than 1.25 (i.e., a half-life of ~14 hours) are less likely to become infectious,
(5) Fast, intermediate, or slow metabolizers with adherence levels between 0-30% have the highest percent resistance and had the fastest time to the emergence of a triple mutation. Whereas slow and intermediate metabolizers

that achieved 40-85% adherence or fast metabolizers with an adherence of 70-90%, had the slowest times to the emergence of a triple mutation.

The next chapter explores how the formation of drug resistance within-host may affect the evolution of HIV drug resistance between-host, or in the population. In aim 2 (Chapter 4), the goal was to utilize the EvoNet modeling framework and model setup of aim 1 (Chapter 3) to vary the proportions of pharmacogenomic parameters (i.e., fast, intermediate, and slow metabolizers), in the population and modify the pharmacokinetic parameter values to predict drug resistance levels, and in the population, given varying proportions of metabolizers in the population. In addition, in aim 2 (Chapter 4), the goal is to investigate the distribution of the different types of drug resistance, transmitted drug resistance and acquired drug resistance, to the overall proportion of HIVDR in the population.

Chapter 4: Population-Based Modeling and HIV-DR

4.1 Introduction

Achieving an accurate estimate of HIV drug resistance (HIVDR) levels at the population level may require the investigation of other contributing factors outside of the World Health Organization (WHO's) early warning indicators. Currently, the WHO's early warning indicators monitor clinical outcomes and quality of service delivery for certain prevention and treatment programs. However, the fields of genetics and pharmacology may provide additional information and factors that contribute to HIVDR levels in a population. Host genetic variation and prevalent drug resistance mutations that circulate in the population may serve as actors or early warning indicators for the prediction of HIVDR. In addition, current HIV population-based models may under or overestimate levels of drug resistance in the population if excluding factors such the frequency of genetic and viral mutations, within-host and in the population. With the rapid replication and evolution of HIV, developing a population-based model that includes both host and viral genetic information can attempt to track and classify population-based scenarios that are more likely to develop HIVDR. Results from this research aim could also help determine the contribution of acquired and/or transmitted drug resistance to the overall drug resistance levels in the population. Acquired drug resistance (ADR) occurs when an HIV-positive individual undergoing treatment develops drug resistant mutations[172]. Transmitted drug resistance (TDR) happens when a treatment-naïve HIV-positive person is infected with a mutant or drug resistant strain of HIV, due to their partner having drug resistant mutations[172].

In this chapter, I discuss (1) the global status of transmitted and pre-treatment drug resistance in low-income versus high-income countries, (2) The inter-ethnic variability of genetic variants such as CYP2B6, and (3) outline the research aims, hypothesis, approach, methodology, and simulation results through the comparison of different population-based HIV models. This is in the context of my overall question of "What strategies and conditions may assist in predicting, reducing, and preventing HIV

transmission and HIV drug resistance within-host and between-host (i.e., in a multi-level system)?" as well as Specific Aim 2 to "model the effect of host genetics and prevalent drug resistance mutations on HIV-DR levels in two sub-Saharan African populations". Specifically, I address whether one should be concerned about the spread of HIVDR in the presence of host genetic information such as the proportion of fast or slow metabolizer in a population or the presence of drug resistance mutations in a population. Furthermore, among two populations with differing frequencies of metabolize types and drug resistance mutations, I address the topic of public health decision making for determining which populations have the greatest risk of spreading HIVDR and where to prioritize resources for the prevention of HIVDR.

4.2 Background

Global Status of Transmitted and Pre-Treatment Drug Resistance. Globally, there have been increasing trends of drug resistance, specifically in places like sub-Saharan Africa and amongst certain demographic populations (i.e., men who have sex with men) where the epidemic is most prevalent39–4139–4139–41. Each geographical region of the world has an increasing trend of pre-treatment drug resistance among people newly diagnosed with HIV. Countries in Africa, South America, Europe, Asia, and North America currently on the World Health Organization's drug resistance surveillance list, have submitted reports on the prevalence of TDR and PDR yearly. For instance, the WHO report predicted that for Southern Africa alone, if PDR exceeded 10% then an additional 890,000 deaths due to AIDS and over 450,000 new infections are predicted to occur between 2016-2030 without an intervention. These predictions also reflect the continued use of NNRTIs as a first-line drug regimen, which was also recommended to be used less and replaced with an ART medication with a higher barrier to resistance. Although the WHO's new guidelines suggest that the first-line drug regimen should no longer include NNRTIs to prevent pre-treatment drug resistance and that the alternative antiretroviral, dolutegravir, should be used instead, and due to stockouts or interruptions such as other disease outbreaks such as COVID-19, the medication may be inaccessible. In addition, no medication is perfect and without the risk of drug resistance. Thus, the WHO adjusted guidelines for prescribing certain drug regimens and supported monitoring pre-treatment drug resistance, globally.

The next sections outline the most recent assessments of pre-treatment drug resistance in the regions of the world outlined in the WHO drug resistance surveillance list and how health officials and organizations are strategizing to combat pre-treatment drug resistance. Additionally, I discuss the economic burden of the HIV epidemic.

PDR in Africa. In differently regions of the continent of Africa, varying levels of PDR have been observed. Gupta et al. 2018 conducted a systematic review and meta-regression analysis that demonstrated from 1996 to 2016, there has been an increase of transmitted drug resistance among countries in southern, eastern, central, and western Africa[166,175]. Figure 4.1 demonstrates the growing increasing trend over the 20-year period in regions of Africa, specifically among those taking NNRTI based medication[175]. Gupta et al. 2018 found that among the 61 studies with results detailing pre-treatment genotypes, exposure to antiretroviral drugs, and accessible sequence data, pre-treatment resistance (i.e., specifically to NNRTI-based drug regimens) odds were increasing yearly by 23% in southern Africa, while other regions like western and central Africa, and eastern Africa increased by 17%[175].

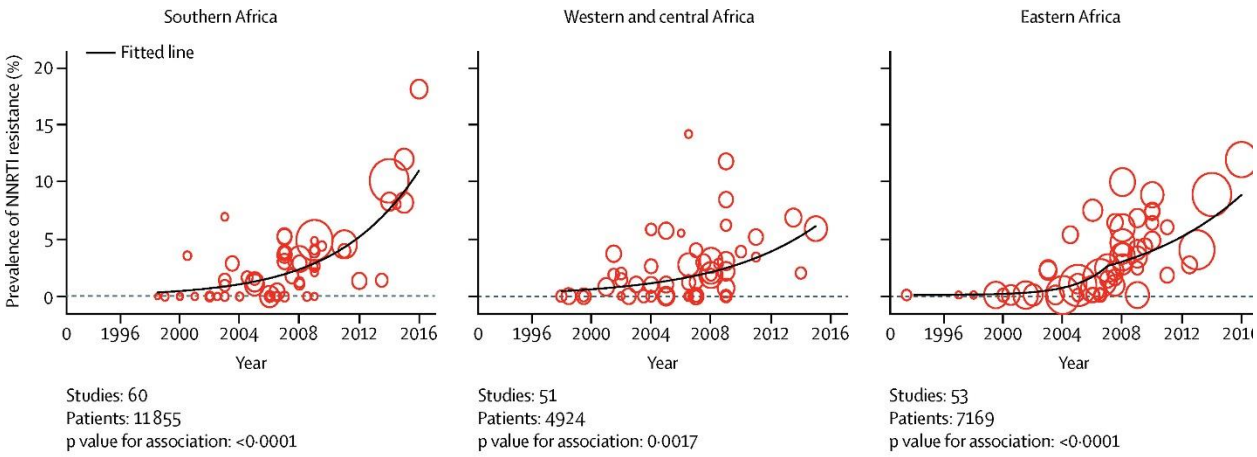


Figure 4.1: The percentage of pre-treatment drug resistance in sub-Saharan African countries as reported by the World Health Organization from 1996-2016 [166,175].

A recent study by Crowell et al. 2020 sequenced genotypes of 972 ART-naïve and experienced individuals from the African Cohort Study (AFRICOS) from the years 2013-2019 in four countries (i.e., Uganda, Kenya, Tanzania, and Nigeria)[176]. The study found that among the 801 ART-naïve participants the surveilled drug resistant mutations had

following prevalence's: 8.2% (n=65) had NNRTI mutations, 4.7% (n=37) had NRTI mutations, and 0.4% (n=3) had PI mutations. Overall, 11% of ART naïve participants had at least one surveilled pre-treatment drug resistant mutation[176]. Among the 90 ART-experienced individuals that experienced virological failure, 36.7% had a known resistance to the antiretroviral medications efavirenz and lamivudine and were still taking the prescribed medication[176]. These prevalence rates and increasing trends for pre-treatment drug resistance in sub-Saharan Africa indicate the need and urgency to implement more effective strategies for reducing PDR levels and stopping PDR from occurring.

PDR in Latin America and the Caribbean. Specific drug resistant mutations have also been identified to be more prevalent in individuals with transmitted drug resistance. Coelho et al. 2019, conducted a study in Sao Paulo, Brazil, sequenced 596 individuals newly diagnosed with HIV and found that 10.9% (n=65) had one or more the globally surveilled TDR mutations[177]. Of those TDR mutations K103N was observed to be the most prevalent, resulting in drug resistance for the NNRTI-based resistance (n=41), then NRTI-based resistance (n=22), and PI resistance (n=13). Lower numbers of dual- and triple -class resistance was found, only two cases were observed for each category. This study also outlined the need to conduct pre-treatment genotype testing prior to the initiation of NNRTI based regimens to ensure viral suppression[177]. In addition, these individuals were found to not be able to suppress the virus and have trouble suppressing the virus when switching to a second-line regimen. Thus, without genotype testing identifying whether TDR mutations exists, the wrong treatment regimen will be given to the patient increasing the risk for virological failure and significantly lessening options for switching to a second-line regimen for individuals living with HIV in Sao Paulo, Brazil[177]. In addition, Gupta et al. 2018 found that pre-treatment drug resistance among those taking NNRTI-based regimens found that increases of about 0.9% were likely between the years 2015-2016 and a prevalence of 9.4% for the year 2016[175]. These values were also consistent with other studies containing empirical data for those selected years.

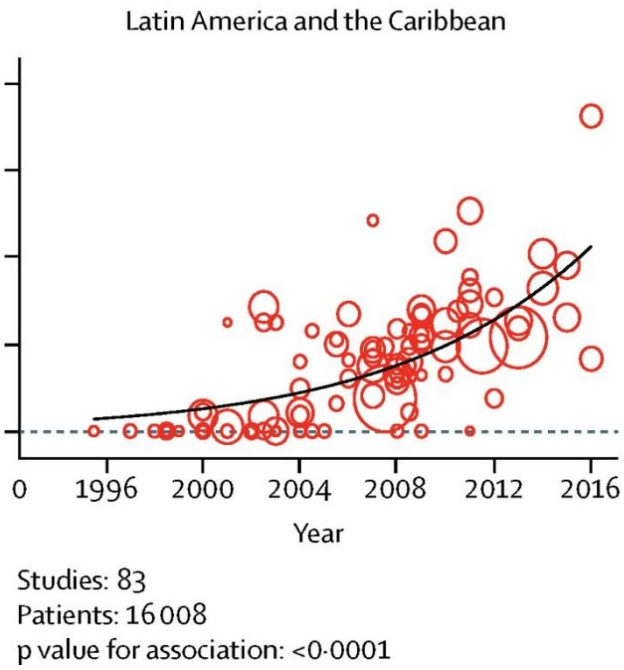


Figure 4.2: The percentage of pre-treatment drug resistance in countries in Latin American and the Caribbean as reported by the World Health Organization from 1996-2016. [166,175] The size of the circles indicates the sizeable population from small to large.

PDR in Asia. In the geographic region of Asia, many studies have been conducted that outline the current state of individuals newly diagnosed with HIV and those experiencing virological failure. A study by Houng et al. 2019 showed that PDR rates among ART naïve individuals were 9.6% and 14.7%, respectively, for the countries of Thailand and Vietnam. These numbers over the WHO threshold of 10% indicated the need to conduct national surveys to track the PR prevalence in other regions/clinics around the continent. In addition, Gupta et al. 2018 reported that the yearly, predicted increases in pre-treatment drug resistance was 11% in Asia, with NNRTI-based drug regimens increasing by 3.6%. Gupta et al. 2018 also noted that the populations surveilled were in urban areas, where there is more accessibility to clinical and laboratory resources like genotyping tests to track PDR.

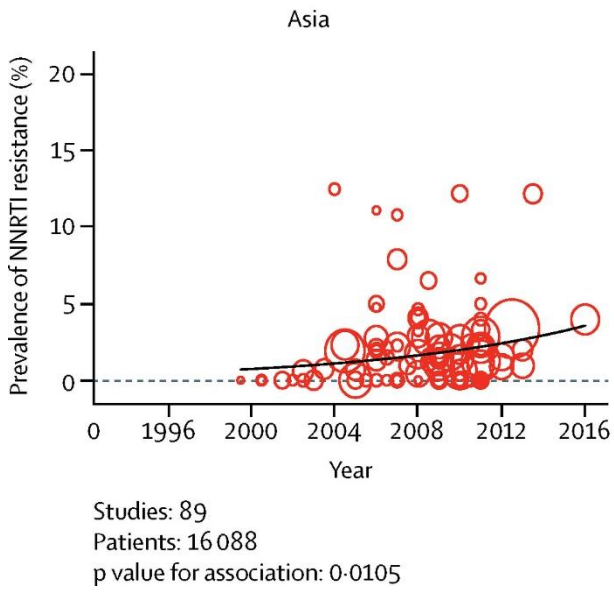


Figure 4.3: The percentage of pre-treatment drug resistance for countries in Asia, as reported by the World Health Organization from 1996-2016. [166,175] The size of the circles indicates the sizeable population from small to large.

Another study by Zou et al. 2020 highlighted the difference in the PDR levels of urban versus rural areas of China[178]. The study indicated that there were higher PDR and ADR patterns in rural communities of central China (i.e., Henan, Hubei, and Hunan provinces). Specifically, the study found that PDR prevalence rates of 4.5% (95%CI: 4.0–5.0), 5.1% (95%CI: 4.1–6.1) and 2.4% (95%CI: 2.1–2.7), respectively, in North, Central and South China. The higher prevalence of PDR in central China versus North and South China is likely a result of historically tragic events that led to an HIV outbreak. This outbreak was a result of unlicensed, commercial blood transfusion centers infecting many individuals after re-using supplies in the early-1990s[178]. Thus, free HIV clinics were established to administer antiretrovirals to anyone affected. However, with the scale-up of antiretroviral use and people on antiretrovirals with a low barrier to resistance for a longer duration of time, drug resistance mutations would inevitably emerge. Additionally, adherence levels to antiretrovirals were relatively low and thus, drug resistance was more likely to occur leading to the transmission of drug resistant virus[178]. However, instances of PDR in regions of Asia, sub-Saharan Africa, and Latin America and the Caribbean currently contrast the levels of PDR found in areas of Europe and the United States in recent years, mostly due to more individuals diagnosed with HIV having access to care and to newer antiretroviral medications.

PDR in Europe. In the Europe, access to genotype testing is more readily available and has become a common clinical practice in starting an individual on a first-line ART regimen[179,180]. However, transmitted drug resistance remains an issue, as most countries do not have a national surveillance of PDR with PDR surveillance occurring within specific regions of the country designated as high-risk areas. A study by Hosfa et al. 2016 tracked PDR levels in Europe since 2001, finding that among 26 countries in Europe the TDR prevalence in was 8.3% with little to no significant change between the years 2002-2007 and 2008-2010[180]. However, another study conducted in Cologne, Germany, where the highest incidence rates for HIV exist among all the European countries, effecting 13.7 per 100,000 individuals, found that regional estimates for pre-treatment drug resistance ranged from 10.4% to 12.5%[181]. Additionally, Stecher et al. 2018 found that transmitted drug mutations were more common in the suburban areas (outside of Cologne) and with the following risk groups in order of the highest prevalence: men who have sex with men, heterosexuals, and people with persons who inject drugs (PWID)[182].

Although the prevalence rates of PDR in some countries in Europe have reached the WHO threshold for intervention, many European countries have also reached the 90-90-90 goals designated to be achieved by the year 2020. Thus, there may be smaller populations of non-suppressed HIV positive individuals and HIV-DR individuals living in Europe. A major contributing factor in the smaller number of non-suppressed HIV positive people due to the availability of second-line drugs with a higher barrier to resistance like dolutegravir. However, there has been reports of drug resistant mutations accumulating for those prescribed a dolutegravir-based regimen[183].

PDR and the United States. The United States like Europe has made significant progress in reducing levels of pre-treatment drug resistance. However, several studies and public health agencies have reported that certain geographical areas of the United States have either a sustained level of resistance over the threshold or an increasing trend of pre-treatment drug resistance, particularly in the southern states[184,185]. Southern states like Georgia, Louisiana, Alabama, Florida, and North Carolina have become the epicenter of the HIV epidemic in the United States. In 2018, the CDC also reported that among the 37,968 newly diagnosed HIV cases in the United States,

approximately 51% are in the southern states. A study by Levintow et al. 2018 reported that 12% of the study population located in North Carolina had PDR and, in their assessment, had trends of increasing percentage yearly[184]. The major risk groups and associations included men who have sex with men (MSM), being a white male, a person younger than 20, higher CD4 cells counts, and those that were a part of a transmission cluster.

In contrast, other studies have found different demographic groups with a higher risk for obtaining PDR. Rich et al. 2021 conducted a retrospective analysis to assess PDR prevalence in the state of Florida and found that between 2012 and 2017, PDR prevalence was 23.5% and specific drug-class resistance ranged from 6.6% to 29.7%[186]. The factors that predicted HIVDR and PDR outcomes included people ages 46 and older, racially identifying as Black, mother-to-child transmission of HIV[186]. Additionally, HIVDR and PDR prevalence rates varied by county, but lower-income areas with a history of job loss and poorly accessible mental health resources had a higher likelihood of HIVDR and PDR [186].

Other non-southern states and cities also show patterns a sustained or increasing trend in transmitted drug resistance. The King County public health department in Seattle, Washington reported that from 2011 to 2020, there was an increasing trend in any high-level resistance among both acquired and transmitted drug resistance for those newly diagnosed with HIV[187]. Among these individuals who tested for HIV-1 and genotyped, 22% had drug resistance in 2020 (an 8% increase from 2011) and 25% had NNRTI drug class resistance (a 10% increase from 2011)[187]. The demographic population of individuals living with acquired or pre-treatment drug resistance in King County, Seattle, Washington included individuals that were homeless, a person who inject drugs (PWID), and/or identified as men who have sex with men (MSM). Thus, the prevalence of drug resistance among special populations in the United States continues to echo the need for interventions and drug resistant monitoring to assist in the eradication of HIV/AIDS.

4.3 Approach

EvoNetC Modeling System.

The EvoNetC modeling system is comprised of several R and C programming scripts, primarily written in the C programming language to ensure computational efficiency and reduce the requirement of loading R packages and is similar in structure and function to the EvoNetHIV system. The EvoNetC modeling system particularly includes addition of the proportion of metabolizer types (fast, intermediate, and slow) and the assignment of the proportion of mutations at one of the five locus positions at the start of the model.

Comparison of Four Epidemic Models

I compared four model scenarios for predicting HIV drug resistance, specifically transmitted, or acquired drug resistance, and other HIV population-based outcomes (see Table 4.1). The first model scenario has conditions that do not include pharmacogenomic factors (i.e., fast, intermediate, or slow metabolizers) or nor does it include prevalent drug resistant mutation data. Thus, the first scenario represents a case-control simulation. The second model scenario incorporates pharmacogenomic data in the existing EvoNetC model. The model scenario includes the proportion of fast, intermediate, and slow metabolizers assigned in the entire population. The third model scenario incorporates only a prevalent drug resistant mutation (i.e., K103N) in the population at the start of the simulation. K103N represents a mutation located at the third locus position and is associated with the drug Efavirenz (EFV). The fourth model scenario incorporates both the pharmacogenomic/metabolizer data and the prevalent drug resistant mutation in the population at the start of the simulation. Each model predicts HIV drug resistance levels in the population and outputs population-based HIV dynamics such as HIV prevalence, HIV incidence (i.e., new infections), and AIDS deaths.

Table 4.1. The four model scenarios demonstrated in a 2x2 table. Each quadrant represents the data added to the HIVDR model simulation.

	No Pharmacogenomic (PGx) data	**PGx data (Fast, Intermediate, Slow) metabolizers**
No Drug Resistant Mutation (DRM)	Model 1 (No PGx, No DRM)	Model 2 (PGx, No DRM)
DRM Present (K103N mutation)	Model 3 (No PGx, DRM)	Model 4 (PGx and DRM included)

A two-way ANOVA test was used to compare model outcomes for HIVDR based on the PGx and DRM data collected. Additionally, the model characteristics were chosen to be representative of HIV-positive study populations in the countries of Zimbabwe and the Democratic Republic of Congo.

4.4 Methods

Data Collection. DRM and PGx data were collected using published works and summarized in Table 4.2. The prevalent DRM used in the model, K103N, frequency data was extracted from the studies by Chimbete et. al 2018[188] and Kamangu et al. 2015[189] for the DRC and Zimbabwe, respectively. While the proportion of fast, intermediate, and slow metabolizer or PGx data was extracted from the studies by Nyakurita et al. 2008 and Peko et al. 2019, for the DRC and Zimbabwe populations, respectively. The pharmacokinetic data collected reflects IC50, fitness costs, drug dosing, and drug decay rates for the first-line regimen: tenofovir, lamivudine, and efavirenz (see Table 4.3). In addition, epidemic modelling parameters reflect a longer epidemic of 10 years or 3650 days. For each model, 20 replications were simulated to account for the stochasticity of each set of results.

Table 4.2. The input parameters and corresponding values and references for the two study populations in Zimbabwe and the Democratic Republic of the Congo

Country	Input Parameter	Value	Reference
Zimbabwe	Proportion of DRM (K103N) in the study population among those with ADR	0.35	Chimbetete et al. 2018
Zimbabwe	Proportion of Fast metabolizer in the study population	0.00	Nyakurita et al. 2008
Zimbabwe	Proportion of Intermediate metabolizer in the study population	0.51	Nyakurita et al. 2008
Zimbabwe	Proportion of Slow metabolizer in the study population	0.49	Nyakurita et al. 2008
DRC	Proportion of DRM (K103N) in the study population among those with ADR	0.13	Kamangu et al. 2015
DRC	Proportion of Fast metabolizer in the study population	0.17	Peko et al. 2019
DRC	Proportion of Intermediate metabolizer in the study population	0.55	Peko et al. 2019
DRC	Proportion of Slow metabolizer in the study population	0.28	Peko et al. 2019

Table 4.3. Model Simulation default parameters.

Input Parameters	Definitions	Default Values
Model Parameters		
nsims	Total number of simulations	20
Epidemic Parameters		
n_steps	Total duration of the model	3750 days
initial_pop	Total population size	3000 agents
initial_infected	Total number of infected agents	800 agents
Treatment parameters		
min_adherence	Minimum level of adherence for Drug1, Drug2, Drug3	0.90
max_adherence	Maximum level of adherence for Drug1, Drug2, Drug3	0.90
prob_eligible_ART	Probability of receiving ART treatment	1.00
start_treatment_campaign	Day to start treatment after infection	Day 100

Hypothesis. I hypothesized that increasing or decreasing drug properties such as drug decay based on pharmacogenomic-pharmacokinetic information associated with the first-line ART regimen will reduce levels of HIV drug resistance in the population using a stochastic, network-based model.

Model Limitations. The model does not reflect a detailed collection of population-based data reflective of social and behavioral characteristics of HIV-positive individuals within the countries of Zimbabwe and the Democratic Republic of the Congo. In addition, the values of the input parameters vary according to the study data provided. The Zimbabwe study population reflects individuals that experienced virological failure and acquired drug resistance while the Democratic Republic of the Congo (DRC) study data reflected treatment-naïve individuals that developed ADR without indication of

virological failure. Other limitations include the differing host genetic, metabolizer data observed for each study population. The Zimbabwe study population only included PGx data on slow and intermediate metabolizers, while the DRC study population included PGx data on fast, intermediate, and slow metabolizers. Thus, the two study populations are not directly comparable in terms of having fast, intermediate, and slow metabolizer frequency data available. This, in turn, may affect the percentage of individuals in the simulated population level results that develop acquired or transmitted drug resistance. However, I provided a general overview of the effect varied frequencies or proportions of fast and slow metabolizers on the levels of drug resistance in the population, using contour plots of a simulated population that either only includes fast or slow metabolizers, in addition to, intermediate metabolizers.

4.5 Results

Percent Drug Resistance, Metabolizer Types, and a Drug Resistant Mutation

The relationship between the proportion of individuals with a certain metabolizer type (i.e., fast, or slow), individuals having a drug resistant mutation, K103N, and developing drug resistance in the population is presented in Figures 4.4 and 4.5. Figure 4.4 shows that among a scenario that includes fast metabolizers in the model, higher levels of drug resistance occur as the proportion of fast metabolizers increase in the population. This increase is not dependent upon the proportion individuals with of K103N mutations in the population, as shown in Figure 4.4. In contrast, Figure 4.5 shows that when slow metabolizers are present in the model, lower levels of drug resistance occur as the proportion of slow metabolizers increase in the population. Additionally, the higher levels of resistance is dependent upon the proportion of individuals with the K103N mutation, as drug resistance increases as the proportion of individuals with K103N mutations increases.

Figure 4.4. The percent of individuals with drug resistance in the population for a duration of 365 days (contour colors) as a function of the proportion of individuals with a K103N mutation in the population (x-axis) and the proportion of individuals that are fast metabolizers in the population (y-axis).

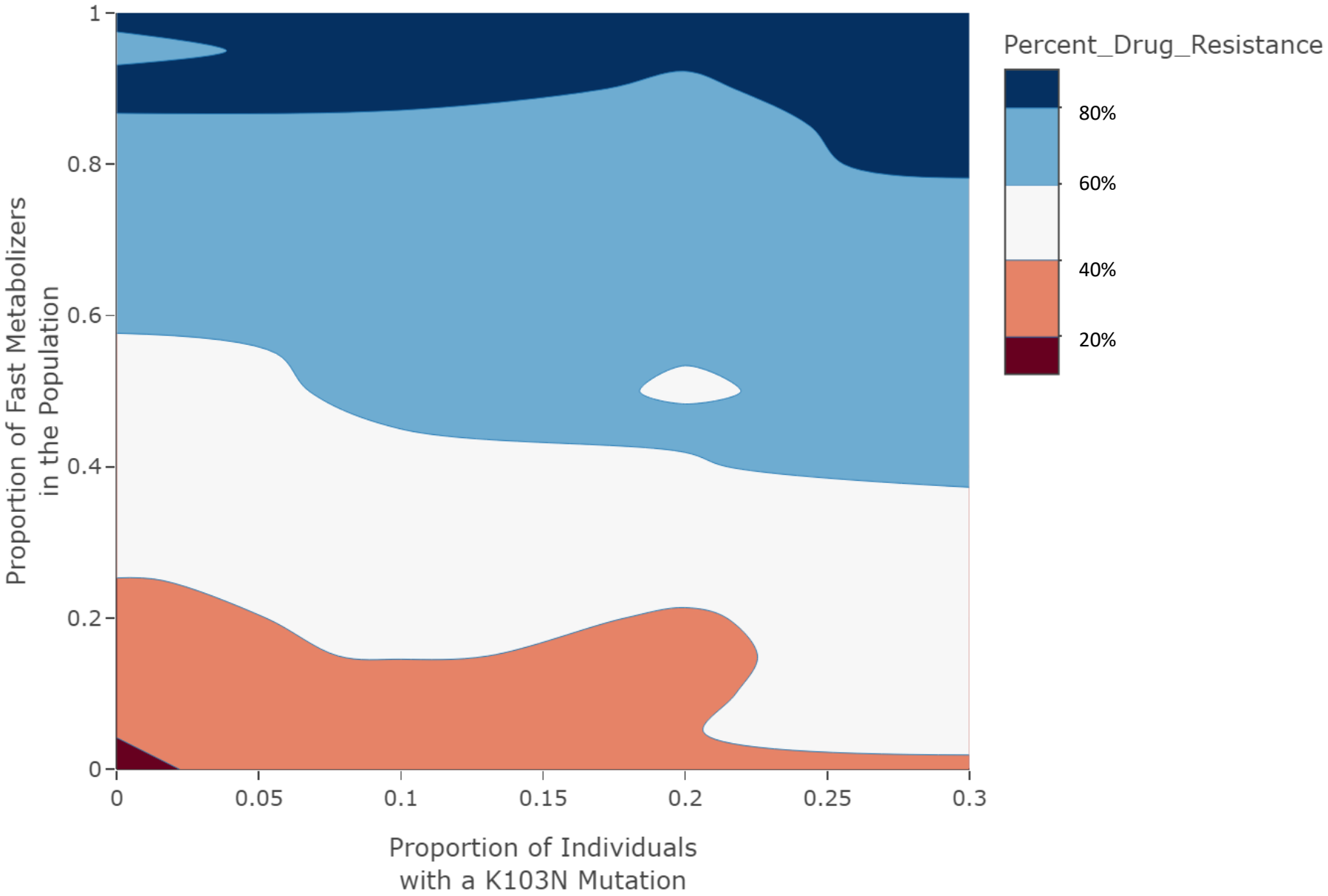
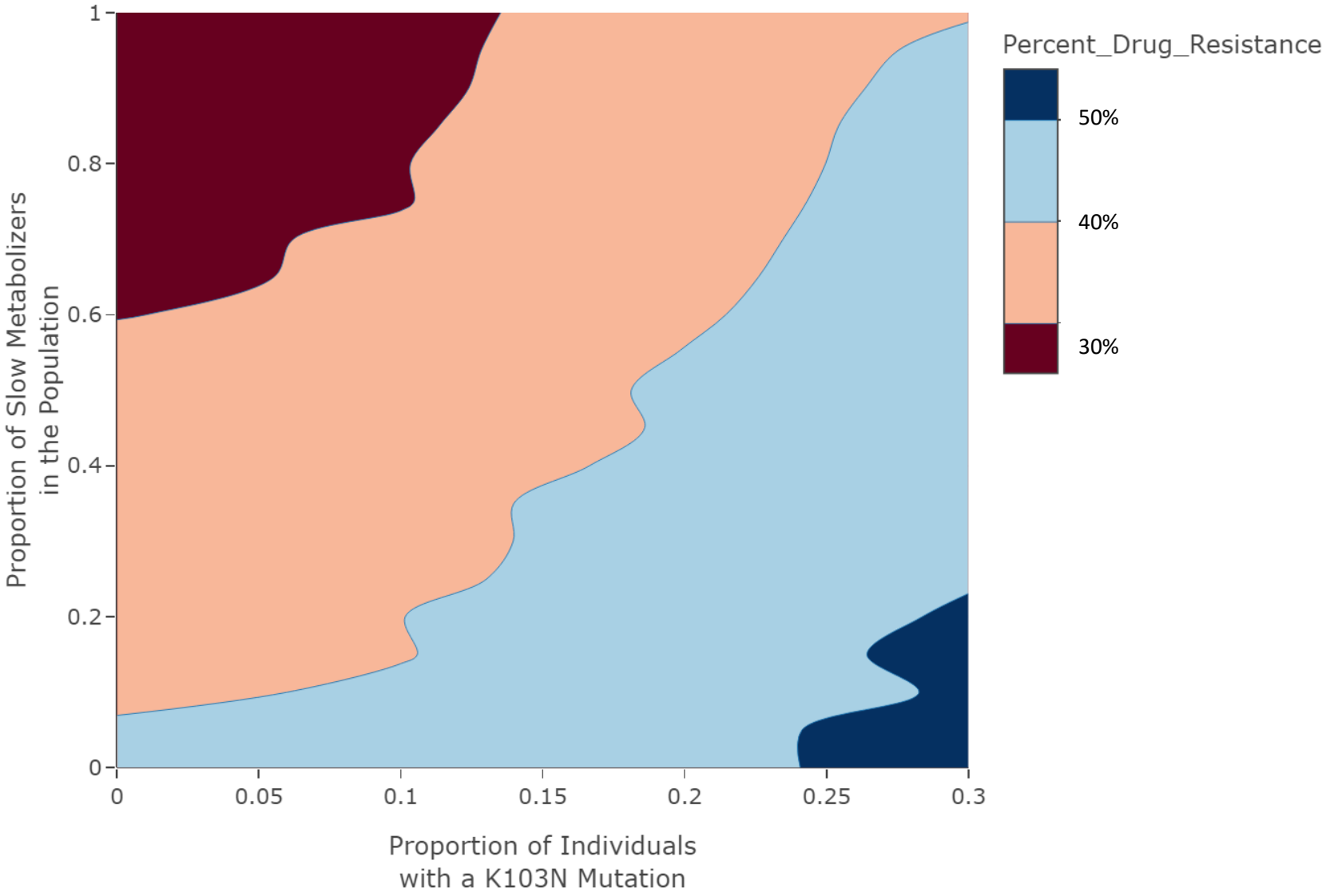


Figure 4.5. The percent of individuals with drug resistance in the population for a duration of 365 days (contour colors) as a function of the proportion of individuals with a K103N mutation in the population (x-axis) and the proportion of individuals that are slow metabolizers in the population (y-axis).



Comparison of 4 Model Scenarios

For each study population, the four models the proportion of indviduals with drug resistant mutations within the population were compared (see Tables 4.4-4.7 and Figures 4.6-4.7). In both study populations of Zimbabwe and the DRC, the lowest prediction of HIVDR levels in the population at the end of the simulation were 0.5 and 0.57, respectively, for model 2, where PGx data was included only. Next, for the Zimbabwean population Model 1 and Model 4 demonstrated HIVDR levels in the population aproaching 0.70 and 0.72, respectively. Finally, Model 3, where the prevalent DRM, K103N, was included in the model among the entire population, had the highest levels of HIVDR at the end of the simulation at 0.81, or 81% of the population. The Congolese population model outcomes showed a higher level of HIVDR in the population for model 1 compared to model 4 at 0.71 and 0.64, respectivly. Model 3 also demonstrated the highest level of HIVDR at 0.75.

Table 4.4 – Zimbabwe study population - Four model comparison of drug resistance levels in the population at 3750 days, or year 10 of the simulation.

DR Levels	Model 1 (no PGx, no DRM)	Model 2 (PGx only)	Model 3 (DRM only)	Model 4 (PGx and DRM)
Model 1	X	M2 ↓	M3 ↑	M4 ↑
Model 2	M1 ↑	X	M3 ↑	M4 ↑
Model 3	M1 ↓	M2 ↓	X	M4 ↓
Model 4	M1 ↓	M2 ↑	M3 ↑	X

† M1 represents "Model 1", no PGx and no DRM included in the model.
† M2 represents "Model 2", PGx only included in the model.
† M3 represents "Model 3", DRM only included in the model.
† M4 represents "Model 4", PGx and DRM included in the model.

Table 4.5 – Congolese study population - Four model comparison of drug resistance levels in the population at 3750 days, or year 10 of the simulation

DR Levels	Model 1 (no PGx, no DRM)	Model 2 (PGx only)	Model 3 (DRM only)	Model 4 (PGx and DRM)
Model 1	X	M2 ↓	M3 ↑	M4 ↓
Model 2	M1 ↑	X	M3 ↑	M4 ↑
Model 3	M1 ↓	M2 ↓	X	M4 ↓
Model 4	M1 ↑	M2 ↓	M3 ↑	X

† M1 represents "Model 1", no PGx and no DRM included in the model.
† M2 represents "Model 2", PGx only included in the model.
† M3 represents "Model 3", DRM only included in the model.
† M4 represents "Model 4", PGx and DRM included in the model.

Figure 4.6 Comparison of Models 1 to 4 for the Zimbabwe study population.

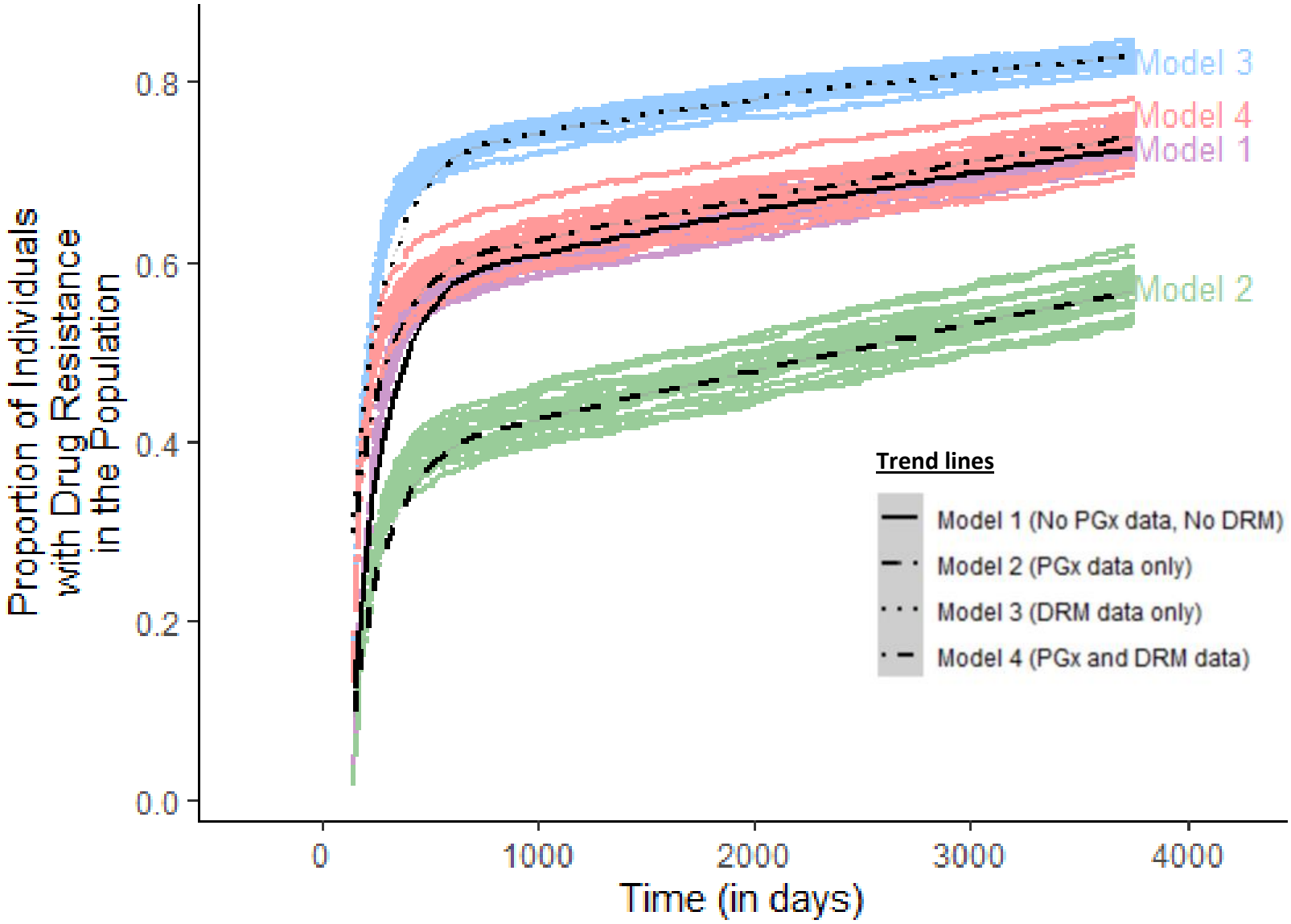


Table 4.6 Models that include Zimbabwean Pharmacogenomic and Drug Resistance Mutation Data - Parameter values for PGx and DRM data for each of the four models.

	Proportion of K103N	Proportion of Fast Metabolizers	Proportion of Intermediate Metabolizers	Proportion of Slow Metabolizers
Model 1	0.00	0.00	0.00	0.00
Model 2	0.00	0.00	0.51	0.49
Model 3	0.35	0.00	0.00	0.00
Model 4	0.35	0.00	0.51	0.49

Figure 4.7. Comparison of Models 1 to 4 for the Democratic Republic of Congo (DRC) study population

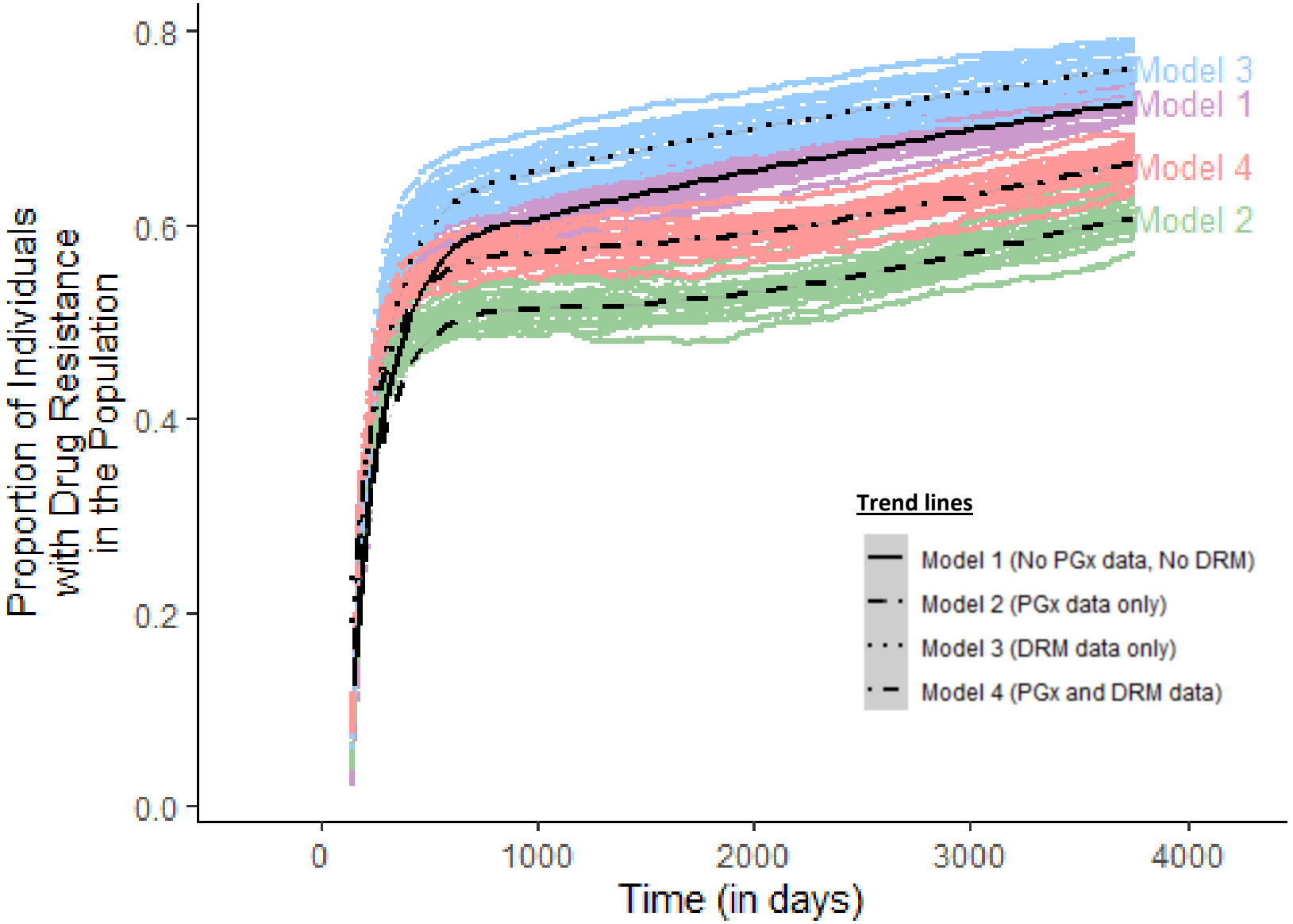


Table 4.7 Models that include DRC Pharmacogenomic and Drug Resistance Mutation Data Parameter values for PGx and DRM data for each of the four models.

	Proportion of K103N	Proportion of Fast Metabolizers	Proportion of Intermediate Metabolizers	Proportion of Slow Metabolizers
Model 1	0.00	0.00	0.00	0.00
Model 2	0.00	0.17	0.55	0.28
Model 3	0.13	0.00	0.00	0.00
Model 4	0.13	0.17	0.55	0.28

Hierarchical Data Visualizations for Drug Resistance and Metabolizer Types

Using hierarchical (icicle) charts, infected, HIV positive individuals were stratified into clusters and groupings based on their: (1) drug resistance status, (2) transmitted or acquired drug resistance status, and (3) genotype/metabolizer status. Figures 4.8-4.9 show the number of infected individuals with hierarchical classification and percent drug resistance levels among two model scenarios (i.e., the model 1 scenario – no PGx, no DRM and the model 4 scenario – PGx and DRM included) for the Zimbabwe study population. Figures 4.10-4.11 show the number of infected individuals with hierarchical classification and percent drug resistance levels among two model scenarios (i.e., the model 1 scenario – no PGx, no DRM and the model 4 scenario – PGx and DRM included) for the DRC study population. 20 replications for each model scenario were generated and a representative profile for each hierarchical charts is presented in Figures 4.8-4.11. Overall, in both study populations, there was a high percentage of drug resistance among those infected, for all four models (see Tables 4.8-4.9). There was also a higher proportion of transmitted drug resistance than acquired drug resistance among infected individuals with drug resistance in all four models, for both populations.

Table 4.8. The percentage of infected individuals with drug resistance, transmitted drug resistance, and acquired drug resistance, in the population, using Zimbabwe pharmacogenomic and drug resistance mutation data.

Model Name	Study Population	Drug Resistance among the infected population	TDR, among infected and drug resistant individuals	ADR, among infected and drug resistant individuals
Model 1 (No PGx, No DRM)	Zimbabwe	75%	60%	40%
Model 2 (PGx only)	Zimbabwe	62%	65%	35%
Model 3 (DRM only)	Zimbabwe	82%	60%	40%
Model 4 (PGx and DRM)	Zimbabwe	75%	59%	41%

† DR represents "Drug Resistance", in the infected population.
† TDR represents "Transmitted Drug Resistance", in the infected and drug resistant population.
† ADR represents “Acquired Drug Resistance”, in the infected and drug resistant population.

Table 4.9. The percentage of infected individuals with drug resistance, transmitted drug resistance, and acquired drug resistance, in the population, using the Democratic Republic of the Congo (DRC) pharmacogenomic and drug resistance mutation data.

Model Name	Study Population	Drug Resistance among the infected population	TDR, among those that are infected and drug resistant	ADR, among those that are infected and drug resistant
Model 1 (No PGx, No DRM)	DRC	75%	62%	38%
Model 2 (PGx only)	DRC	75%	61%	39%
Model 3 (DRM only)	DRC	79%	59%	41%
Model 4 (PGx and DRM)	DRC	75%	60%	40%

† DR represents "Drug Resistance", in the infected population.
† TDR represents "Transmitted Drug Resistance", in the infected and drug resistant population.
† ADR represents "Acquired Drug Resistance", in the infected and drug resistant population.

For the distribution of metabolizers found in Zimbabwe, model 3 had the highest number of infected individuals, followed by model 1> model 4> model 2. Model 3 included only intermediate metabolizers since the model was simulated to reflect only the K103N mutation frequency information being known. Although, Model 3 had no slow or fast metabolizers in the simulation, having 35% of the simulated population with the K103N mutation at the beginning of the model run led to more people developing drug resistance overall, suggesting that the presence of a drug resistant mutation present in 35% of the population, may contribute to increasing drug resistance levels overall. None of the agents with transmitted drug resistance had the K103N mutation at the beginning of the model simulation. For the increase in ADR for model 3, majority (i.e., 207/447, or 46%) of the agents with ADR at the end of the simulation, or year 10 were agents with the K103N mutation.

Also, for models 2 and 4, which included slow metabolizers, there was a decrease in total drug resistance in the population compared to the control, or Model 1. This suggests that slow metabolizers are effective in suppressing the virus and as such, including slow metabolizers in the simulation may result in a lower estimate of total drug resistance. Surprisingly, 3 out of the 4 models-maintained levels of transmitted and acquired drug resistance of 60% and 40%, respectively, regardless of the percentage of total drug resistance in the population, or if including PGx or DRM data. However,

Model 2 which only included slow and intermediate metabolizers demonstrated a decrease by 5% in the acquired population and an increase of 5% in the transmitted population. Agents with high log SPVL of 5 or higher, even as slow metabolizers, developed drug resistance. Whereas agents that were slow metabolizers with a low log SPVL did not appear to develop drug resistance unless an intermediate metabolizer with a high SPVL developed drug resistance and transmitted the virus to the slow metabolizer. In this case, whether the slow metabolizer had a high or low SPVL, the agent developed drug resistance due to the partner's mutant virus. Thus, there may be a trade-off between intermediate and slow metabolizers in the model related to timing of the emergence of drug resistance, set point viral load, drug concentration that relates to more transmitted drug resistance developing in the population.

For the distribution of metabolizers for the Democratic Republic of Congo (DRC), model 3 also had the highest number of infected individuals, followed by model 1> model 4> model 2. Fast, intermediate, and slow metabolizers were present in the model. However, model 3 had a lower K103N mutant frequency (i.e., 13%) than the mutation data gathered form the Zimbabwe population (i.e., 35%). The low frequency of the drug resistant mutation in the population may have led to a lower total drug resistance level in the population. Although, there was a difference between the percentage of TDR and ADR levels. In comparison to the control model, or model 1, there was an increase in ADR and decrease in TDR for model 3. By the end of the simulation, among the agents that had the K103N mutation, 83% had ADR, 5% had TDR, and 12% had no drug resistance for model 3. The increase in ADR for model3 and the high percentage of ADR among the agents with the K103N mutation may be due to more agents having a log SPVL between 4 and 5 (n=47), whereas there were fewer agents with TDR and the K103N mutation with a log SPVL between 4 and 5 (n=12). For model 2, which contained the fast, intermediate, and slow metabolizers, there is a clear relationship between metabolizer type and developing drug resistance. For agents that are fast metabolizers, the drug concentration for one of the three antiretroviral medications (i.e., efavirenz) decreases to sub-therapeutic concentrations (i.e., below 1000 nMols), and as such, develops drug resistance since in the presence of the drug, the agent is not able to suppress the virus and mutant viruses are able to replicate. Whereas slow metabolizers

can suppress the virus and have supratherapeutic levels that are slightly above the therapeutic range for EFV.

Figure 4.8: Zimbabwe, Model 1 HIVDR Hierarchical Chart – No PGx, No DRM included, with only intermediate metabolizers in the population.

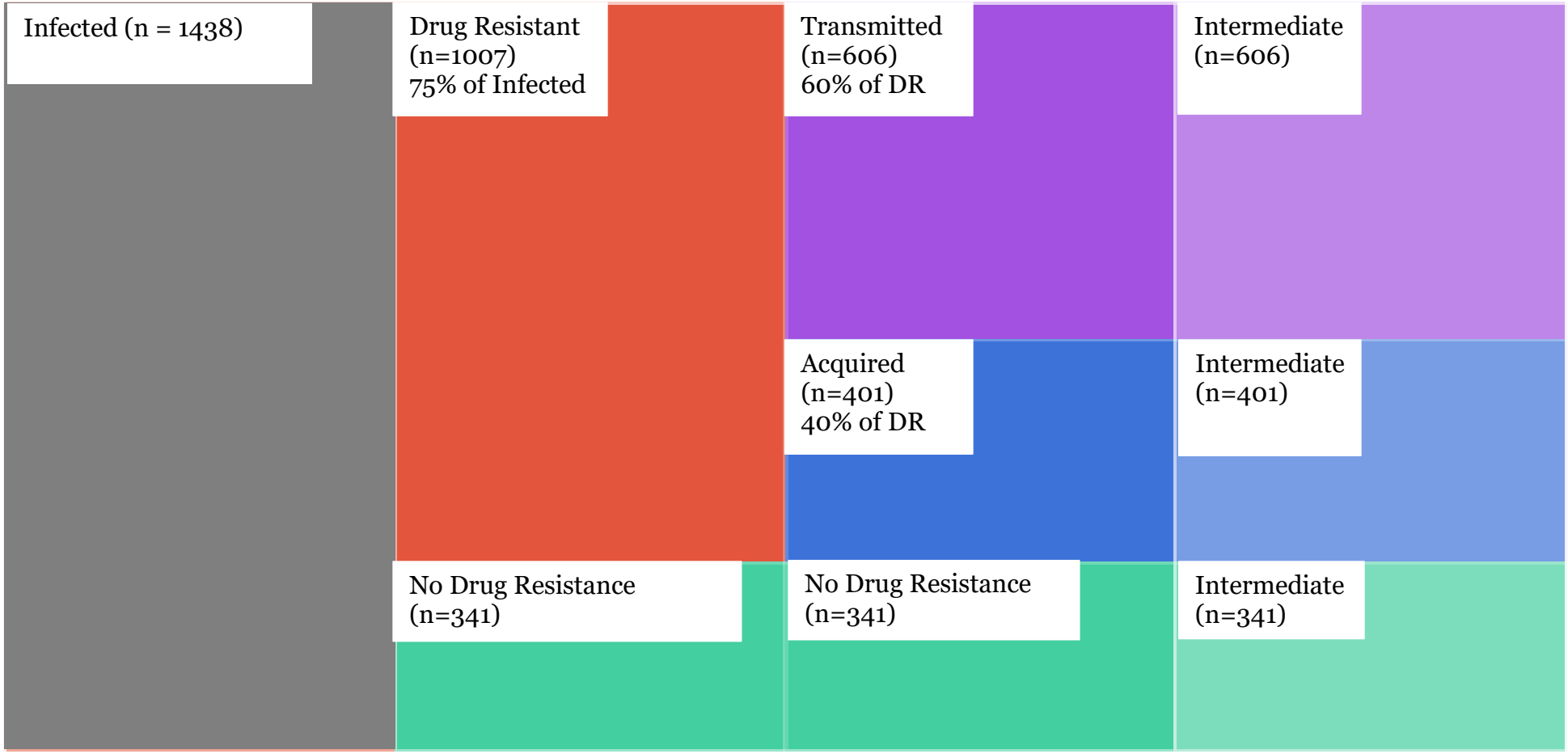


Figure 4.9 Zimbabwe, Model 4 HIVDR Hierarchical Chart – PGx and DRM included, with intermediate and slow metabolizers in the population.

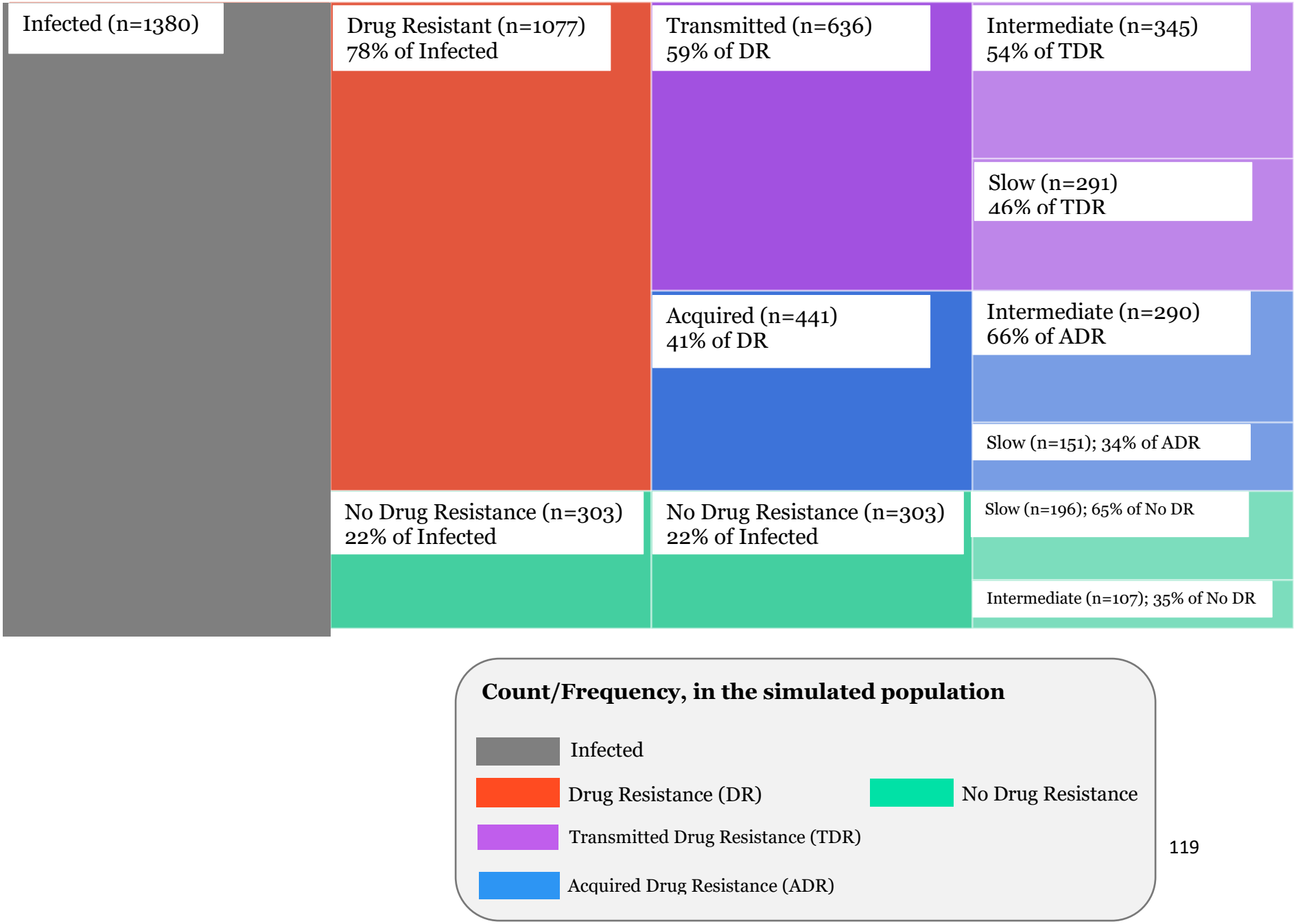


Figure 4.10: Democratic Republic of the Congo, Model 1 HIVDR Hierarchical Chart – No PGx, No DRM included, with agents assigned as intermediate metabolizers.

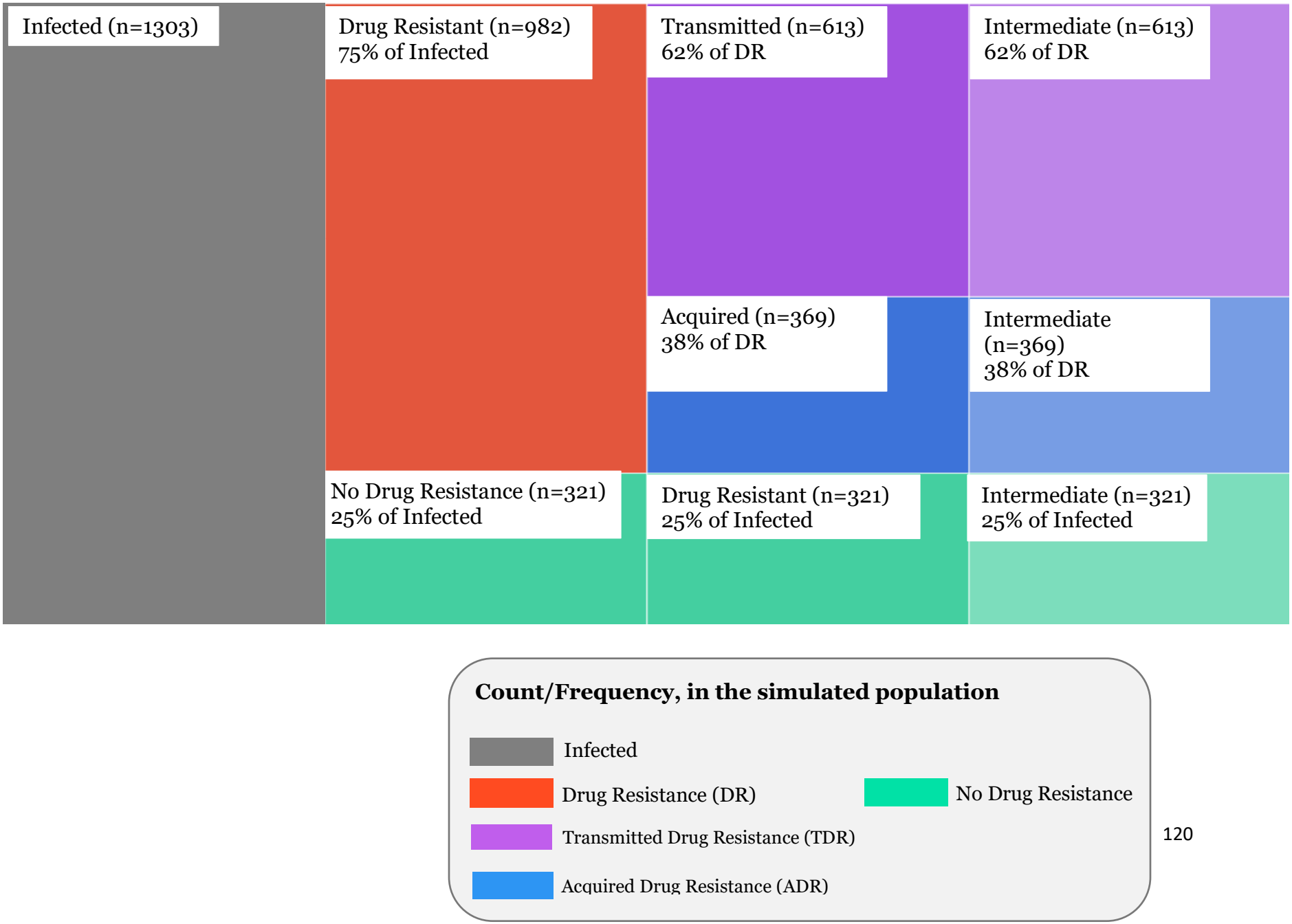
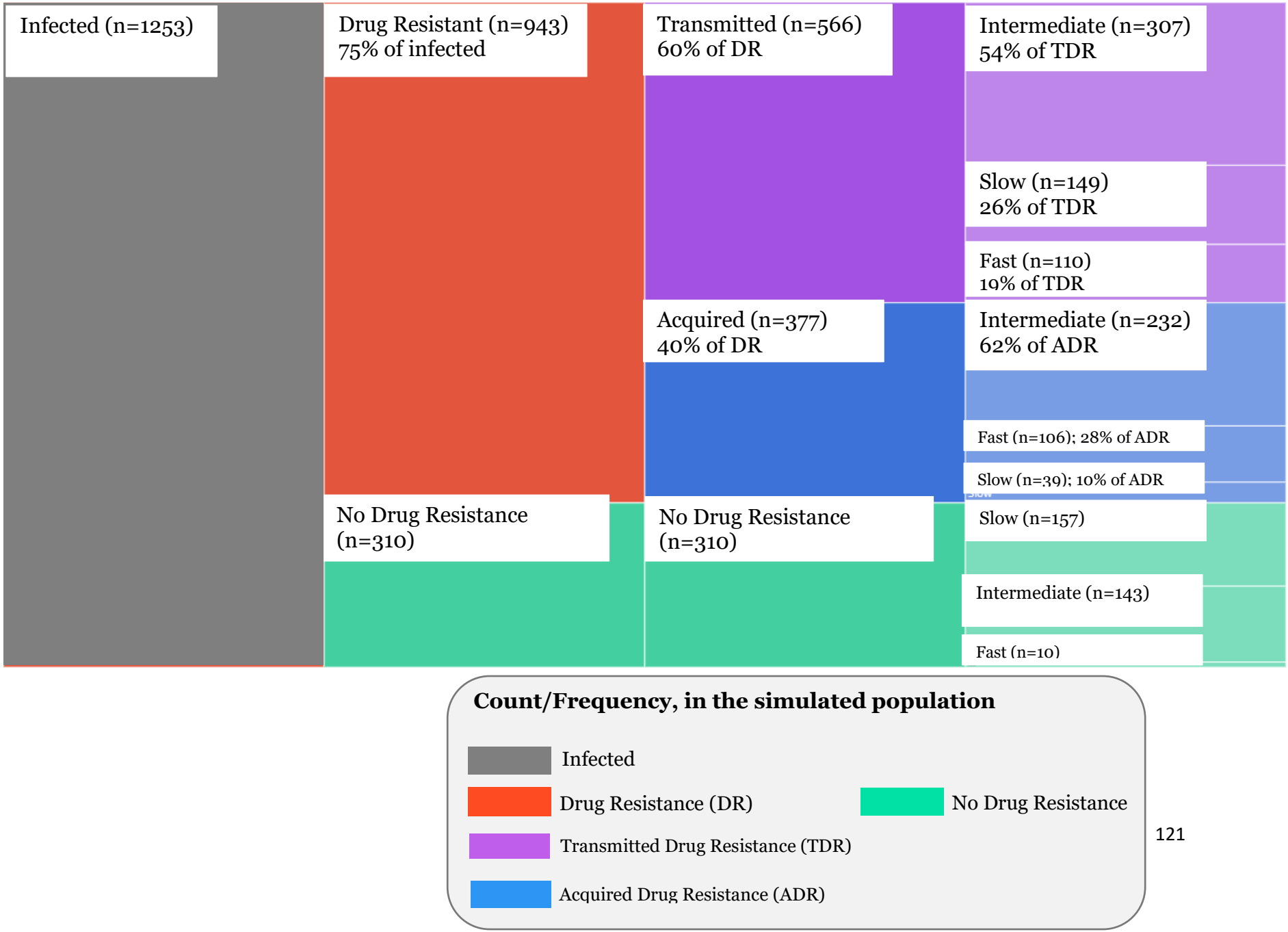


Figure 4.11 Democratic Republic of the Congo, Model 4 HIVDR Hierarchical Chart – PGx and DRM, included, with agents assigned intermediate, slow, and fast metabolizers.



Number of AIDS Deaths and Number of Newly Infected Individuals

The number of AIDS deaths and newly infected individuals demonstrates a similar rate of growth. For the Zimbabwean population, each of the models 1 to 4, the number of AIDS deaths increase between 1.55-fold and 2.41-fold. In parallel, the newly infected individuals increase between 0.74-fold and 3.30-fold. For the Congolese population, each of the models 1 to 4, the number of AIDS deaths increase between 1.57-fold and 2.75-fold. In parallel, the newly infected individuals increase between 1.90-fold and 3.16-fold.

Figure 4.12: Number of AIDS Deaths, , using the Zimbabwean and Congolese study PGx and DRM data

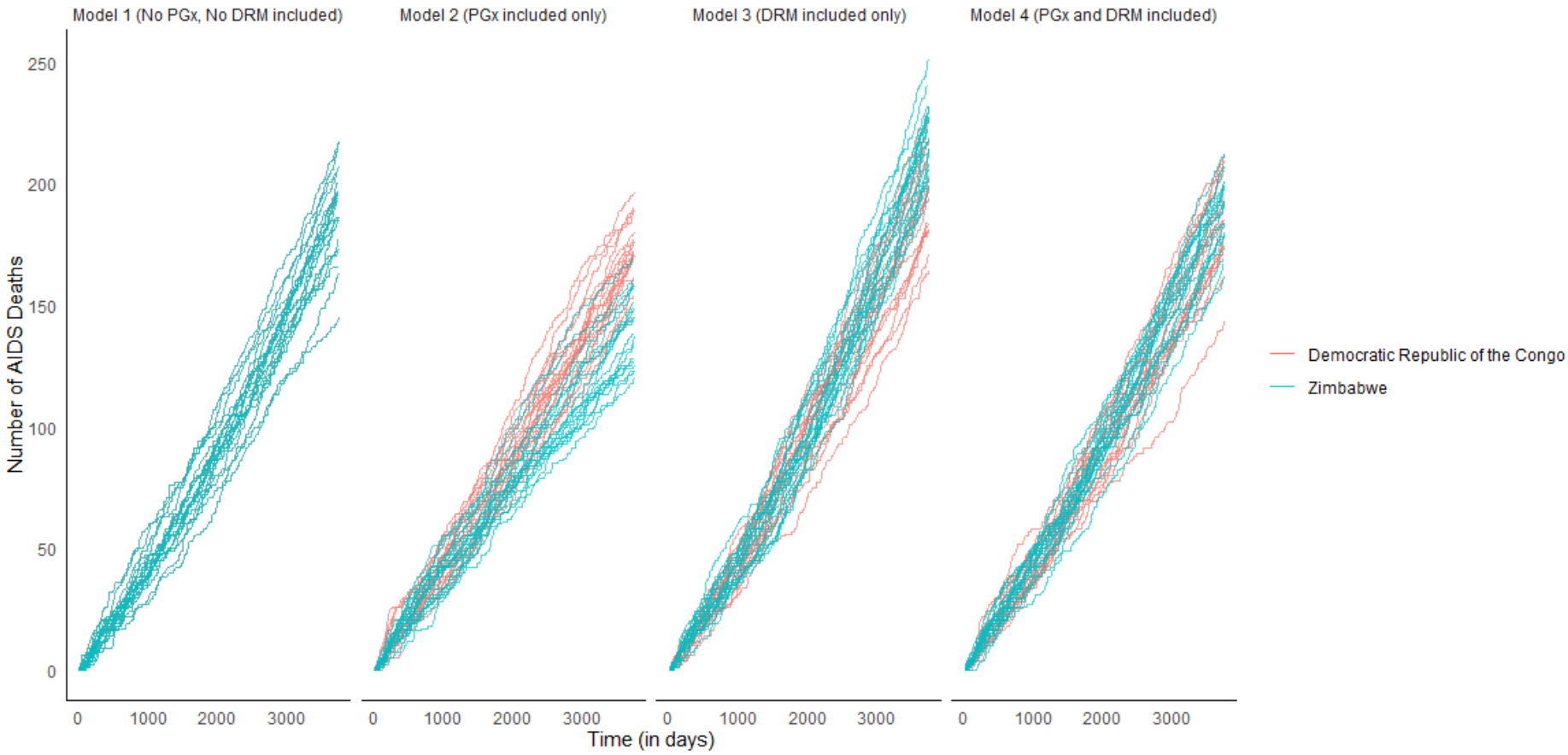
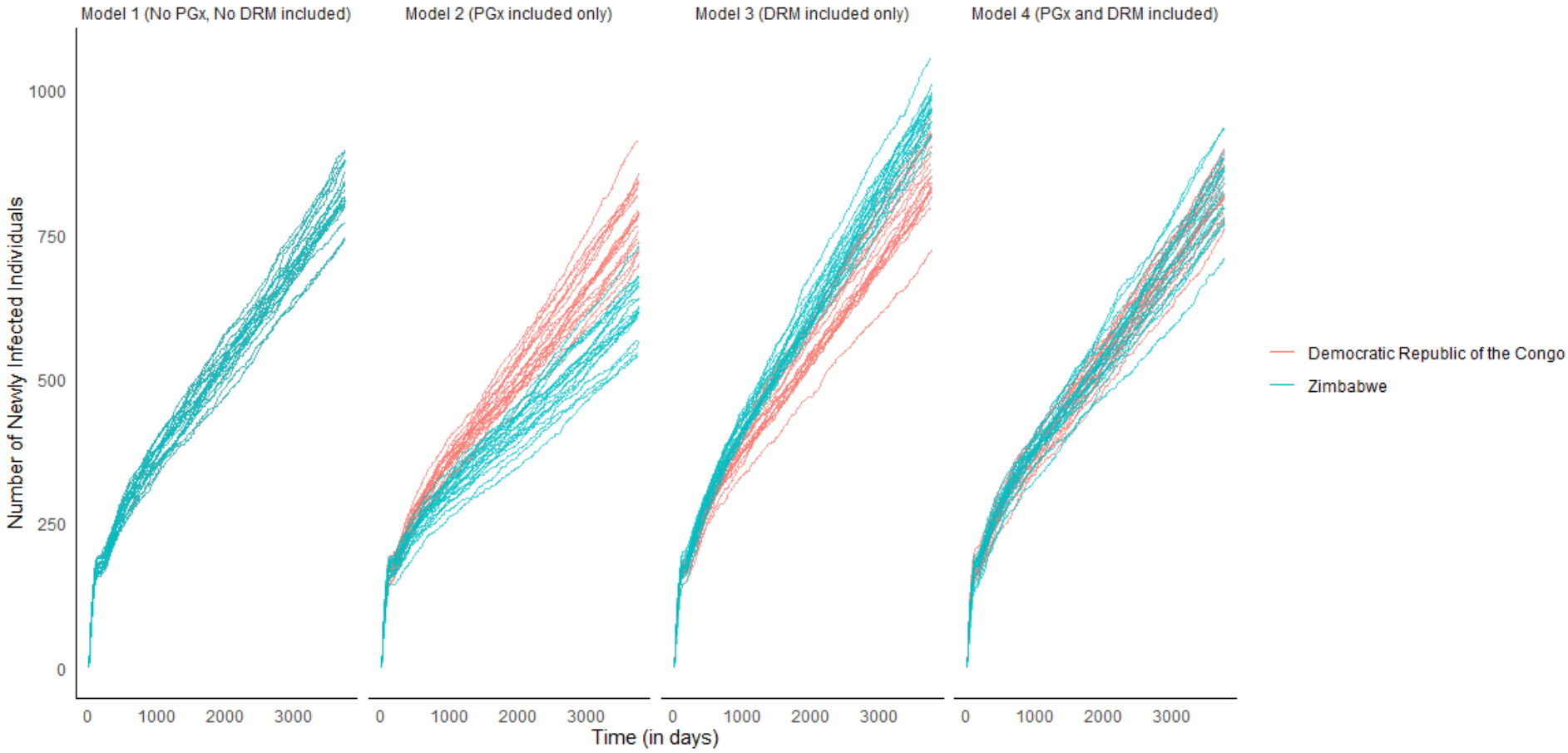


Figure 4.13: Number of Newly Infected Individuals, using the Zimbabwean and Congolese study PGx and DRM data



Number of Individuals w/ a Mutation at Specific Locus Positions (Greater than 1 copy/mL)

The frequency of the number of mutations at the five locus positions for each viral strain were tracked and computed over the span of a ten-year period for each of the four models. A trend indicating a significant increase in the accumulation of EFV- and TDF-related drug resistant mutants was evident in all four models and in both study populations. Within the first 200 days of the simulation, most of the viral population consisted of single mutant strains circulating in the population above 1 copy per milliliter. The cutoff of 1 copy per milliliter was used as an arbitrary number to capture at least 1 viral copy circulating within-host.

After treatment was introduced (i.e., after 200 days) the triple and quadruple mutations outcompeted the number of people with single and double-mutant strains. Table 4.5 details the number of individuals with specific viral mutations, for all four models among the Zimbabwean and Congolese population. The quadruple mutation, K65R+K103N+gTDF+gEFV was the most prevalent mutation for both populations and among all the models. For the Zimbabwean population, for model 3, when the prevalent K103N DRM was included in the model, the quadruple mutation increased exponentially after day 500 of the simulation, with a frequency in the population among greater than 1000 people by 2500 days, or 6.8 years.

However, for model 3, the Congolese population demonstrated an increase of 998 individuals with above 1 copy/mL for the K65R+K103N+gTDF+gEFV quadruple mutant by 3750 days, or year 10. For models 2 and 4, the triple mutants, K65R+K103N+gTDF and K65R+gTDF+gEFV viral strains were both found at low frequencies for each study population for the first 200 days of the simulation. However, the Congolese population had a significant increase in the number of individuals with the K65R+gTDF+gEFV viral strain, plateauing at ~200 individuals with 1 copy of the mutation from day 200 until day 3750 or year 10. Model 1 demonstrated no significant rise among individuals in the population with any other viral mutations, other than with the K65R+K103N+gTDF+gEFV viral strain, in both study populations.

The M184V mutation was acquired at low frequencies that did not outcompete the triple and quadruple EFV-related and TDF-related mutations. Specifically, the M184V mutation emerged at low levels of adherence and at low percentages of the viral

population (i.e., between 0.002-0.03%). These low frequencies of the M184V mutation may be due to the short half-life of the drug and viral fitness.

Figure 4.14. Zimbabwe study population – Number of Individuals with a mutation in the population– above 1 copy/mL
Model 2 – PGx included only

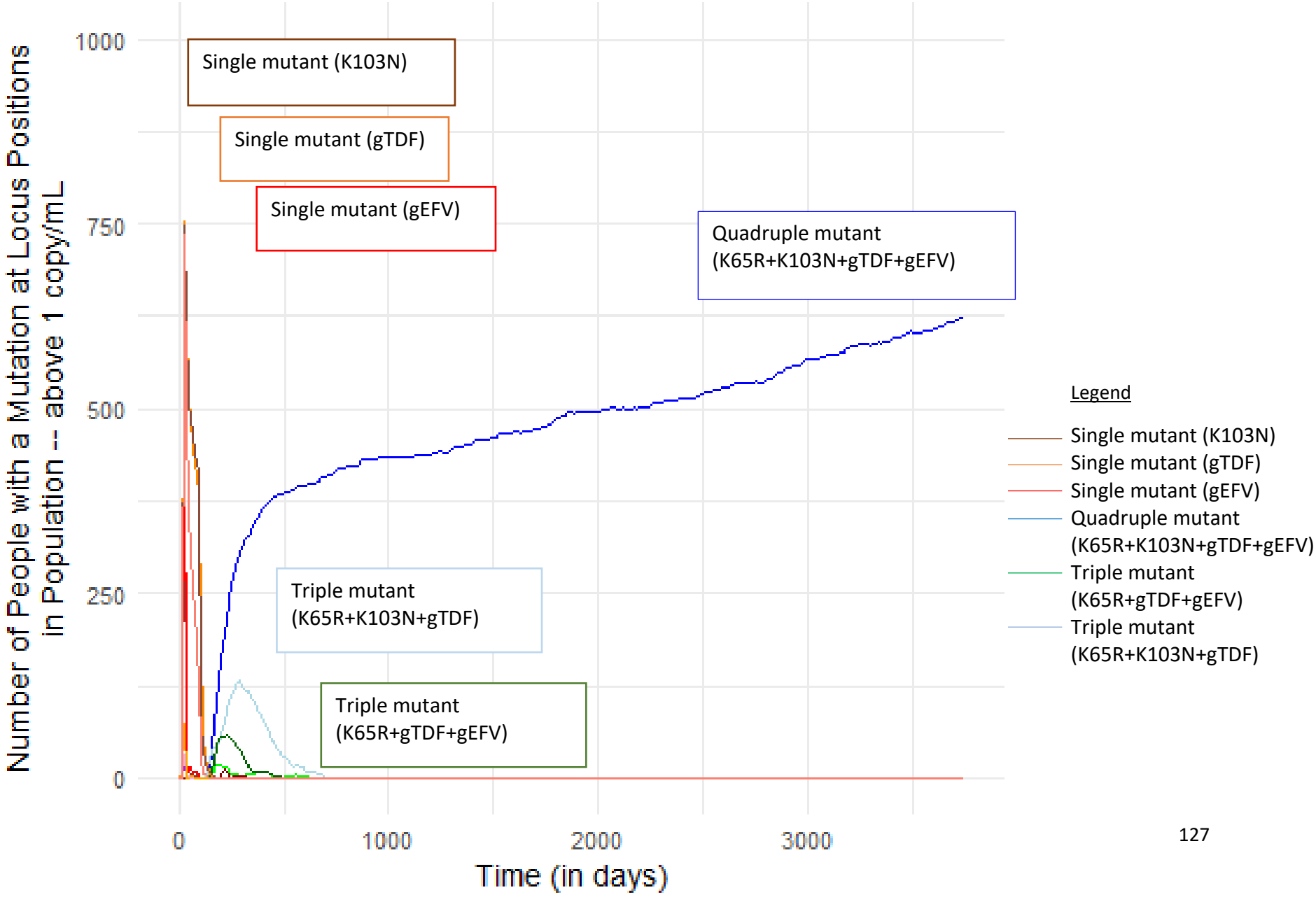


Figure 4.15: Zimbabwe study population – Number of Individuals with a mutation in the population– above 1 copy/mL
Model 4 – PGx and DRM included

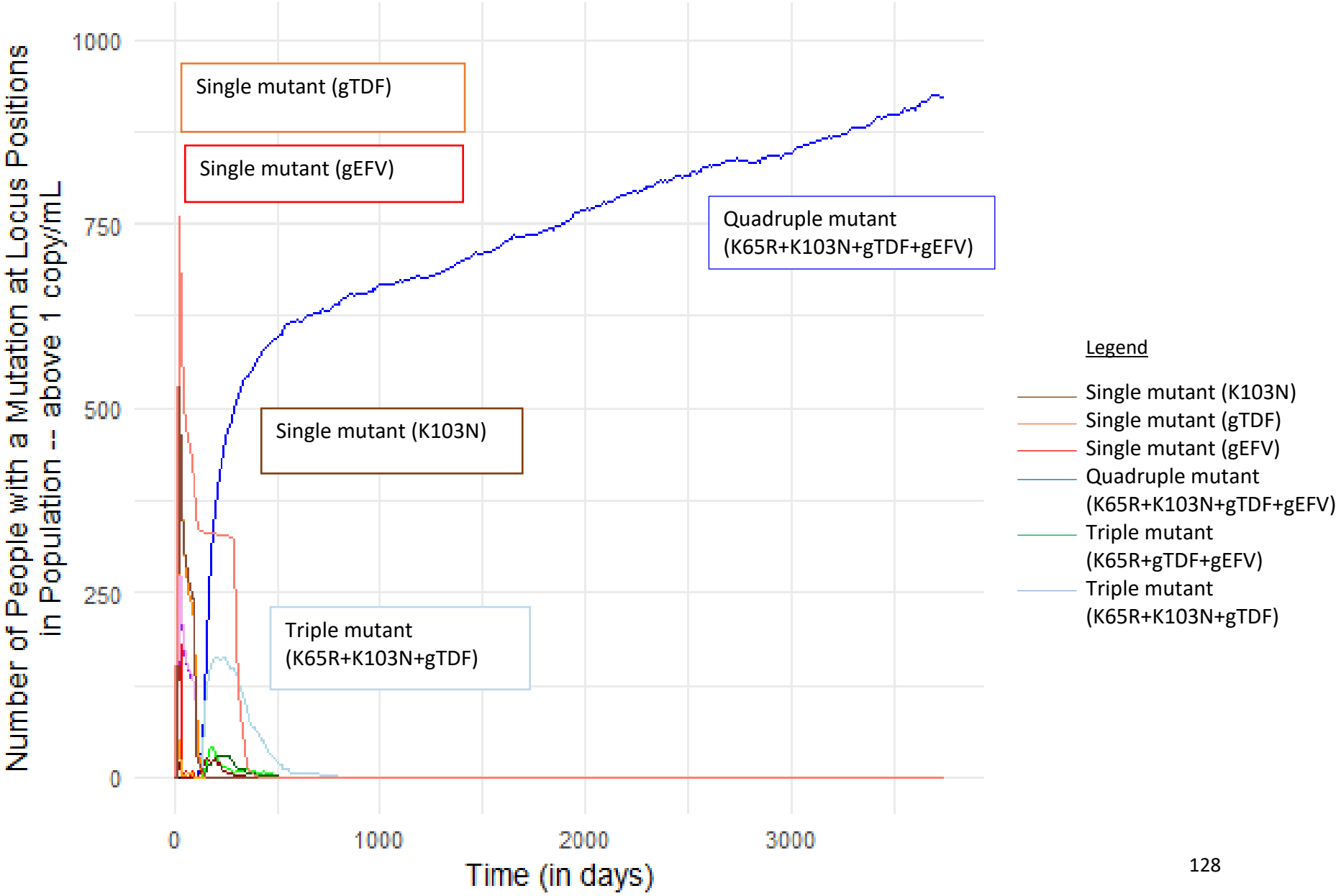


Figure 4.16: Congolese study population – Number of Individuals with a mutation in the population– above 1 copy/mL
Model 2 – PGx included only

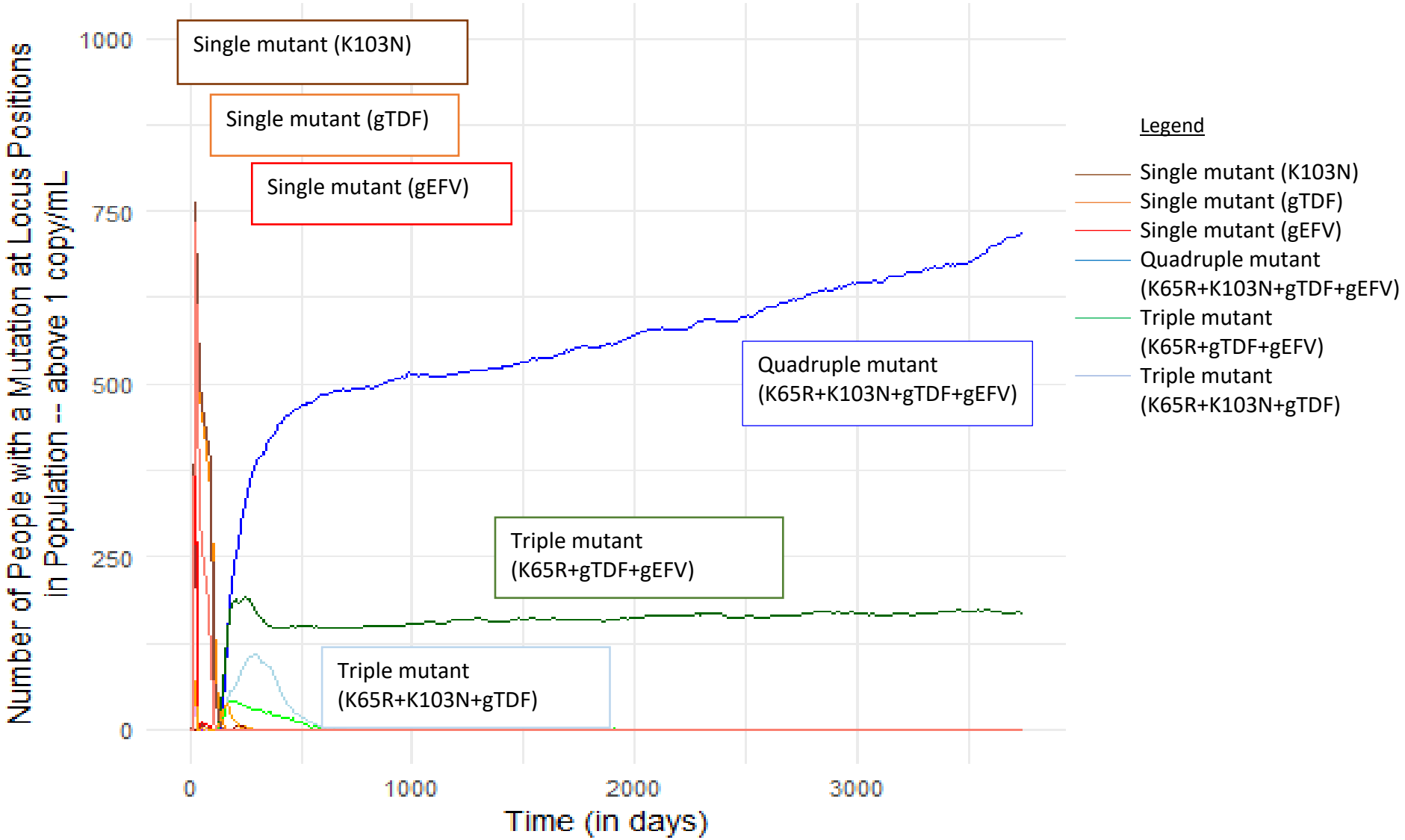
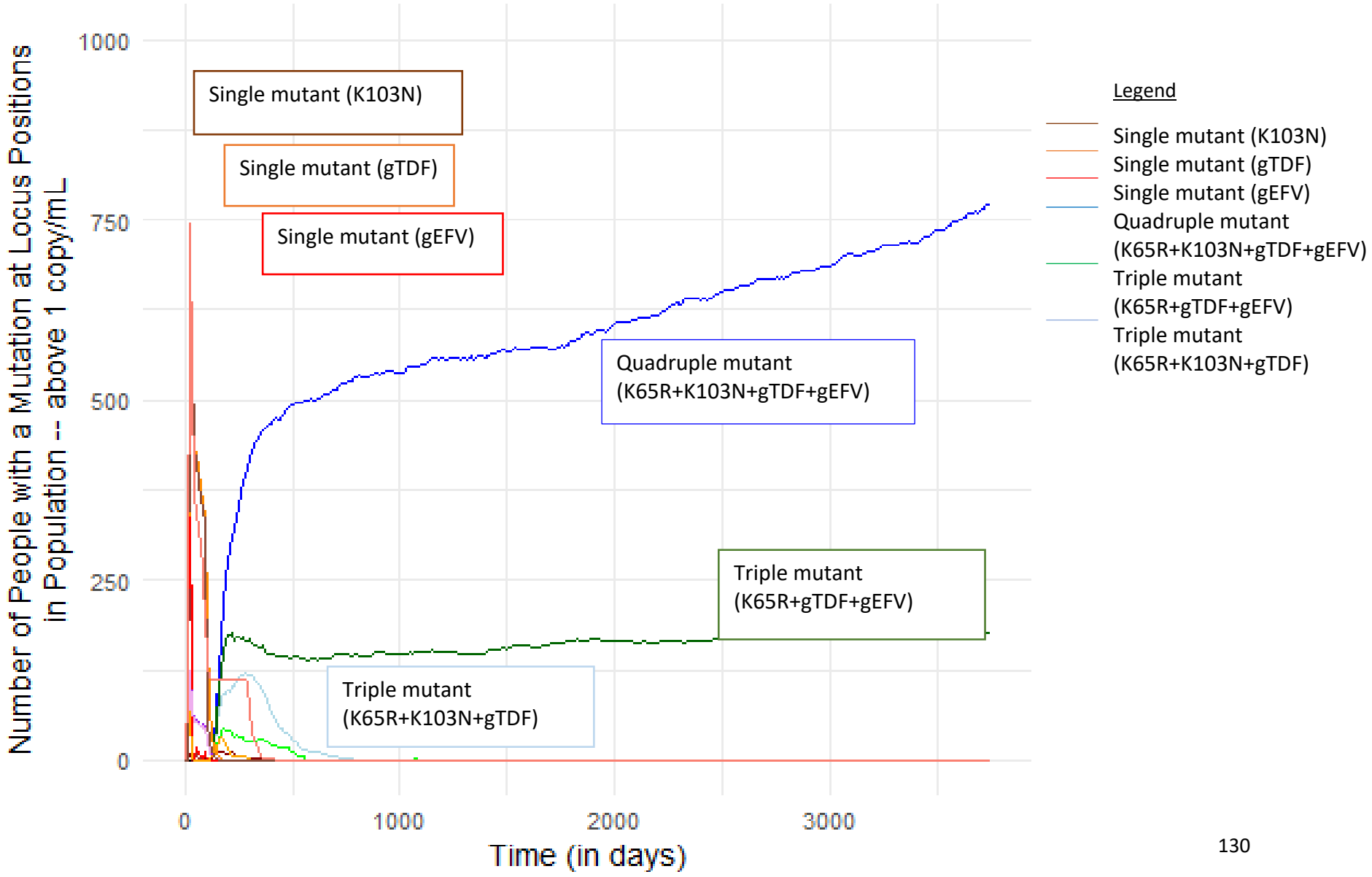


Figure 4.17: Congolese study population – Number of Individuals with a mutation in the population– above 1 copy/mL, Model 4 – PGx and DRM included



4.6. Discussion

The potential for populations to develop drug resistance was dependent upon proportion of individuals that either were fast metabolizers or had the presence of individuals with a K103N mutation. Populations that had a higher proportion of fast metabolizer demonstrated higher levels of percent resistance regardless of how many individuals had the K103N mutation. However, in a scenario where slow metabolizer are more prevalent in the population, the main concern was the proportion of individuals with a K103N mutation, as the slow metabolizers would effectively suppress the virus, although at toxic levels of drug concentration. Thus, tailoring interventions for fast metabolizers and individuals with a mutation like K103N may reduce drug resistance levels in a population, where both metabolizer and mutation data are known.

Among the four models, the models containing only a prevalent drug resistance mutation (i.e., models 3) demonstrated the highest levels of people with drug resistance, regardless of the proportion of people with the DRM, K103N, viral strain. When the pharmacogenomic parameters were included in the model (i.e., models 2 and 4), the estimated levels of drug resistance decreased significantly from the control model containing no pharmacogenomic metabolizer information and no prevalent DRM, due to the presence of the slow metabolizers. Slow metabolizers effectively suppressed the virus amongst those infected. However, slow metabolizers also created more supra-therapeutic levels of the drug EFV which leads to toxic drug concentration levels. The higher number of people with supra-therapeutic drug concentrations for PGx-related models with slow metabolizers in the Zimbabwean and Congolese study populations show the importance of targeting populations that may have a higher proportion of individuals as slow metabolizers. Thus, the Zimbabwean study population would likely have more individuals than the Congolese population to switch to a second-line regimen due to toxic drug concentrations. In contrast, the Congolese population would have a higher number of people with sub-therapeutic drug concentration levels for the drug EFV.

In my simulations, transmitted drug resistance was the dominant type of drug resistance among all model types. In particular, the percentage of intermediate, slow, and/or fast metabolizers were similar to the proportional values assigned at the

beginning of the simulation. For the Zimbabwean population, model 2 also demonstrated the highest levels of transmitted drug resistance in the population, among all other three models. Model 2 also contained approximately 50% of the population as slow metabolizers. The assumption is that slow metabolizers would not have acquired drug resistance as the virus would effectively be suppressed. However, given that within the first 100 days of the simulation, where treatment is not occurring, transmission happens rapidly. Additionally, each metabolizer increased steadily over the duration of 10-years, possibly stemming from individuals that were newly infected.

Lastly, among the viral population, the triple and quadruple mutations predominated in the population for individuals with at least one copy per mL, within-host. Specifically, the TDF- and EFV-associated drug resistant mutations, K65R, K103N, a general EFV and a general TDF mutants drive the drug resistant levels. As such, the number of people with a triple or quadruple mutant mirrored the accumulation of drug resistance over a duration of ten years, as demonstrated in both the hierarchical charts and four model comparisons.

4.7. Summary

The type of drug resistance, either acquired or transmitted can differ depending on several factors such as age, medication adherence, gender, and geographic location. This is particularly relevant when comparing resource limited countries and high-income countries, as disparities between acquired and transmitted resistance exist largely due to social factors. The model incorporating prevalent drug resistance mutations indicated a higher account of drug resistance in the population and thus, interventions involving testing for specific drug resistant mutations may still offer advantages such as reducing the number of people with transmitted drug resistance. However, using the DRM data alone, may also over-estimate the levels of HIVDR. Thus, finding more studies that have genotyped certain populations where HIV is most prevalent could assist in tracking the likelihood of transmitted versus acquired drug resistance in the population. In all, transmitted drug resistance is dominate among infected individuals with drug resistance. However, no clear evidence suggests that having a fast or slow metabolizer in the model increases or decreases the levels of transmitted drug resistance significantly.

In addition, having a sub-therapeutic or supra-therapeutic drug concentration in the transmitter can affect whether susceptible individuals get infected with a mutant virus and thus, has transmitted drug resistance. In the context of my overarching question "What strategies and conditions may assist in predicting, reducing, and preventing HIV transmission and HIV drug resistance within-host and between-host (i.e., in a multi-level system)?" and Aim 2 - model the effect of host genetics and drug resistance mutations on HIV-DR levels in two sub-Saharan African populations, the key findings were:

(1) Predicting HIVDR in the population is highest when incorporating a drug resistant mutation in the model.

(2) Depending on the proportion of metabolizers in the population, the model may estimate lower levels of drug resistance, which may be closer to empirical observations.

(3) SPVL also plays a role in the viral load rebounding and the accumulation of drug resistance both ADR and TDR.

These findings led me to explore the prioritization of second-line therapy and the question of Aim 3 of "which patients to prioritize for second -line therapy, given the availability of resources under specific cost and conditions?" Furthermore, given that HIV-DR levels in specific populations have risen above the WHO recommended level of 10% (as described in Chapters 2 and 4) and could stop efforts (i.e., like switching to second-line drug treatment) made to reduce HIV transmission and HIV-DR, the following questions are modeled in the next chapter (Chapter 5, Aim 3):

- Which optimization algorithm for treatment switching is best to use for finding policies to prioritize infected individuals, in the population, for second-line drug treatment?
- What parameter conditions of set point viral load (SPVL) and adherence levels yield the lowest levels of HIV-DR in the population when second-line treatment is either expensive or inexpensive?

Chapter 5: An Optimization Routine for HIV Treatment Switching

5.1 Introduction

In public health and medicine, optimization techniques and strategies can improve the impact of health interventions. Optimization is a process, action, or method that is functionally efficient for finding the best solution for a specific condition, scenario, or resource[190]. Using optimization many healthcare facilities and providers can improve quality of care, reduce costs associated with the demand and supply of healthcare resources, and other external factors such as the need for mobility of the patient to healthcare facilities and healthcare costs, in general. Within the context of computer science, optimization is defined in terms of a system of inputs and outputs, within a search space; as such, optimization is a process or algorithm used to find the minimal or maximum value of a function by selecting given a set of input values and constraints [191]. Mostly recently, public health studies have utilized optimization algorithms to improve the quality of biomedical missing data in electronic health records [192], the detection of cancer [193], the prediction of diabetes type II [194], and the prevention of the spread of infectious diseases [195].

In this chapter, I discuss: (1) basic optimization algorithms such as grid search, hill climbing, gradient descent, and simulated annealing, (2) health policy modeling and the impact of HIVDR on the economy, and (3) the research aims, hypothesis, approach, methodology and optimization results for the prediction of HIV population-based dynamics. The goal of aim 3 (Chapter 5) ("Create an antiretroviral treatment (ART) optimization routine that finds the best policy for a desired HIV outcome, using a stochastic, network-based model that modifies drug dosing and/or switch treatment regimens, given the patient's clinical and behavioral characteristics and the threshold level of a certain parameter in the model.") was to find the best policy for switching

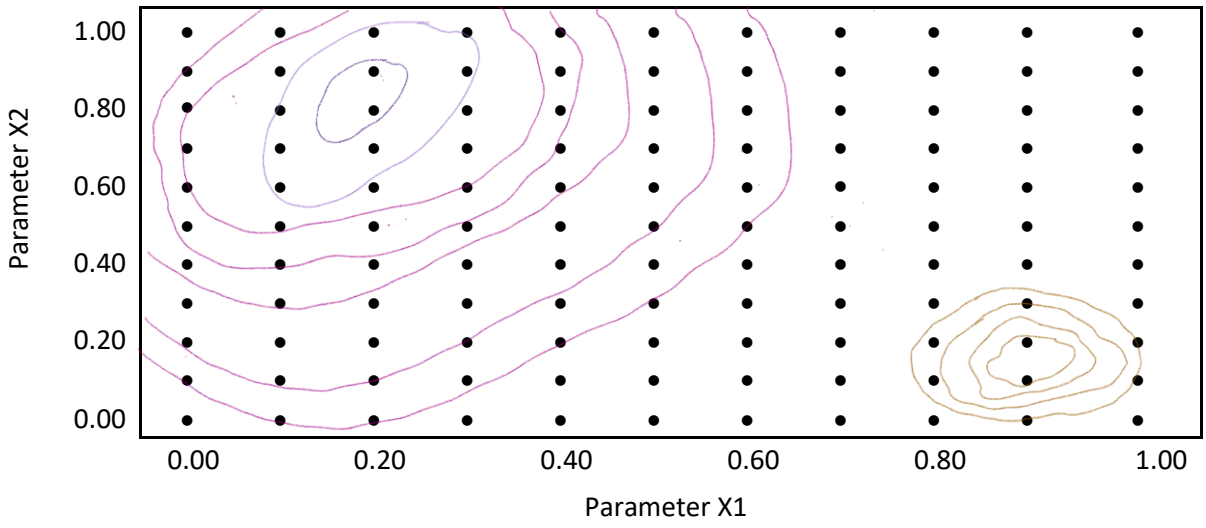
individuals to second-line treatment, given their individual attributes or parameter values such as levels of set point viral load and drug adherence) to achieve the lowest, or optimal HIVDR level for the simulated population. The optimization routine was developed using the within-host and between host modeling structure (i.e., the multiscale model) from Aims 1 and 2 to simulated HIV outcomes from a simulated population. During the simulation of the multiscale model, a computer programming algorithm that switched individuals based on whether they adhered to policy guidelines. Second-line treatment was generalized to act as an effective treatment that could suppress the virus better than the first-line treatment. Additionally, in aim 3 (chapter 5), I compared the benefits of using two optimization methods: the adapted grid-search method and simulated annealing. This is the final piece of my work answering the overall question of "What strategies and conditions may assist in predicting, reducing, and preventing HIV transmission and HIV drug resistance within-host and between-host (i.e., in a multi-level system)?"

5.2. Background

In chapters 3 (Aim 1) and 4 (Aim 2), I investigated the within-host and between-host attributes necessary in predicting HIV-DR and HIV transmission. Both aims indicated key parameters and parameter values that effected the increase or decrease of HIV mutants emerging, HIV-DR (within-host and in the population), and other dynamics like and specific mutations in the viral population, drug concentrations, number of infectious individuals, and the number of AIDS deaths. However, finding strategies and the best conditions that would likely prevent adverse health conditions and death, would involve more time to investigate each scenario, individually. Using modelling, researchers can conserve the resources needed to perform test and evaluate an intervention without jeopardizing the safety and harm of those infected or susceptible to infection. As such, using optimization techniques and algorithms for the outcomes of modeling routines of Aim 1 (the within-host model) and Aim 2 (the between-host model), would also create an environment that may conserve time and effort needed to find the optimal conditions for reducing, in this case, HIV-DR and HIV transmission. The next sections discuss several optimization algorithms that may assist in finding the best scenario for reducing HIV-DR and transmission.

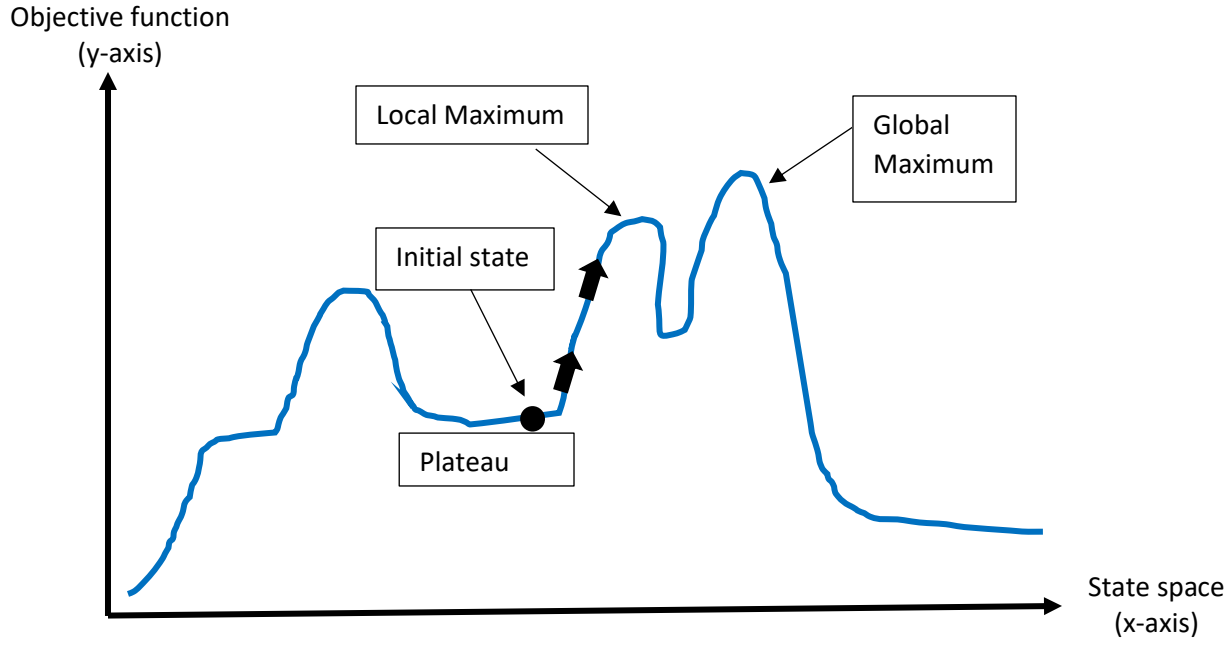
Optimization Algorithms: Grid Search. The grid search method is an exhaustive search algorithm that is commonly used to tune hyperparameters of a learning model[196]. Hyperparameters are parameters that are not learned by the model and are set prior to the training of a model [196,197]. The grid search method may also aid in determining the size of a neural network or learning rate for a model. The grid search method utilizes a lower and upper bound to explore all possible combinations of hyperparameter values and the associated model outcome when each value is in the model (Figure 5.1). The result of the grid search is to identify the best set of hyper-parameters and hyperparameter values that ensure that the model is both efficient, and accurate [196]. However, a limitation of the grid search method is that when the number of parameters included is large, then finding all possible solutions may result in a computational time burden [197]. Thus, for Aim 3, using a grid search method approach, where the parameters are chosen, but the values for each parameter combination is within a certain range, may aid in finding all the outcomes for each parameter combination, that would otherwise be entered manually.

Figure 5.1. A grid search using two hyper-parameters to identify the optimal value for the model indicated in purple with the brown circles demonstrating non-optimal solutions.



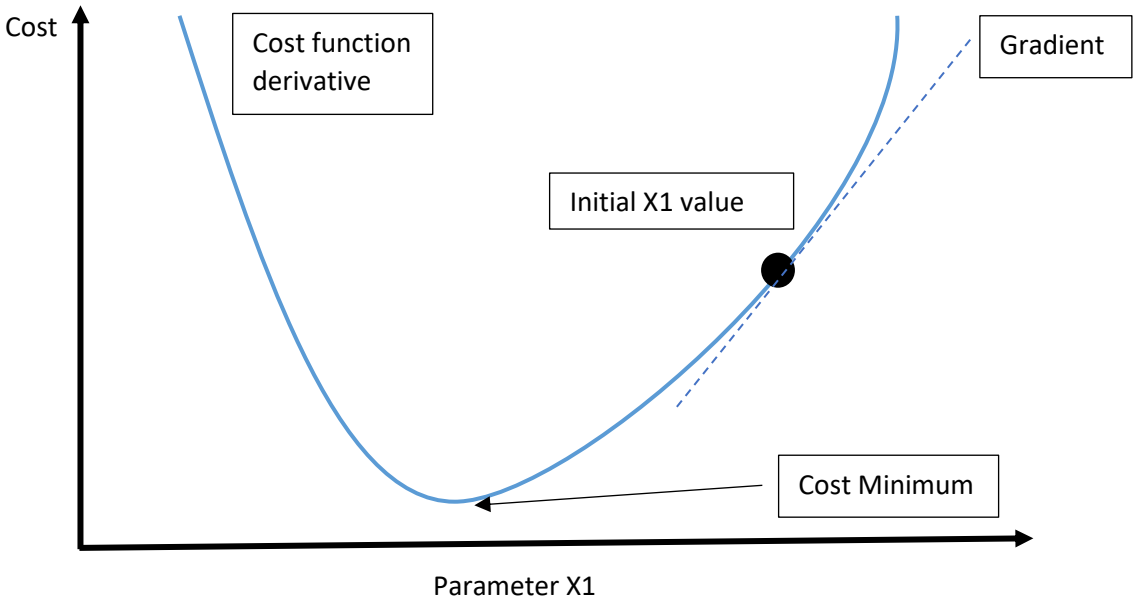
Optimization Algorithms: Hill Climbing. The hill climbing algorithm is a heuristic search and optimization algorithm. By definition, a heuristic is an approach or process for decision-making based on the history or previous outcomes[198,199]. The purpose of heuristics is to find the optimal solution to a problem without expending costly resources and time. Thus, a heuristic search such as hill climbing finds an optimal solution in a search space. Figure 5.2 illustrates the hill climbing algorithm. Hill climbing begins with an initial solution or current state and a goal state [198] .The current state is modified incrementally using an objective function then compared to a goal state to evaluate whether the current state has improved or moved closer to the goal state. If the goal state is achieved the algorithm quits and outputs the current state as the optimal solution, given the properties associated with the current state. Specifically with hill climbing, the final state is the peak of the search space [198]. Although hill climbing achieves finding an optimal maximum of a search space, there are several pitfalls to the algorithm that may result in finding a suboptimal solution such as reaching a plateau/ridge or finding the local maximum. In this chapter, for Aim 3, the objective is to create a treatment switching routine; however, drug dosing may also help with optimizing the values of a certain HIV outcome. For example, using the hill climbing method may also serve as an approach to finding the optimal drug dosing for certain medications in the model, while trying to achieve the maximum number of individuals that are suppressed.

Figure 5.2. Simple Hill Climbing algorithm illustration



Optimization Algorithms: Gradient Descent. Gradient descent is a commonly used machine learning and optimization algorithm that uses a cost function to evaluate the performance of the model. Gradient descent works by iteratively moving along a search space to find the local minimum of a cost function. The cost function works in a similar logic to the objective function, except the cost function is minimized, indicating descent[200]. In general, gradient descent incrementally changes the value of the parameters associated with the cost function by choosing steps away from the current point of gradient (i.e., slope) of the function. Figure 5.3 illustrates the concept of gradient descent. Thus, the gradient decent algorithm generally follows two steps that include: calculating the gradient or slope, then stepwise, moves downward along the gradient. This process happen iteratively until the minimum cost is achieved[200].

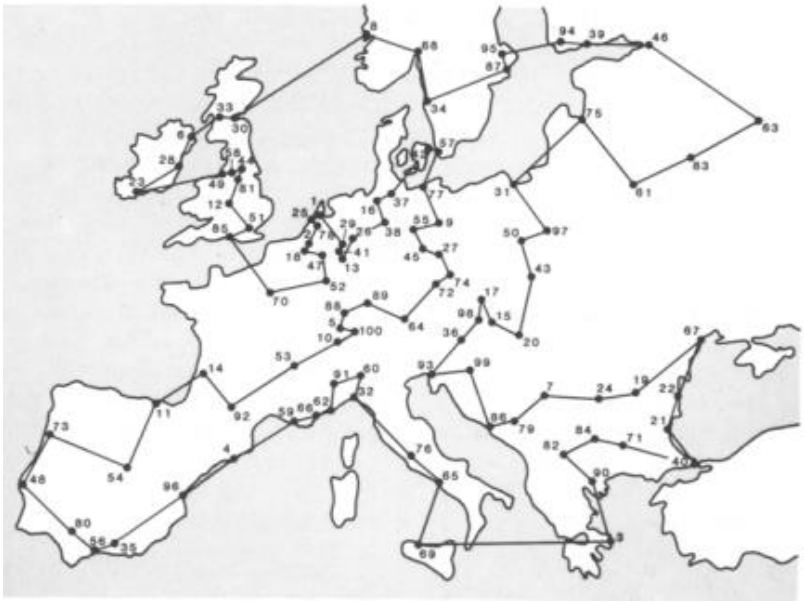
Figure 5.3. Gradient Descent illustration, where the blue line indicates the cost derivative, the dashed line represents the gradient, or slope.



Optimization Algorithms: Simulated Annealing. The simulated is a machine learning optimization algorithm that finds the global optima (i.e., minima) of the objective function. Thus, simulated annealing is typically used when the search space may involve many local minima. As such, simulated annealing performs similarly to hill climbing and gradient descent, except the algorithm modifies the parameters associated with the objective function randomly instead of determining if the next change should be based on the changes of the previous result of the objective function. If the modification to the parameter values of the objective function is closer to the solution (i.e., goal state), then the solution and parameter value set is always accepted with each iteration. However, if the modification to the objective function's parameter values do not improve the solution, then the solution may still be accepted with using a probability value. This probability of acceptance also decreases for every modification that does not improve the solution. The concept of the simulated annealing algorithm is borrowed from the field of metallurgy, where the temperature of a metals is cooled slowly until a desired physical structure is produced[201–203]. In all, simulated annealing is used when other exact algorithms such as gradient descent perform poorly due to the complexity

(i.e., large combinatorial search space). Simulated annealing is a popular solution to the traveling salesman problem, where a list of cities and distances between the cities are known and the salesman is instructed to visit every city, only once and return to the original city, using the shortest route. Figure 5.4 illustrates the use of simulated annealing to solve the travelling salesman problem. Simulated annealing has also been used in other discipline such as healthcare and medicine[202,204,205]. Thus, for Aim 3, the simulated annealing algorithm may provide both the ability to find new solutions for predicting optimal HIV outcomes and the time efficiency needed to explore the parameter space.

Figure 5.4. Simulated annealing algorithm solution example for the traveling salesman problem, with 100 European cities from the study by Aarts et al. 1988. Each city is assigned a numbered node and the solid lines indicate the optimal route for the problem.



Pre-Treatment Drug Resistance (PDR) Effects on the Economy. Efforts to optimize resources is essential to not only reducing HIV transmission and HIV-DR, but also the economic burden of spending money to fund HIV interventions and strategic plans. In Aim 3, the objective is to create and develop a treatment switching routine that optimizes HIV outcomes; however, prior to building the optimization routine to reduce

HIV-DR outcomes, I aim to understand what impact HIV-DR and particularly pre-treatment drug resistance has on the economy and financial spending, in general. Pre-treatment drug resistance happens when an HIV positive person has drug resistance but have started treatment for the first time or are re-starting treatment[176,206].

The disparity between certain countries and across specific special populations have also highlighted the importance of understanding how access to healthcare and the increasing global trends of HIVDR relate to economic stability. PDR has lasting effects on the health of a population and the ability of that individuals in the population to afford other second- and third-line regimens. PDR also impacts HIV programs and budgets, creating a cost burden that otherwise would not exist if PDR was not prevalent in the population of those diagnosed HIV and on ART. To this end, studies have outlined and displayed the severity and urgency of combatting PDR. Phillips et al. 2019 modeled the impact of PDR prevalence in African countries meeting or exceeding 10% and predicted that an extra expenditure of $6.5 billion dollars in the cost of ART therapy caused by transmitted drug resistance.[207] The study also suggests that even levels slightly below 10% predicted approximately $5 billion dollars in ART therapy spending.[207]

Other studies have investigated the cost needed to surveil and test for drug resistance prior to administering ART and/or when a patient may be failing treatment, especially in areas where the epidemic is most prevalent and drug resistance testing is not common practice. Gachogo et al. 2020 found that the cost of HIVDR testing in Kenya would be approximately $271.78 per test compared to studies that estimate the cost to be between $55 to $550 (i.e., some costs do not include the cost of personnel) in the United States [208–210].Thus, there are obvious disparities in the cost of HIVDR testing between LMICs and HICs leading to the economic burden of treating PDR leading to virological failure that could create a surge in the epidemic globally if immediate action is not taken. However, HIV program planning, and preparedness needed to avert the costs of PDR and HIVDR is expensive to implement in-person and in certain situations, may be unethical. One of the best approaches to finding strategies that will be the most effective in combating HIVDR and PDR is mathematical and epidemic modeling. In the next section, I will discuss the approach, methodology, and results for the treatment switching prioritization and optimization routine.

5.3 Approach

The epidemic modeling system, EvoNet[211–215], was used to create and simulated HIV within-host and between-host dynamics for the research aims in this Chapter. Using the C programming language to improve performance for running the model, additional parameters were added to the EvoNet system to account for parameter threshold values and rule-based logic. An adapted grid search algorithm was used to find all parameter value combinations for predicting the HIV-population based dynamic outcomes and a simulated annealing algorithm package was also explored. These two optimization algorithms were used because both methods allowed for exploration of a greater range of parameter values and finding the global optima – in relation to simulated annealing. The R package, *optimization*, and function, *optim_sa()*, were used to improve upon the adaptive grid search algorithm due to the model parameters values creating a large search space. The following parameters were chosen for determining prioritization of groups in the treatment switching optimization routine: adherence, set point viral load, viral load, fast metabolizer status, slow metabolizer status, single mutations, double mutations, triple mutations, and drug concentration for each of the three HIV medications (tenofovir, lamivudine, and efavirenz). However, for Aim 3, only the adherence and set point viral load were explored in the optimization routine for treatment switching. Table 5.1 includes the list of parameters and logic used in the C program of the HIV model simulation. These parameters each were assigned a threshold value, which given the logic or rule-based condition, would attempt to either switch or keep the agent in the simulation model on the first-line drug regimen of TDF-3TC-EFV, or the alternative highly effective medication (i.e., representative of an effective but costly medication, either financially or in terms of accessibility).

Table 5.1. EvoNetC modeling threshold parameters for the treatment switching optimization routine.

EvoNetC Modeling System, Optimization parameter	**Rule-based logic, executed in the C program**
Adherence threshold	If the agent has an adherence that is less than the adherence threshold, then the agent is switched to the second-line drug.
Set Point Viral Load (SPVL) threshold	If the agent has a spvl that is greater than the spvl threshold, then the agent is switched to the second-line drug.
Fast Metabolizer threshold	If the agent has a fast metabolizer status and the fast metabolizer proportion in the population is less than the fast metabolizer threshold, then the agent is switched to the second-line drug.
Slow Metabolizer threshold	If the agent has a slow metabolizer status and the slow metabolizer proportion in the population is less than the slow metabolizer threshold, then the agent is switched to the second-line drug.
Viral Load threshold	If the agent has a vl that is greater than the vl threshold, then the agent is switched to the second-line drug. The default value for the VL threshold was 1000 copies per mL.
Drug concentration for Drug 1 (TDF) threshold	If an agent has a viral load greater than 1000 copies/ml and the drug 1 concentration is less than the drug 1 concentration threshold, then the agent is switched to the second-line drug
Drug concentration for Drug 2 (3TC) threshold	If an agent has a viral load greater than 1000 copies/ml and the drug 2 concentration is less than the drug 2 concentration threshold, then the agent is switched to the second-line drug
Drug concentration for Drug 3 (EFV) threshold	If an agent has a viral load greater than 1000 copies/ml and the drug 3 concentration is less than the drug 3 concentration threshold, then the agent is switched to the second-line drug
Single mutation threshold	If the number of single mutation viral strains are greater than the single mutation threshold, then the agent is switched to the second line drug
Double mutation threshold	If the number of double mutation viral strains are greater than the single mutation threshold, then the agent is switched to the second line drug
Triple mutation threshold	If the number of triple mutation viral strains are greater than the single mutation threshold, then the agent is switched to the second line drug

5.4 Methodology

Treatment Switching Optimization Routine. I developed two algorithms based on the concepts of a grid search and the use of the *optim_sa()* function in R. Each algorithm contained epidemiologic, clinical, pharmacogenomic and drug specific

parameters for creating the HIV simulated epidemic using the EvoNet modeling system. The following parameters were used for each method of the treatment switching optimization routine:

Table 5.2: Model Parameters for treatment switching optimization routine.

Input Parameters	**Definitions**	**Default Values**
Model Parameters		
nsims	Total number of simulations	20
Epidemic Parameters		
n_steps	Total duration of the model	365 days
initial_pop	Total population size	3000 agents
initial_infected	Total number of infected agents	800 agents
Treatment parameters		
min_adherence	Minimum level of adherence for Drug1, Drug2, Drug3	0.90
max_adherence	Maximum level of adherence for Drug1, Drug2, Drug3	0.90
prob_eligible_ART	Probability of receiving ART treatment	1.00
start_treatment_campaign	Day to start treatment after infection	Day 100
Target_treated	Start prioritization of treatment switching	14 days, 31 days, and intervals of 60-90 days

I used a nested loop to create the adapted grid search method. The nested loop used the following two parameters: adh_threshold_2nd_line and spvl_threshold_2nd_line. Parameter constraints were also assigned with values for the adherence threshold of 0.0 to 1.0, incrementing by 0.05 and for the SPVL threshold of a log10 value from 2 to 8,

incrementing by 0.25. The following pseudocode details demonstrates the simple, nested for-loop:

1. Set the adherence range to 0 to 100%, by increments of 5%.
2. Set the set point viral load (SPVL) range of values from a log10 of 2 to 8,
3. Run the EvoNet model based on the selected values for adherence and SPVL

The optim_sa() function also used the model function, or cost function with given lower and upper bounds for the adherence and SPVL thresholds that mirrored the constraints of 0.0 to 1.0 for the adherence and 2 to 8 for the SPVL. The following code details the optim_sa() function using the adherence and SPVL threshold parameters.

```
evonet_sa <- optimization::optim_sa(fun = evonetc_run_optim_sa,
                                    start = c(0.5, 4.4),
                                    lower = c(0.00, 2.0),
                                    upper = c(1.0, 8.0),
                                    trace = TRUE,
                                    control = list(t0 = 7000,
                                                   nlimit = 100,
                                                   t_min = 0.09,
                                                   dyn_rf = TRUE,
                                                   rf = 0.5,
                                                   r = 0.3
                                                  )
                                     )
```

The simulated annealing controls of the *optim_sa()* function include variables such as *nlimit*, which restricts the number of inner loops where the parameter values change; *t_min,* which determines where the outer loop stops; *dyn_rf*, which if set to true ensures a wide search domain that will shrink over time, rf, or the standard deviation in relation to the parameter; *r,* which determines if the temperature will be reduced at the end of the outer loop.

In addition, the cost function for each method included the reduction of the following:

Cost function = Percent Resistance + Pill Count(1/weight),*

where the weight refers to how much you are willing to pay in number of second-line pills to reduce the HIVDR by a given percentage. A low weight means second-line pills

are expensive (i.e., less willing to give out pills). A high weight means second-line pills are cheaper (i.e., more willing to give out pills). The objective of the cost function is to:

- Lower the value of HIVDR in the population
- Lower total number of second-line pills given in the population.

Data Analysis. The results of each simulation for each optimization method were stored in a separate file detailing the threshold values for adherence and SPVL. The corresponding output for each model simulation parameter value set included:

(1) The second-line pill count of the alternative HIV medication
(2) Percent resistance in the population
(3) Number of infectious agents, and
(4) The Cost Function values for both lowering HIV-DR and second-line pills given in the total population

Contour plots for each method were generated based on the simulation results. The contour plots indicate which agents to prioritize based on gradient color schemes from a dark red (high value) to a dark blue (low value) associated with the parameter value modifications conducted for each model simulation.

Limitations. The limitations of the optimization routine included the limited set of methods developed that could incorporate the EvoNetC modeling system. For example, the optim_sa() function was developed as an R routine for a fixed equation and parameter set of 2. However, EvoNetC has over 175 parameters and approximately 25 affect the simulation model runs for Aim 3. Thus, inputting more than two parameters in the optim_sa() function cause the algorithm to crash or have a longer run time. Additionally, more optimization and machine learning techniques like gradient descent and hill climbing could have been used. However, we noted that these optimization techniques may have been better suited for conditions that involve changing drug dose, for instance. Other limitations included that the results reflected a simplified weight value, not based on economic data or actual access to treatment. Including econometric data would have more accurately depicted the trade-off for prioritizing specific groups of individuals based on their SPVL and adherence levels. In addition, the optim_sa (simulated annealing) results did not fully explore the search space for parameter

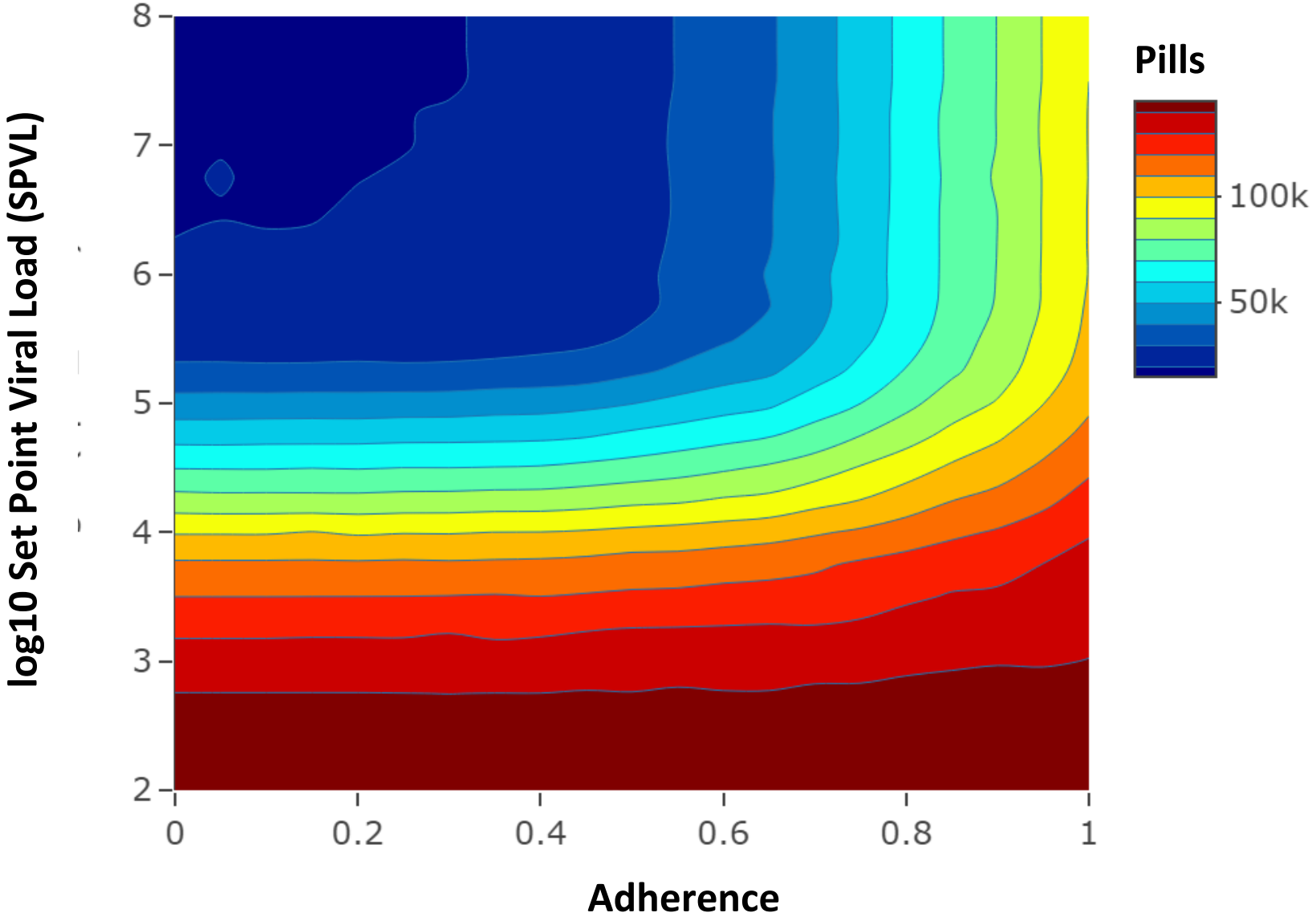
combinations leading to a narrow range in SPVL and adherence values simulated in the model. Thus, when using the optim_sa routine, increasing the range of values is important to ensuring the local optima (i.e., minimum) is found.

5.5 Results

Each optimization method indicated a group of individuals in the population that could be prioritized for the second-line regimen, given their host attributes such as their drug adherence level and set point viral load values. However, the contour plots also indicate the which agents that would benefit from treatment switching to the second-line regimen, thus lowering the cost function in scenarios when the second-line treatment is more or less expensive. The computational time taken to execute these two methods were 2.5 days for the grid search method and ~2 hours for the *optim_sa()* function or simulated annealing method in R/RStudio.

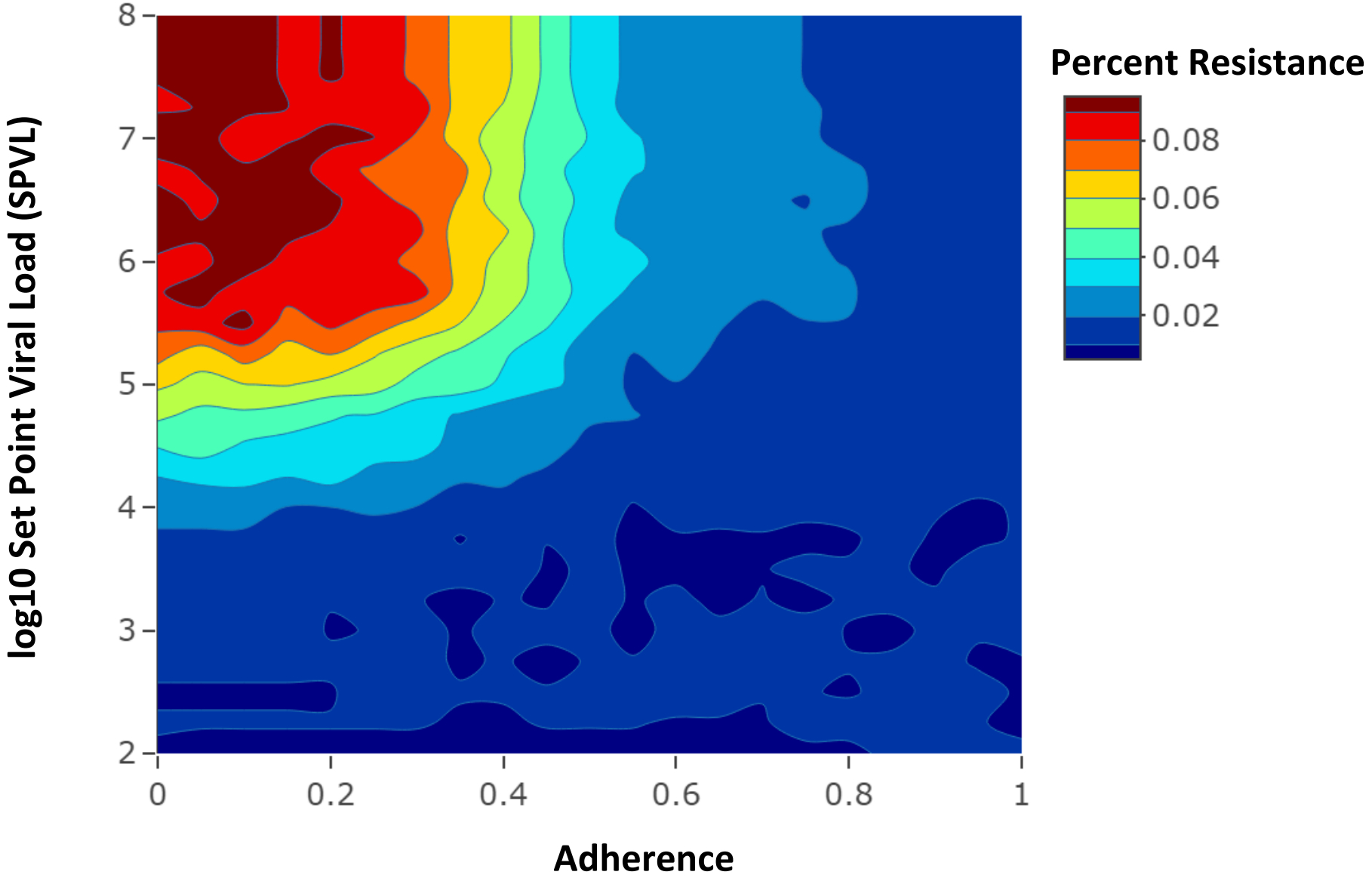
Using the Adapted Grid Search Method: Optimizing Number of Pills Given. Figure 5.5 shows that the number of second-line drugs given in the population over the duration of a year, given a set range of adherence level thresholds from 0% to 100% and a log10 SPVL thresholds ranging from 2 to 8. Using the grid method, the figure displays that when prioritizing individuals with adherence levels are below 0.30 or 30% and their log10 set point viral load is approximately 6 to 8, then the lowest number of pills are given. This may be due to the low number of people that had a high log10 SPVL within the simulation, based on the normal distribution of individuals with majority having an average SPVL between 4 and 5. As such, individuals within a high log SPVL of 6-8 are low in number among all individuals in the population. In contrast the highest number of pills were given if prioritizing individuals that had a low log10 SPVL ranging between a 2 and 3. This finding suggests that prioritizing individuals with a low log10 SPVL to switch to second-line treatment include the majority of people in the population of individuals that have a low SPVL. Also, people within this population may also live longer given their low SPVL may not acquire drug resistance while on first-line treatment and thus affording more second-line pills to be given in the population for those with higher SPVLs.

Figure 5.5. The total number of second-line drug pills given in the population for the duration of 365 days (contour colors) as a function of adherence threshold (x-axis) and SPVL threshold (y-axis), using a the nested for-loop/grid search method.



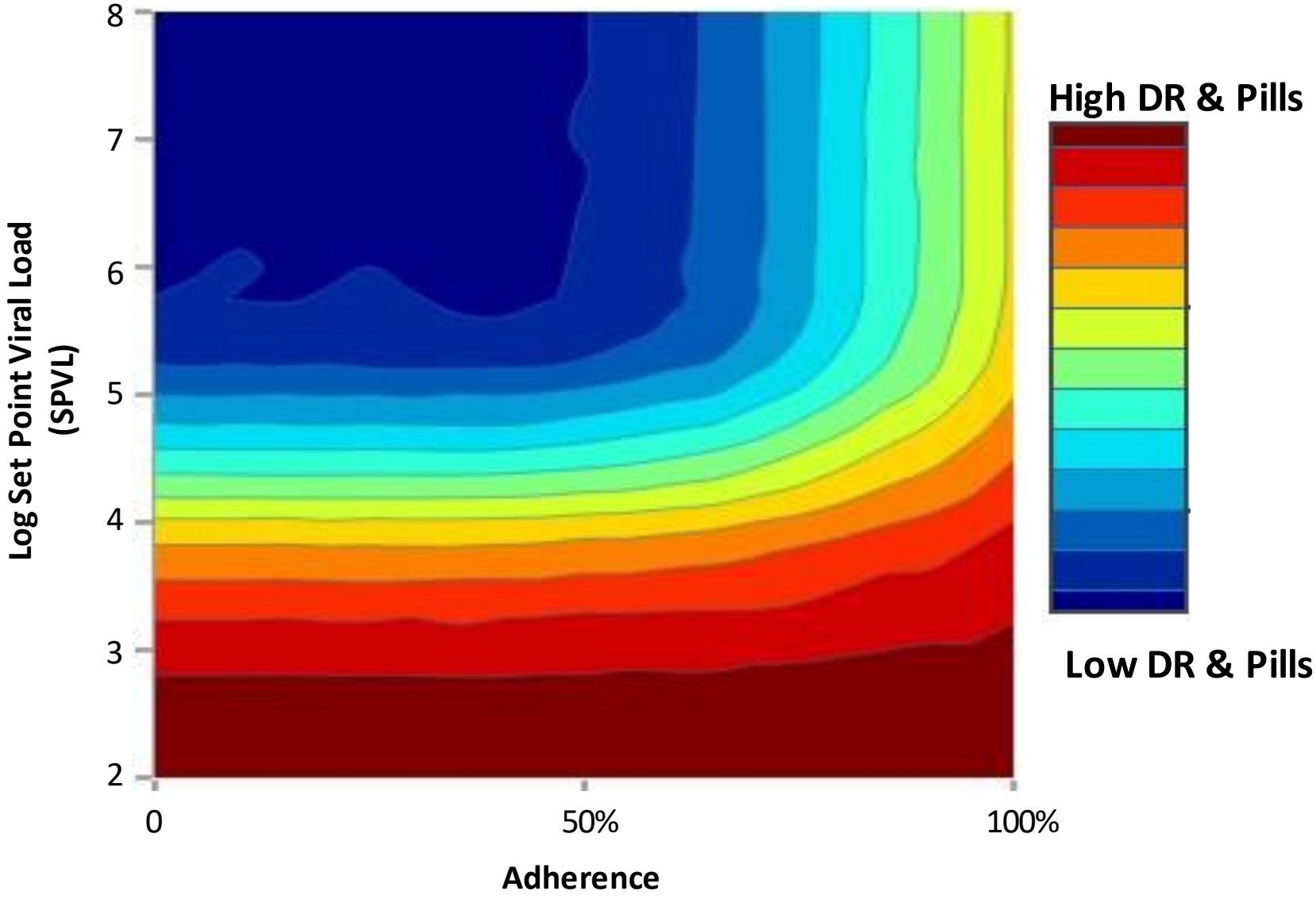
Using the Adapted Grid Search Method: Optimizing Percent Drug Resistance in the Population. Figure 5.6 shows that percent drug resistance in the population over the duration of a year, given a set range of adherence level thresholds from 0% to 100% and a log10 SPVL thresholds ranging from 2 to 8. Using the grid method, the figure displays that when prioritizing individuals with a log10 SPVL below 4, the percent drug resistance decreases to levels between 0% and 1% of the population. The lowest levels are seen, sporadically, along the wide range of adherence levels if the individuals prioritized were below a log10 of 4. However, these sporadic points showing the lowest levels of persistent resistance in the population may also be either imputed as a function of the contour plot or outliers due to the stochasticity of the model. Additionally, as the SPVL increases, those with a high SPVL above 5.0, would need to at least have an adherence level above 0.8 or 80%. In contrast, prioritizing individuals for second-line treatment with a high SPVL above 5 and below 0.20 or 20% resistance have the highest percent resistance. This finding suggests that when trying to lower percent resistance in the population based on SPVL and adherence, the individuals with the lowest chance of acquiring drug resistance due to viral dynamics, genetic factors, and behavior should be prioritized.

Figure 5.6. The proportion of drug resistance in the population for the duration of 365 days (contour colors) as a function of adherence threshold (x-axis) and SPVL threshold (y-axis), using a the nested for-loop/grid search method.



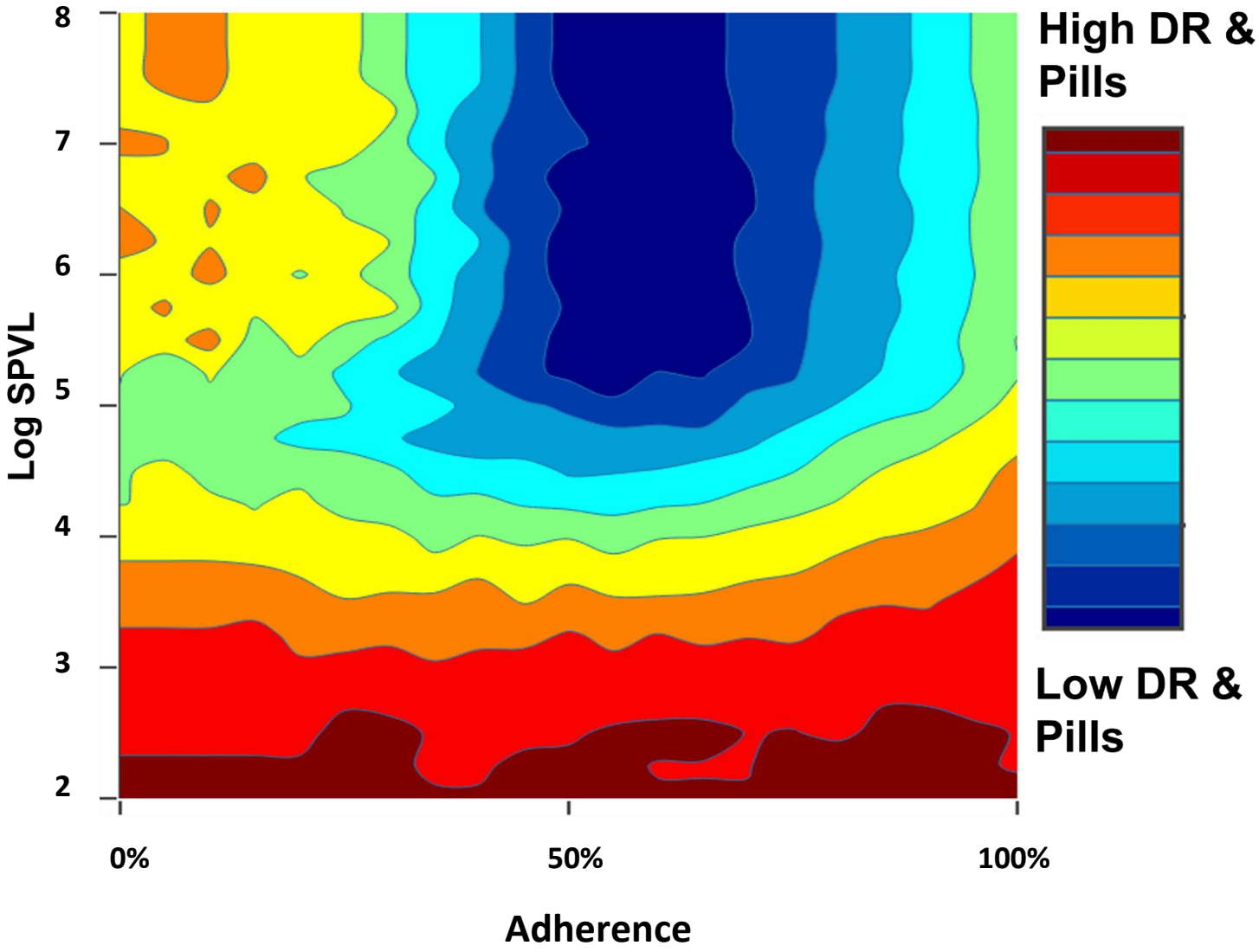
Using the Adapted Grid Search Method: Optimizing the Cost Function when Second-Line Pills are Expensive. As stated in the methods section, "weight" is an arbitrary number and refers to how much you are willing to pay to reduce the HIVDR by a given percentage, in terms of price and the value representing the willingness to administer pills (i.e., due to external factors such as stock-outs, stigma happening in the community, and other social factors). Figure 5.7 shows that when the cost function with a weight of 100,000, indicating that pills are more expensive, given the range of adherence level thresholds and SPVL thresholds, for the duration of a year. Figure 5.7 demonstrates that when prioritizing individuals with an adherence level of at most 50% and a log SPVL between 6 and 8 for second-line treatment, both the level of drug resistance and number of second-line pills given are low. Whereas the highest value of the cost function occurs when prioritizing individuals that had a low log10 SPVL ranging between a 2 and 3. Thus, when medication is more expensive, individuals with a high set point viral load are prioritized so that they may suppress the virus. Also, if there are less pills that are given in the population, then the group that is most likely to acquire drug resistance and have the lowest number of pills should be prioritized. Thus, prioritizing individuals that would output the lowest number of pills is most important, since in a scenario when second-line pills are more expensive, conserving medication to individuals that are at the highest risk of acquiring HIVDR is important to maintaining or lowering HIVDR in the population.

Figure 5.7. The cost function with a weight of 100,000 for the duration of 365 days (contour colors) as a function of adherence threshold (x-axis) and SPVL threshold (y-axis), using a the nested for-loop/grid search method.



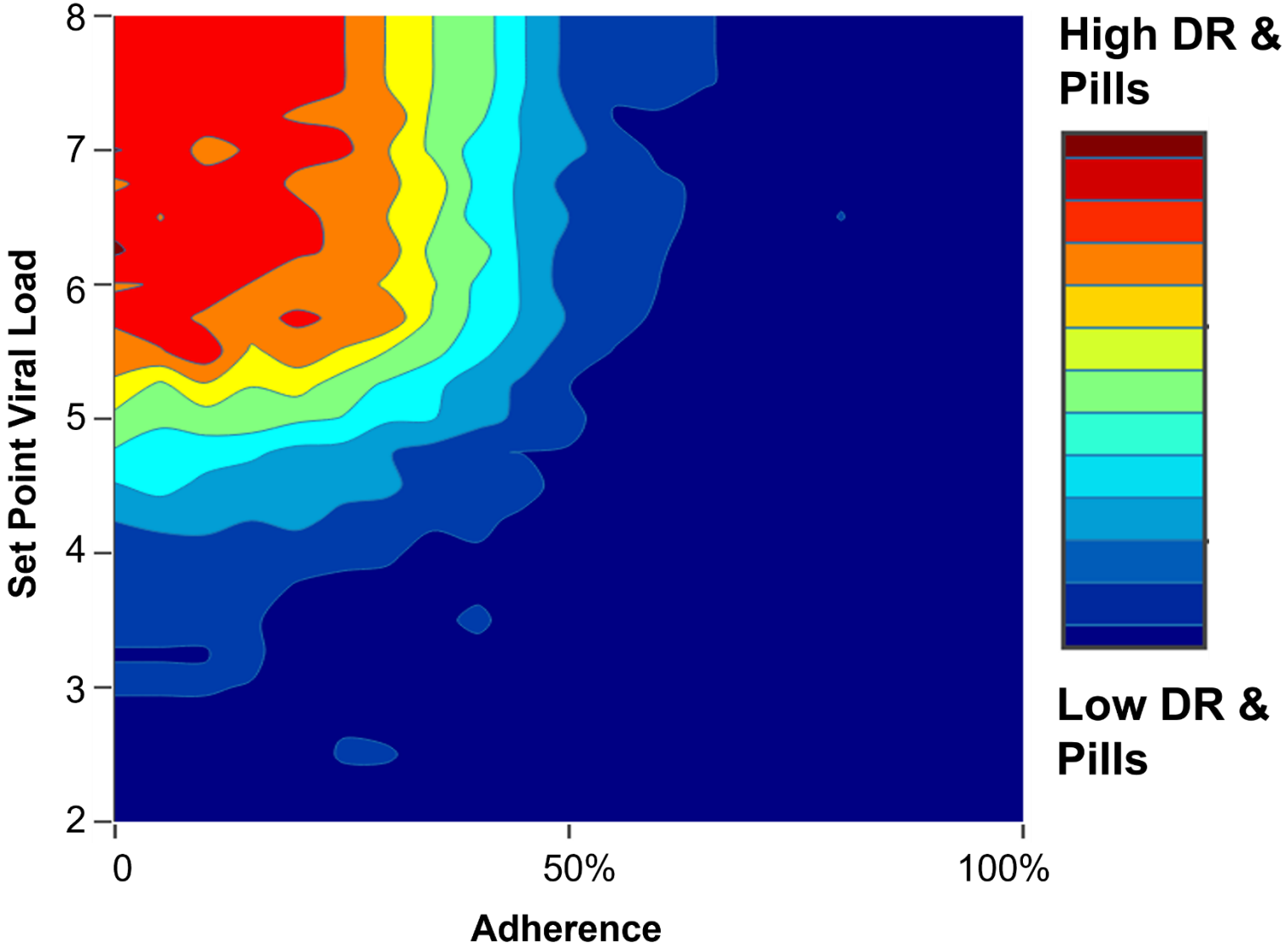
Using the Adapted Grid Search Method: Optimizing the Cost Function when Second-Line Pills are Moderately Expensive. Figure 5.8 showed that when the cost function with a weight of 1 million, there is a willingness to give more pills to the population, but pills are still expensive, given the range of adherence level thresholds and SPVL thresholds, for the duration of a year. The cost function is highest, indicating a high drug resistance and high second-line pill count, prioritizing individuals that had a low log10 SPVL ranging between a 2 and 2.5, regardless of adherence level. The cost function is lowest when adherence level of at most 50-80% and a log SPVL between 5.5 and 8. Thus, more second pills can be given out. However, since second-line pills are still expensive, prioritizing individuals that are more likely to obtain HIV-DR if not switched to second-line treatment is important, as well as the likelihood that those individuals will adhere to treatment. This finding suggests that when pills are moderately expensive and there is a willingness to give out more pills in the population, switching individuals that have the highest chance of reducing HIV-DR within host and that are the most likely to acquire HIV-DR if not given an alternative regimen should be the priority of the health policy.

Figure 5.8. The cost function with a weight of 1 million for the duration of 365 days (contour colors) as a function of adherence threshold (x-axis) and SPVL threshold (y-axis), using a the nested for-loop/grid search method.



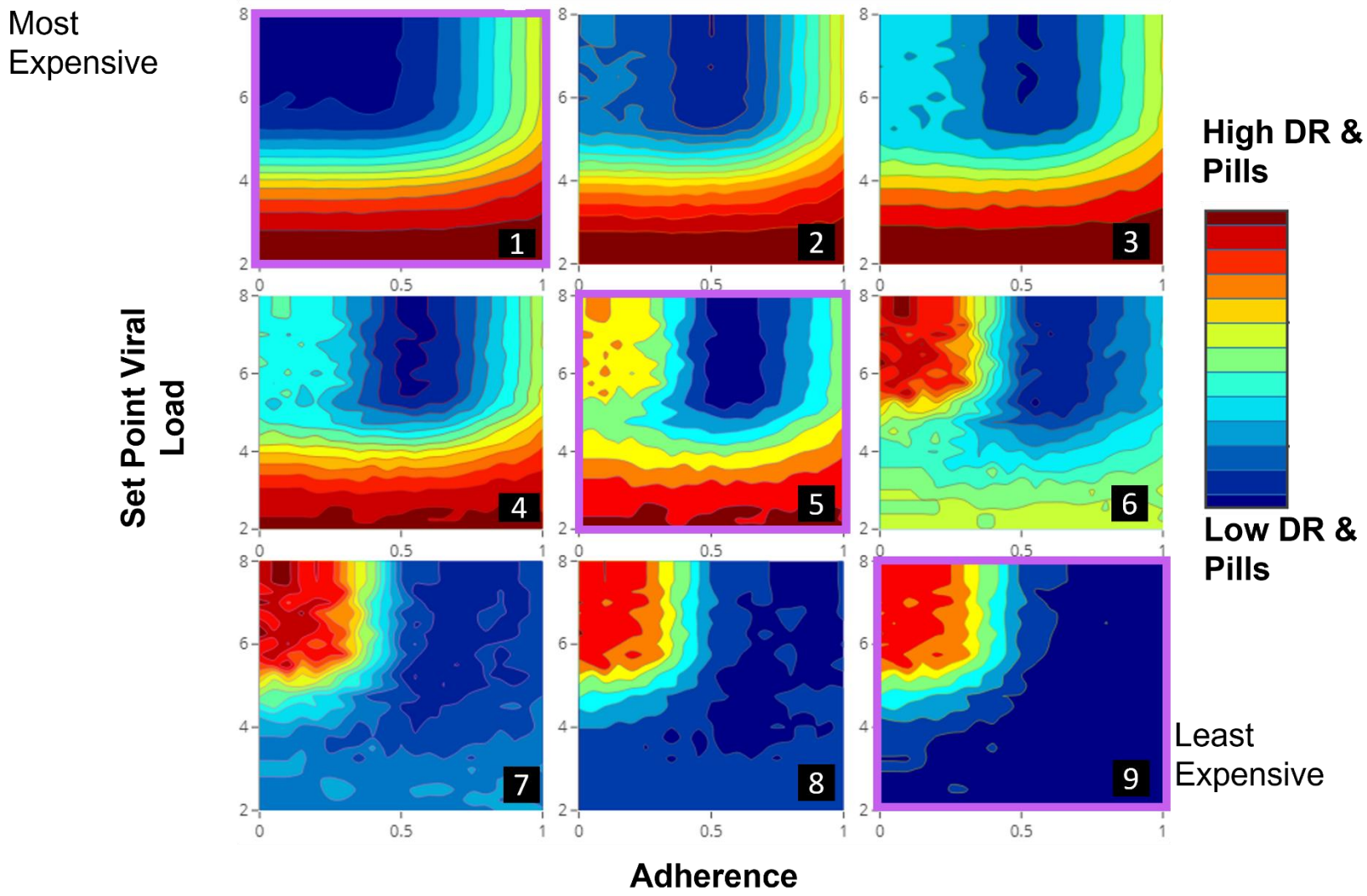
Using the Adapted Grid Search Method: Optimizing the Cost Function when Second-Line Pills are Least Expensive. Figure 5.9 shows that when the cost function with a weight of 8 million, there is a willingness to give more pills to the population, and pills are not expensive, given the range of adherence level thresholds and SPVL thresholds, for the duration of a year. The cost function is highest, indicating a high drug resistance and high second-line pill count, when prioritizing individuals that had a high log10 SPVL ranging between a 6.5 and 8, regardless of adherence level. The cost function is lowest when prioritizing individuals with a high SPVL of 5 or higher, but only when those individuals have at least an adherence of 45%. The cost function value is also lowest when prioritizing individuals who have a low SPVL between 2 and 3, regardless of the adherence level. This finding suggests that when second-line medication is inexpensive, more pills are given and thus, there is more opportunity to give effective treatment to a wider range of individuals that will likely not obtain HIV drug resistance. Thus, when pills are inexpensive prioritizing individuals that are less likely to obtain HIV-DR may be an effective health policy. Whereas prioritizing only those individuals with a high SPVL may prove to be ineffective when more second-line treatment can be given to more individuals in the population.

Figure 5.9. The cost function with a weight of 8 million for the duration of 365 days (contour colors) as a function of adherence threshold (x-axis) and SPVL threshold (y-axis), using a the nested for-loop/grid search method.



Using the Adapted Grid Search Method: Trends for Optimizing the Cost Function. Figure 5.10 displays a contour plot grid of the different scenarios starting with a condition where second-line treatment is most expensive (indicated by plot #1) to moderately expensive (indicated by plot #5) to least expensive (indicated by plot #9). The trend found among the nine plots included the following: as the second-line treatment gets less expensive the priority groups change and shift to add more individuals from varying ranges of low to high SPVL and adherence. Thus, the prioritization of allotting pills to the most at-risk groups in scenarios when pills are expensive seems to logically result in both low HIVDR and total pills given in the population. However, as pills get cheaper, the prioritization of individuals most likely to not acquire HIV-DR may be included, in addition to, other at-risk groups with conditions such as low adherence and high log SPVL. This suggest that health policies that allow individuals to switch to second-line treatment may varying depending on the scarcity and availability of the medication, in addition to, who may assist in the highest reduction of HIV-DR levels in the population.

Figure 5.10. A grid of the cost function displaying scenarios when the second-line pills are the most expensive in the population (plot#1) to when the second-line pills are moderately expensive in the population (plot#5) to when the second-line pills are the least expensive in the population (plot#9). Each plot is a function of adherence threshold (x-axis) and SPVL threshold (y-axis), using the adapted grid search method.



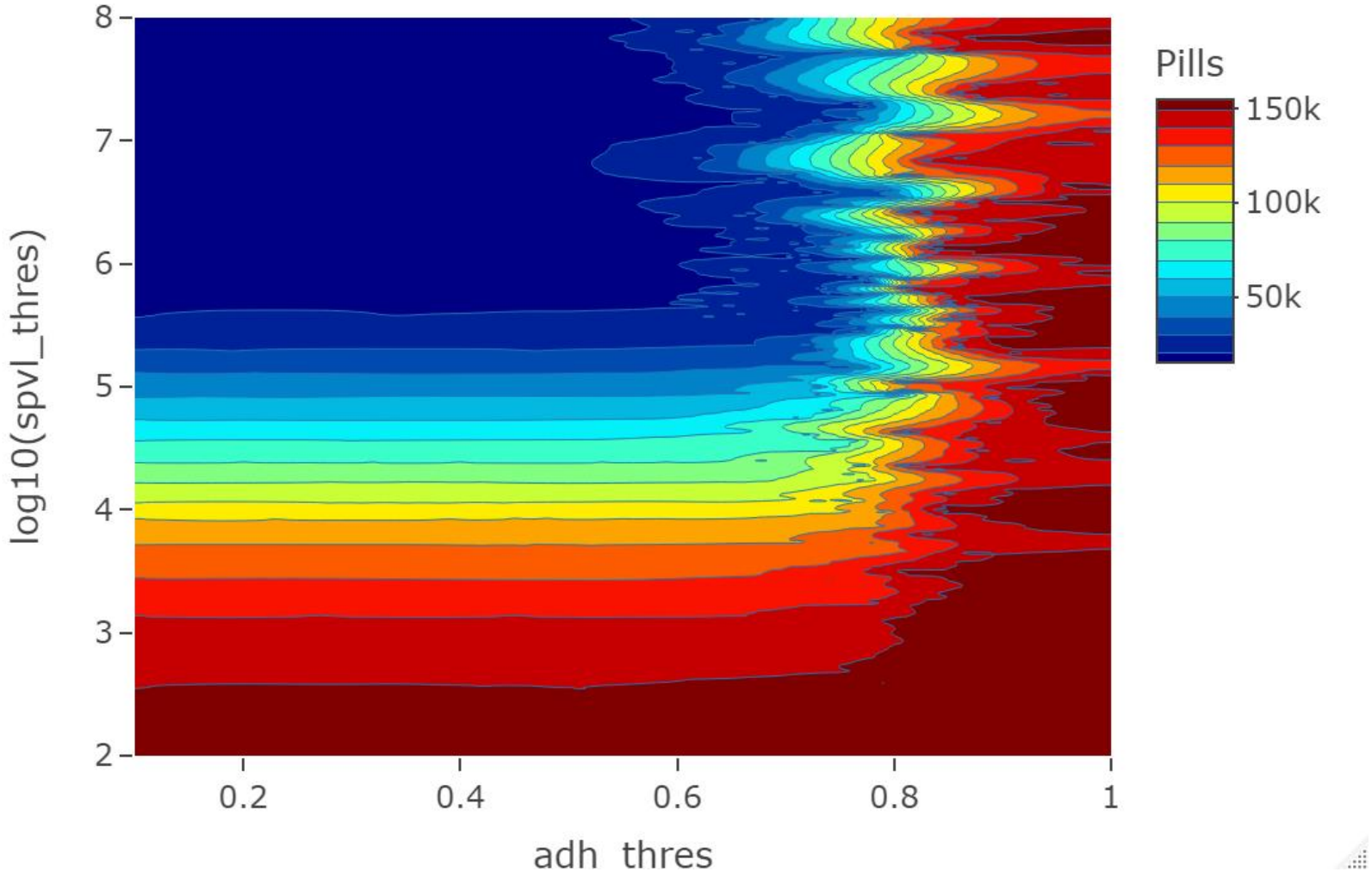
Investigations of simulated annealing

I next explored the use of simulated annealing to optimize the allocation of second-line therapy. Before showing the simulated annealing results, I show the EvoNetC data that our simulated annealing algorithm operated on. Doing so should help to reinforce the underling tradeoffs shown if Fig 5.10 and illustrate the nature of the fitness landscape that simulating annealing algorithm had to traverse.

Using EvoNetC data fed into the simulated annealing function, *optim_sa()* function: The Number of Second-line Pills Given, in the population.

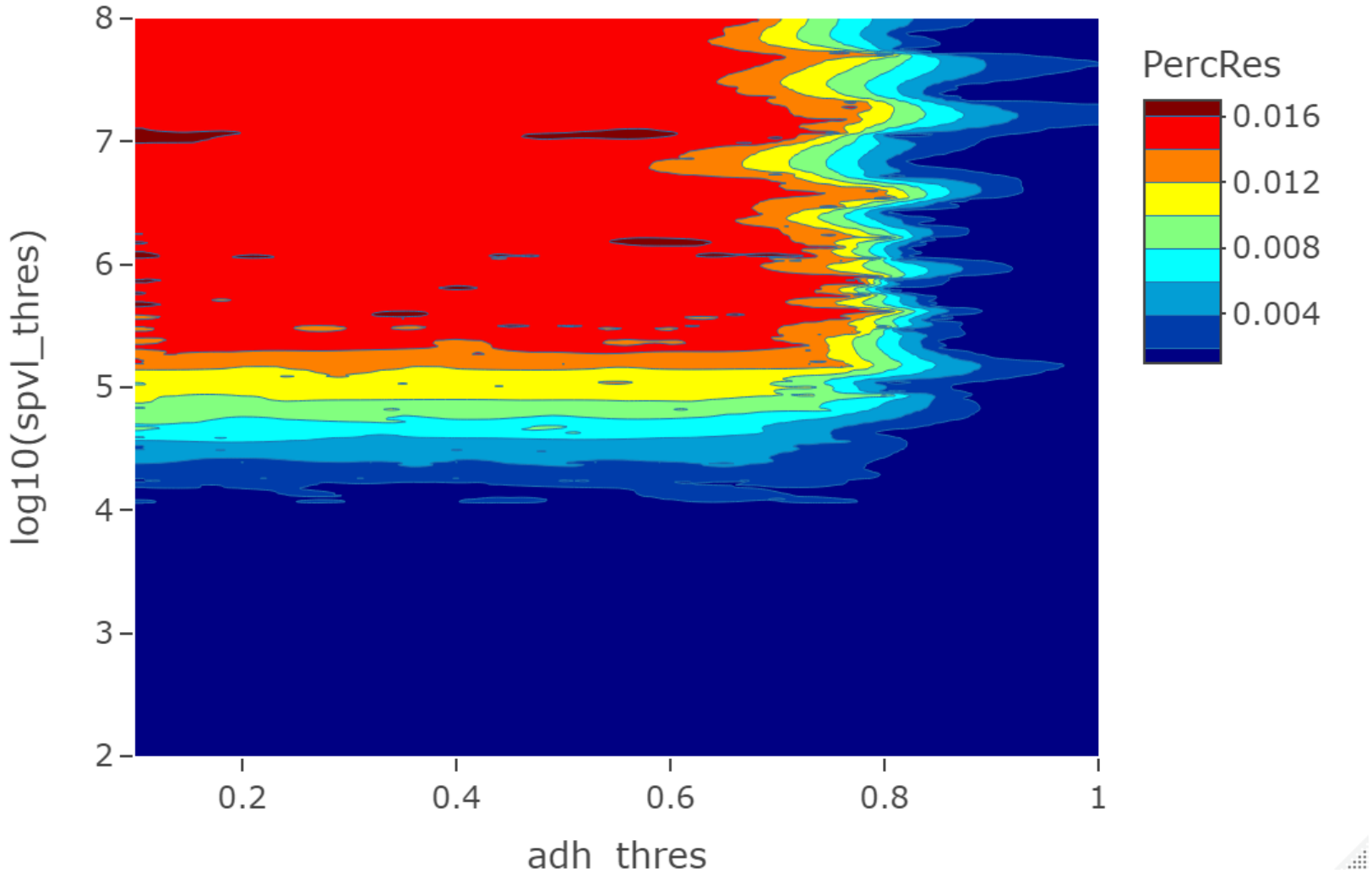
Figure 5.11 displays the predicted number of second-line pills given in the population over the duration of a year, given the range of adherence level threshold, that were utilized by the simulated annealing function, *optim_sa()* in Figure 5.14 below. This graph shows two extremes that capture priority groups for the lowest and highest number of pills. Prioritizing and switching treatment for individuals with a log10 SPVL of 5.5 to 8.0 and an adherence between 0 to ~60% have the lowest pill count for the population. Whereas individuals with an adherence between 0-1% and a log10 SPVL of 2.0-2.5 displays the highest pill count. This suggests that the simulated annealing algorithm may have similar results to the adapted grid search method that demonstrated similar results for the number of pills given in the population. In addition, within the model there may not be as many individuals with a high log SPVL between 6-8 due to the normal distribution of Assigned log SPVLs.

Figure 5.11. The total number of second-line drug pills given in the population for the duration of 365 days (contour colors) as a function of adherence threshold (x-axis) and SPVL threshold (y-axis) using the simulated annealing function, optim_sa() in R.



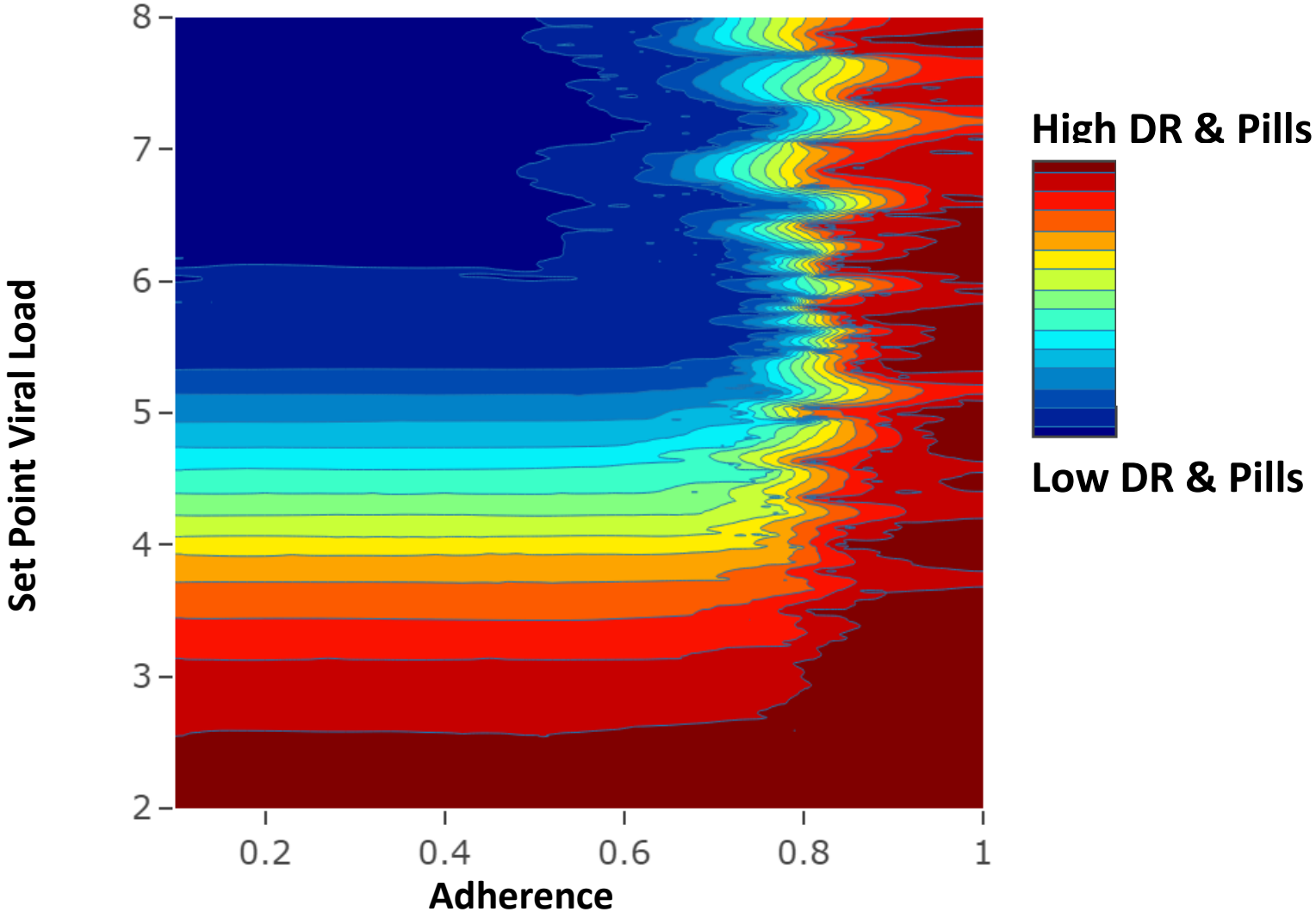
Using EvoNetC data fed into the simulated annealing function, *optim_sa()* function: Percent Drug Resistance, in the population. Figure 5.12 shows the number people with drug resistance in the population over the duration of a year, given the range of adherence level thresholds and SPVL thresholds, that were utilized by the simulated annealing function, *optim_sa()* in Figure 5.14 below. The dark blue color scheme indicates a low levels of drug resistance while a dark red indicates a high levels of drug resistance. This data shows a clear pattern: when prioritizing individuals have a log10 SPVL of greater than 5.2 and an adherence of ~75% or less then percent drug resistance in the population is at the highest level. In contrast, when prioritizing individuals have an adherence level between 0-80% and log SPVL of 5.2-8.0 or an adherence level of 80-100% and log SPVL of 2-8, the percent drug resistance in the population is at the lowest. This means that at high levels of adherence, prioritizing everyone regardless of their SPVL between 2 and 8, will benefit the population by lowering HIVDR levels.

Figure 5.12: The proportion of drug resistance in the population for the duration of 365 days (contour colors) as a function of adherence threshold (x-axis) and SPVL threshold (y-axis), using the simulated annealing function, optim_sa() in R.



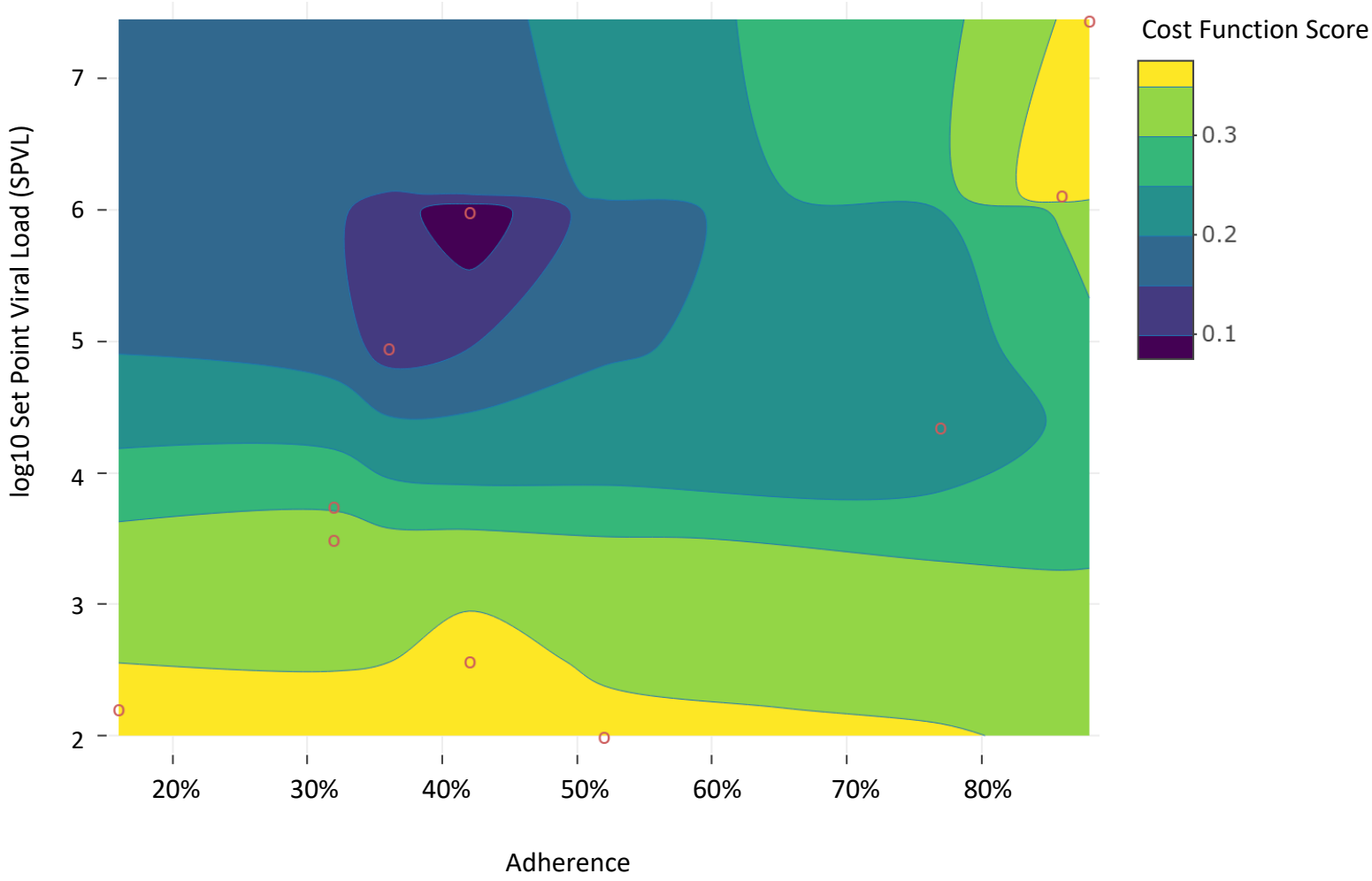
Using EvoNetC data fed into the simulated annealing function, *optim_sa()* function: Cost Function when Second-line Pills are Expensive. Figure 5.13 shows that the cost function with a penalty of 100,000, (indicating that pills are more expensive), given the range of adherence level thresholds and SPVL thresholds, for the duration of a year. There is a similar pattern to the contour plot that only included the number of pills given in the population. The upper left quadrant indicating a lower cost function is among those that have a SPVL of 5.0-8.0 with adherence level at or below 55-65%. Thus, in this case individuals that have a higher likelihood of having a viral rebound and drug resistance because they are poorly adherent to their pill-taking should be prioritized for a more effective treatment.

Figure 5.13. The cost function with a penalty of 100,000 for the duration of 365 days (contour colors) as a function of adherence threshold (x-axis) and SPVL threshold (y-axis), using the simulated annealing function, optim_sa() in R.



Simulated Annealing: Optimizing the Cost Function when Second-line Pills are Less Expensive. The contour bands in Figure 5.14 show the underlying cost function with a penalty of 100,000, indicating that pills are cheaper, given a range of adherence level thresholds and SPVL thresholds, for the duration of a year. The simulated annealing algorithm, optim_sa, indicated that cost was minimized when prioritizing individuals with a log10 SPVL of 6.0 and an adherence between 40-50% (red dot within the dark blue zone). This optimal value lies within the dark blue zone identified using the grid search method shown in Figure 5.7. While the simulated annealing algorithm quickly identified a spot within parameter space that leads to an optimal outcome, this speed came with a clear tradeoff: *optim_sa* missed key details revealed in the more exhaustive search shown in Figure 5.7. For example, optim_sa suggests that for individuals who are around 40% adherent that it would be more cost-effective to target individuals with SPVLs of 6 than those with SPVLs of 7, a clearly false prediction that is an artifact of the fact that optim_sa did not explore cases regions of parameter space with adherence < 80% and SPVL > 7.

Figure 5.14. The cost function score with a penalty of 100,000 for the duration of 365 days (contour colors) as a function of adherence threshold (x-axis) and SPVL threshold (y-axis), using the simulated annealing function, optim_sa() in R. The red o's represent values of adherence and SPVL for which the optim_sa calculated a cost function score. The "o" in the blue zone had the lowest cost function score.



5.6 Discussion

The adapted grid search and simulated annealing methods both identified optimum HIV-DR and the number of second-line medication given in the population. However, the computational time needed to execute each optimization routine did differ between the two methods. The adapted grid search method took the longest time to execute (2.5*24 thus 60 hours) for the treatment switching optimization routine; whereas the *optim_sa()* function only took a few hours (i.e., ~4 hours, or about 15x faster than the adapted grid search method) to complete. This time burden of the adapted grid search method may be due to large number of parameter value combinations that create a large search space. However, the simulated annealing method allows for a faster time in executing the optimization routine due to the ability of the simulated annealing algorithm to jump to different parameter value combinations in the search space as the algorithm approaches a desired value or optimum result. In this case, once the lowest levels of HIV-DR and number of second-line pills given in the population were achieved, the simulated annealing algorithm continued to search within a narrow window of parameter values close to the values found at the lowest levels of HIV-DR and second line pills given in the population until completion.

Both methods (the adapted grid search and simulated annealing) resulted in the reduction of HIV-DR and second-line pills given in the population. However, the adapted grid search method provided more health policy options for every scenario of expensive, moderately expensive in expensive second-line treatment or a willingness to pay/give out second line treatment in the population. However, this method also does not explore other values outside of the systematic parameter value combinations and may miss a parameter value combination that achieves a lower or the lowest desired outcome. In contrast, the simulated annealing method had a shorter and more narrow result of parameter value combinations of adherence and SPVL due to the ability for simulated annealing to randomly select parameter value combinations that achieve the desired result before narrowing in on a search space. Although the SA method provided concise results, these results were not similar to the findings of the exhaustive adapted grid search method. The difference in results also could be due to the random selection of values and stochasticity in the EvoNet model. However, increasing the number of

iterations in the simulated annealing algorithm may also ensure that more values are explored in the search space.

In terms of health policies found, the adapted grid search method demonstrated that in an expensive scenario, individuals with high SPVLs and low adherence levels should be prioritized. This is result may be due to the need to lower pills given in the population and HIV-DR, when there is less willingness to give second-line treatment. Thus, the health policy when second-line treatment is most expensive is to switch individuals at the highest risk of HIV-DR, virological failure, and the need for second-line treatment. Whereas, when pills are inexpensive or there is more willingness to give out second-line treatment, HIV-DR may be lowered by giving out more second-line treatment and switching treatment for individuals that may also be at risk to develop HIV-DR (i.e., individuals that have a high SPVL and at least moderate-to-high adherence levels and individuals with a low SPVL at any adherence level), but also have a higher likelihood of suppressing their viral load and thus not acquiring HIV-DR. The simulated annealing optimization routine had a similar pattern for when second-line treatment may be expensive or inexpensive. However, when second-line treatment is expensive prioritizing individuals with a very narrow range of high SPVL (i.e., 5.2-5.6) and adherence levels (5-10%) resulted in the lowest levels of HIV-DR and second-line treatment given to the population. Then, when second-line treatment was inexpensive, the criteria for individuals to prioritize in the population increased including those that had a high SPVL (i.e., 4.6-5.6) and an adherence of 5-50%. Thus, as the pills became cheaper, a wider range of people with low to moderate adherence levels were prioritized to lower HIV-DR levels in the population and second-line pills given in the population.

5.7 Summary

Optimization algorithms serve as a method and tool for helping shape health policies that assist in finding the best solution to complex problems by exploring for a range of parameters what impact they have on outcomes. This in turn allows development of policies to optimize outcomes using those parameters influenced by policies (e.g., viral load, drug concentration levels, adherence levels, set point viral load, and cd4 count). Algorithms such as the grid search method, hill climbing, gradient descent, and

simulated annealing each serve a purpose depending on the complexity of the objective or cost function and the size of the search space. Optimization routines may also assist with determining how financial costs effect the overall goal of a health intervention and the tradeoffs between prioritizing one group of individuals over another. Using the grid method, many of the prioritization groups were clearly identified based on the adherence and set point viral load thresholds. However, the weighted costs in the model does not reflect the cost of determining the SPVL and adherence of everyone in the simulated population. Overall, the adapted grid method was computationally taxing on the hours needed to complete the treatment switching optimization routine.

The simulated annealing algorithm was able to capture the best parameter set for the desired outcome and demonstrate time efficiency. Thus, moving forward, using the adapted grid method may assist with having a well-defined health policy when the HIV conditions of a group of people are known including but not SPVL, viral load, adherence levels because the method is both systematic and exhaustive, finding all possible solutions for the search space. However, the simulated annealing method may either spend less time finding the optimal solution if clinical and laboratory results are known for individuals, and/or may provide new policies based on values that were not systematically explored in the parameter search space. In all, this summary is based on the treatment regimen recommendation of 2016 for treating individuals diagnosed with HIV and may differ when modeling and optimizing HIV outcomes like HIV-DR and second-line treatment. The key findings related to Aim 3 " Create an ART optimization routine for a stochastic, network-based model that will modify drug dosing and/or switch treatment regimens, given the patient's characteristics and the threshold level of a desired HIV outcome", related to the overall question of "What strategies and conditions may assist in predicting, reducing, and preventing HIV transmission and HIV drug resistance within-host and between-host (i.e., in a multi-level system)?" was,

(1) Although each optimization method was run on a simulation cluster server (sim cluster) "…, featuring simulation- specific software intended for computationally intensive work…" [216] and the compiler for the C programming script in the EvoNet modeling system included an -O3 variable that was intended for optimizing compile time and memory usage, the adapted grid

search method was computationally taxing with respect to time, taking 2.5 days to complete the algorithm. In general, the model conditions have a large combinatory parameter search space.

(2) The simulated annealing algorithm, *optim_sa()*, was more efficient on time.

(3) Individuals with a log10 SPVL of 4.6 and higher and an adherence threshold of 0.7 and lower should be prioritized for treatment switching.

(4) When pills are more expensive, individuals with a log10 SPVL lower than 4.5 should not be prioritized for treatment switching. If these individuals are prioritized when pills are expensive, a higher percentage of drug resistance in the population and higher pill count for the alternative second-line medication may occur. In contrast, the cost function values indicated that when pills are more expensive, individuals with an adherence level between 0.37-0.65 and a log10 SPVL of 5.5 or greater should be prioritized. This result was also reflected similar low-cost function scores when prioritizing individuals having a SPVL of at least between 5.5-6 and between 40-45% adherence, for both the adapted grid search method and simulated annealing algorithm.

(5) When pills are cheaper, individuals a log10 SPVL of 4.9 or higher should not be prioritized, unless they have an adherence of at least 45% drug adherence. In addition, cheaper pills may also generate more total second-line pills in the population, as well, because more individuals would be eligible to receive second-line treatment due to people with high log10 SPVLs and drug adherence levels above 45%, having priority in addition to those with lower SPVLs (i.e., those below 4.9).

Chapter 6: Summary of Research Findings and Implications

6.1 Introduction

As described in Chapter 2 and demonstrated in Aims 1-3 (Chapters 3-5) mathematical modeling can be used to aid in the prevention, control, and treatment of infectious disease. Mathematical models help enable the determination of optimal policies for use in conditions with limited resources (i.e., time, money, medication) by predicting the impact of outcomes based on varying parameters that reflect different and new policy approaches. The scenario of finding optimal policies for combating debilitating and sometimes fatal infectious diseases such as influenza, the coronavirus 2019, and still, the HIV epidemic, while having limited resources is a common issue of today. As mathematical models become more complex and sophisticated, more efforts need to be put into achieving model accuracy and optimizing parameters. In turn, these efforts will help researchers, public health professionals and clinicians make decision and policies that create a more personalized approach to healthcare and medicine.

For health professionals and researchers deciding how to treat each patient based on their clinical, laboratory, and genetic information is important in solving complex problems like eradicating an infectious disease such as HIV. The integration of clinical, laboratory and genetic information may lead to new treatment guidelines that do not compromise the efficacy of the medication prescribed for, in this case, the suppression of HIV. Although HIV transmission has significantly been reduced globally, there are many locations around the globe that are still affected by the epidemic and have also seen an uptick in cases among certain populations, especially in the areas of sub-Saharan Africa and in some cities in the United States such as Atlanta, Georgia[217,218]. The Joint United Nations Programme on HIV/AIDS (UNAIDS) has a goal of ending the HIV epidemic by 2030[72,73]. As described in Chapter 2, UNAIDS established goals to

achieve by the year 2020 which include: (1) to test and diagnose 90% of all people living with HIV/AIDS (PLWHA), (2) provide those diagnosed with HIV antiretroviral treatment, and (3) to ensure those receiving antiretroviral treatment achieve viral suppression[73]. However, the UNAIDS 90-90-90 plan was not achieved by 2020[219]. If health interventions such as HIV epidemic modeling and health policies do not address issues such as HIV drug resistance and in particular, transmitted drug resistance, then eradicating HIV will become more challenging. Organization such as the World Health Organization (WHO) have identified early warning indicators (EWIs) such as treatment supply, follow-up care/retention, and other clinical factors. However, other contributing factors to HIVDR such as pharmacogenomics should be investigated more widely and adopted in mathematical models for the prediction of HIV within-host and between-host dynamics.

The field of pharmacogenomics has contributed to better tailoring treatment using patient information such as lifestyle, their environment and genetic makeup, leading to a more accurate account of drug effectiveness [150]. In Chapter 3, I define pharmacogenomics (PGx) as the study of how host genes affect drug response[138,139]. In other words, PGx details the impact of host genetic variants on drug pharmacokinetics, or the way the drug moves through different compartments in the body[138]. Consequently, either adverse drug events such as toxicity may occur and/or HIV drug resistant strains may accumulate and transmit HIV to other individuals in the population, prolonging the fight to end the HIV/AIDS epidemic. CYP2B6 variants impact metabolism of the antiretroviral drug efavirenz. Depending on the genetic variant, the drug is metabolized rapidly or slowly, resulting in supra- or sub-optimal concentrations in the body. Ngaimsisi et al. 2013 and Nyakurita et al. 2008 demonstrated the difference in plasma concentrations for individuals with the CYP2B6 variant that led to high EFV drug concentrations of >4000 ug/mL (i.e., above the therapeutic range) indicating a slower metabolism of the drug and ultimately accumulating to toxic levels[148,150]. These findings led to reductions in dosing by 35% for patients that were slow metabolizers, which is also the current recommended range of dosing by the WHO for individuals taking efavirenz. In contrast, Maseng et al. 2020 analyzed a sub-sample of HIV patients in Gaborone, Botswana administered an

EFV/NVP regimen and found that individuals with a CYP2B6 polymorphism and associated with being a fast metabolizer, were more like to have EFV/NVP drug resistance than the slow metabolizers[220].

In this thesis, I aimed to predict and reduce HIV transmission and the levels of HIVDR, including acquired and transmitted drug resistant strains, within-host and in the population, using a network-based, stochastic epidemic model called EvoNet. In aim 1 (Chapter 3), the goal was to predict the emergence and proportion of total drug resistant mutations, within-host, and other outcomes such as viral load, drug concentration, and the other viral mutant strains that makeup single, double, triple, quadruple, or quintuple HIV mutations, within-host. Additionally, in aim 1, we parameterize the model with pharmacogenomic, and pharmacokinetic information related to a first-line HIV drug regimen (i.e., tenofovir, lamivudine, and efavirenz) to predict the emergence of drug resistance, given the presence or absence of a genetic variant, within-host, variation. In Aim 2 (Chapter 4), the goal was to model the effect of viral genetics and prevalent drug resistance mutations on HIV-DR levels in two sub-Saharan African populations. Lastly, in Aim 3 (Chapter 5), the goal was to develop an optimization routine for treatment switching to find model parameter values that would optimize an HIV outcome such as percent resistance in the population, and the supply of antiretroviral medication in the population. Next, I will conclude by outlining my thesis research findings and discussing future implications.

6.2. Summary of Findings

The overall goal of this dissertation project is to explain and predict possible conditions that are likely to lead to HIVDR, within-host and in the population. In addition, this dissertation project aims to use that knowledge to implement methodologies that optimize resources such as antiretroviral medication to assist researchers, clinicians, and public health professionals in reducing HIVDR and HIV transmission in the population, given a patient's clinical lab test results and/or their genetic makeup. For each aim I utilized the HIV-related information and methodologies discussed in Chapters 1 and 2. The following summarizes the goal/objective of each aim and the corresponding research findings in the context of the overarching question of "What

strategies and conditions may assist in predicting, reducing, and preventing HIV transmission and HIV drug resistance within-host and between-host (i.e., in a multi-level system)?"

Aim 1 (Chapter 3) Research Findings

For aim 1, the main goal was to predict the emergence and proportion of total drug resistant mutations, within-host, and other outcomes such as viral load, drug concentration, and the emergence of specific viral mutants, given the presence of host genetic variants and varying drug decay levels.

Specifically, I sought to model the effect of host genetic variation on the concentration of HIV-DR mutations, within-host, by varying "patient-specific" pharmacokinetic properties of antiretroviral medication, and drug adherence.

To accomplish this, I used a multi-scale epidemic model to determine the effect of within-host viral evolution, given the presence of a host allele of a fast, intermediate, or slow ART metabolizer. The purpose of the aim was to also quantify the frequency of drug resistance mutations in an infected individual and examine how and when the drug resistant mutation emerges. The following patterns were found with the within-host model:

(1) Metabolizer type effects drug concentration in the body. As such, fast metabolizers were found to quickly clear the antiretroviral medication and slow metabolizers accumulated high levels of the antiretroviral.
(2) The time to drug resistance also varied by metabolizer type. For fast metabolizers, drug resistance accumulation within-hosts, happened more quickly than for the intermediate and slow metabolizers.
(3) The emergence of drug resistant mutations also demonstrated a hierarchical order. For instance, the pattern of drug resistant mutations appearing included the single>double>triple>quadruple mutations.
(4) Drug adherence drives the accumulation of drug resistant mutations at low adherence levels, particularly at 50% adherence.

Aim 2 (Chapter 4) Research Findings

In aim 2 (Chapter 4), the goal was to utilize the EvoNet modeling framework and model setup of aim 1 (Chapter 3) to vary the proportions of pharmacogenomic parameters (i.e., fast, intermediate, and slow metabolizers), in the population and modify the pharmacokinetic parameter values to predict drug resistance levels, and in the population, given varying proportions of metabolizers in the population. In addition, in aim 2 (Chapter 4), the goal is to investigate the distribution of the different types of drug resistance, transmitted drug resistance and acquired drug resistance, to the overall proportion of HIVDR in the population.

Specifically, I sought to model the effect of viral genetics and prevalent drug resistance mutations on HIV-DR levels in two sub-Saharan African populations.

To accomplish this, I collected pharmacogenomic and mutation frequency data from two studies conducted in sub-Saharan African countries: Zimbabwe and the Democratic Republic of the Congo (DRC), respectively. The data did not utilize all epidemiological findings related to HIV population-based outcomes in either country and only reflected values specific to the frequency of fast, intermediate, and slow metabolizers and the K103N mutation in each study's sample population. The model used for aim 2 also incorporated the antiretroviral first-line drug regimen: tenofovir, lamivudine, and efavirenz, which was the regimen recommended for clinical use in 2016 and part of 2017. As such, the model parameter values for this drug regimen that PGx and PK information was used to simulate the EvoNet HIV model outcomes for this dissertation. However, the EvoNet modeling system is parameterized to modify pharmacologic parameters for old and new medications, and as such the new model regimen that includes dolutegravir (DTG) instead of EFV, may be parameterized in the EvoNet model as well.

The outcome of each model simulation included quantifying the levels of drug resistance in the population and the contribution of acquired and transmitted resistance to the overall drug resistance. Four model conditions were used to compare and evaluate how the drug resistance, transmitted, and acquired drug resistance, and metabolizers changed over the duration of 10 years, using the frequency of PGx and K103N mutant

data for each study sample population conducted in Zimbabwe and the DRC. The following patterns were found with the between-host model:

(1) The model simulation that included only assigning the prevalent drug resistant mutation to a proportion of the population at the start of the simulation (i.e., model 3) predicted the highest level of drug resistance at the end of the 10-year duration. The second highest drug resistant mutation level occurred with the model having no a priori drug resistant mutation proportion assigned to the population or metabolizers in the population (i.e., model 1). The lowest drug resistance levels predicted occurred with the metabolizers in the population only (i.e., model 2). Finally, the model included both the prevalent drug resistant mutation in the population and the metabolizer proportions in the population (i.e., model 4) had drug resistance mutations levels similar to the model without the drug resistant mutation and metabolizer proportions (model 1).
(2) A higher proportion of transmitted drug resistance among infected and drug resistant individuals in the population was found in all the four model types. The levels of transmitted drug resistance were particularly all above 10% in the population and within a 59-60% range for transmitted drug resistance among those who had drug resistance in the population.
(3) Sub-therapeutic drug concentrations were more prevalent in the population for models 1 and 3.
(4) Supra-therapeutic drug concentrations were more prevalent in the population for model 2 and 4.

Aim 3 (Chapter 5) Research Findings

In aim 3 (Chapter 5), the goal was to find the best policy for switching individuals to second-line treatment, given their individual attributes or parameter values such as levels of set point viral load and drug adherence to achieve the lowest, or optimal HIVDR level for the simulated population. The optimization routine was developed using a computer programming algorithm that switched individuals based on whether they adhered to policy guidelines. For instance, if a person had a drug adherence level

below the policy guideline, then the person was switched to second-line treatment. Second-line treatment was generalized to act as an effective treatment that could suppress the virus better than the first-line treatment. Additionally, in aim 3 (chapter 5), I compared the benefits of using two optimization methods: the adapted grid-search method and simulated annealing.

Aim 3 Specific objective: Create an ART optimization routine for a stochastic, network-based model that will modify drug dosing and/or switch treatment regimens, given the patient's characteristics and the threshold level of a desired HIV outcome.

The third thesis aim investigated the use of optimization and machine learning technique to inform an optimization routine for switching treatment and reducing the levels of drug resistance in the population. Other HIV outcomes such as pill count and number of infectious people were also calculated. The methodology for aim 3 involved using a nested for-loop algorithm that mirrored the concept of the grid search method and a second algorithm that involved using the simulated annealing R package, optimization, and the function, optim_sa(). Both methods were used to determine the best parameter set values for reducing drug resistance and/or conditions that included a high to low cost of second-line treatment.

The following information was gained from creating and exploring the optimization routine:

(6) Although each optimization method was run on a simulation cluster server (sim cluster) "…, featuring simulation- specific software intended for computationally intensive work…" [216] and the compiler for the C programming script in the EvoNet modeling system included an -O3 variable that was intended for optimizing compile time and memory usage, the adapted grid search method was computationally taxing with respect to time, taking 2.5 days to complete the algorithm. In general, the model conditions have a large combinatory parameter search space.

(7) The simulated annealing algorithm, *optim_sa()*, was more efficient on time. but missed key details about the parameter space identified by the grid search

method. While optim_sa might be able to provide more of these details with parameter tweaking, the grid search method, while more computational demanding, uncovers the entire space without the need to consider the complex details of how optim_sa works.

(8) For the parameter values in our model, individuals with a log10 SPVL of 4.6 and higher and an adherence threshold of 0.7 and lower should be prioritized for treatment switching.

(9) When pills are more expensive, individuals with a log10 SPVL lower than 4.5 should not be prioritized for treatment switching. If these individuals are prioritized when pills are expensive, a higher percentage of drug resistance in the population and higher pill count for the alternative second-line medication may occur. In contrast, the cost function values indicated that when pills are more expensive, individuals with an adherence level between 0.37-0.65 and a log10 SPVL of 5.5 or greater should be prioritized. This result was also reflected similar low-cost function scores when prioritizing individuals having a SPVL of at least between 5.5-6 and between 40-45% adherence, for both the adapted grid search method and simulated annealing algorithm.

When pills are cheaper, individuals a log10 SPVL of 4.9 or higher should not be prioritized, unless they have an adherence of at least 45% drug adherence. In addition, cheaper pills may also generate more total second-line pills in the population, as well, because more individuals would be eligible to receive second-line treatment due to people with high log10 SPVLs and drug adherence levels above 45%, having priority in addition to those with lower SPVLs (i.e., those below 4.9).

6.3 Limitations

For aim 1 (Chapter 3), the limitations included variance in the values of pharmacokinetic properties from different resources for the antiretrovirals. For instance, the inhibition concentration 50 (IC50) is a pharmacokinetic property that, depending on the source of each pharmacokinetic parameter, had differing value ranges (see Chapter 3, Table 9). For instance, different IC50 ranges were observed for 3TC between for the FDA drug labels, the Rosenbloom et al. 2012 study[132], and the Parkin et

al. 2004 study. Additionally, drug resistance mutation data was obtained using the genotype-treatment information on the Stanford HIV Drug Resistance database (HIVdb) and other published works [164–166]. We resolved the discrepancy in the values used for IC50 by simulating the within-host model for each set of IC50 values and chose the values that were found to result in the HIV outcomes that were similar to empirical HIV within-host, clinical and published peer-reviewed studies. A sensitivity analysis was conducted to demonstrate how the model outcomes may vary due to the uncertainty of model parameter values and sources of information.

For aim 2 (Chapter 4), the limitations included the absence of fitting the model to the study populations and ensuring that the EvoNet model outcomes are approximate to the empirical data or expected data in the field. The model conditions only included frequency data from two studies conducted to geographically different countries, Zimbabwe, and the Democratic Republic of the Congo. The frequency data gathered were for the proportion of HIV-positive individuals who experienced virological failure and had the K103N mutation in the population and the proportion of fast, intermediate, and slow metabolizers in the population. However, the model could have included other parameters (i.e., related to social and behavior attributes, for instance) that may have resulted in a better fit for both countries. Social and behavioral parameters may have included contraceptive use or age distribution. These parameters, alone, may have modified the model results and reflected a more accurate model for both countries.

For aim 3 (Chapter 5), the limitations included the limited set of methods developed that could incorporate the EvoNetC modeling system. For example, the optim_sa() function was developed as an R routine for a fixed equation and parameter set of 2. However, EvoNetC has over 175 parameters and approximately 25 affect the simulation model runs for Aim 3. Thus, inputting more than two parameters in the optim_sa() function cause the algorithm to crash or have a longer run time. Additionally, more optimization and machine learning techniques like gradient descent and hill climbing could have been used. However, we noted that these optimization techniques may have been better suited for conditions that involve changing drug dose, for instance.

6.4 Other Findings

The models in this dissertation, included infected individuals in the simulated population that were eligible for and treated with the first-line drug regimen: tenofovir, lamivudine, and efavirenz.

For aim 1 (Chapter 3), I found that the following items from my research findings were consistent with other empirical data from the fields of HIV and pharmacogenomics and in relation to HIVDR:

(1) Metabolizer type effects drug concentration in the body. As such, fast metabolizers were found to quickly clear the antiretroviral medication and slow metabolizers accumulated high levels of the antiretroviral.

AND

(2) Drug adherence drives the accumulation of drug resistant mutations at low adherence levels, particularly at 50% adherence.

Other studies have shown that at the CYP2B6 516G>T variant, specifically TT genotypes (i.e., indication of a slow metabolizer) is closely associated with high EFV plasma concentrations and increased toxicity related to adverse neurological effects[148,221–226]. Nyakurita et al. 2008 examined the plasma levels of a sample population in Zimbabwe and found that 50% of the HIV-positive individuals had a plasma level greater than 4000ng/mL and were slow metabolizers, having the TT genotype or CYP2B6 516G>T(*6) variant[150]. 49% of individuals in the population were slow metabolizers. Gengiah et al. 2012 that found that individuals that were fast metabolizers of EFV-based regimens, were clearing the virus because of their GG genotype and the co-administration of the drug rifampicin, used to treat tuberculosis[227]. Gengiah et al. 2012 study found that slower metabolizers demonstrated a clearance of EFV almost half the rate of fast metabolizers[227].

(2) The time to drug resistance also varied by metabolizer type. For fast metabolizers, drug resistance accumulation within-hosts, happened more quickly than for the intermediate and slow metabolizers.

The finding that slow metabolizers accumulate drug resistance more slowly than fast metabolizers is consistent with empirical data. Maseng et al. 2021 conducted a study in Botswana with HIV-1 patients given an EFV-based regimen and found that drug resistance mutations were most prevalent in individuals that were fast versus slow metabolizers at 30.8 and 16.8% respectively[228].

(3) The emergence of drug resistant mutations also demonstrated a hierarchical order. For instance, the pattern of drug resistant mutations appearing included the single>double>triple>quadruple mutations.

A study by Feder et al. demonstrated that drug resistance mutations had a particular emergence order[229]. The study examined other published works that investigated the emergence of drug resistance mutations. For an EFV-based regimen, the patients developed EFV-based drug resistance (i.e., a single mutation), first. Then, the patients developed EFV-3TC drug resistance (i.e., a double mutation), followed by the e triple mutant[226]. Another study by Zou et al. 2020 stated single mutations in the presence of an NNRTI-based drug like EFV will eventually lead to multiple and high-level resistance[178]. The study also found that K103N/S was the most frequently observed mutation. Our research finding also demonstrated that K103N was a frequently observed mutation within-host. Feder et al. 2021 and Pennings et al. 2014 both have observed that HIV drug resistance mutations evolve in a stepwise fashion with some viral strains harboring 1-, or 2- drug resistant mutations and with a high propensity to evolve into triple drug resistance over time[168,229].

For aim 2 (Chapter 4), I found that the following items from my research findings were both consistent and inconsistent with other empirical data from the fields of HIV and pharmacogenomics and in relation to HIVDR:

(1) The model simulation that included assigning the prevalent drug resistant mutation to a proportion of the population at the start of the simulation (i.e., model 3) predicted the highest level of drug resistance at the end of the 10-year duration.

The model condition #3 from the thesis aim 3 (Chapter 5) was similar to the research findings of pre-treatment drug resistance found in a population. As described by the

WHO surveillance drug resistance mutations, the presence of common mutations like K103N in the model may lead to increased drug resistance levels when the existing drug resistance levels are over 10% in the population[33,34,170,230]. For model #2 that included pharmacogenomic data, studies such as Maseng et al. found that EFV/NVP drug resistance was most prevalent with in the population when accounting for fast metabolizers, which is also consistent with our findings[228]. However, to date the model condition describing both the K103N mutation and inclusion of the proportion of fast, intermediate, and slow metabolizers in the model (i.e., model condition #4) was not found to have an equivalent empirical observation in a population.

(2) A higher proportion of transmitted drug resistance among infected and drug resistant individuals in the population was found in all the four model types. The levels of transmitted drug resistance were particularly all above 10% in the population and within a 59-60% range for transmitted drug resistance among those who had drug resistance in the population.

The higher levels of transmitted drug resistance were not consistent with other research findings with a similar first-line drug regimen. Lee et al. 2014 found that study populations in Uganda, had low to moderate levels of TDR (7% to 8%) in the population. Additionally, a study Weng et al. found that levels of TDR did not increase significantly between the four years patients were tracked in the study, between 2007-2011 and 2012-2015. Specifically, TDR was reported at 10.6 and 7.6% in the population.

(3) Sub-therapeutic drug concentrations were more prevalent in the population for models 1 and 3.

AND

(4) Supra-therapeutic drug concentrations were more prevalent in the population for model 2 and 4.

The aim 2 (Chapter 4) research findings related to sub-therapeutic and supratherapeutic drug concentrations were consistent with empirical and clinical/lab data. Nwogu et al. 2021 analyzed EFV-based first line regimen and found that within their study population conducted in Nigeria, slow metabolizers, or individuals with the CYP2B6 516GT and CYP2B6 983CC homozygotes had three- and two-fold higher plasma concentrations of the drug efavirenz compared to intermediate metabolizers[231]. Thus,

individuals were observed to have a range of between 46.8 to 15592.3ng/ml, where less than 1000 ng/mL is subtherapeutic, 1000-4000 ng/mL is therapeutic, and above 4000 ng/mL is supratherapeutic[231]. Thus, for this thesis, aim 2 (Chapter 4) the models 2 and 4 both contain pharmacogenomic data and the presence of either slow and/or fast metabolizers and have drug concentrations with more people having supratherapeutic drug concentrations than subtherapeutic concentrations, as was found in the study by Nwogu et al. 2021[231].

Lastly, in aim 3 (Chapter 5), I found that the following items from my research findings that were consistent with other studies from machine learning literature:

(1) Although each optimization method was run on a simulation cluster server (sim cluster) "…, featuring simulation- specific software intended for computationally intensive work…" [216] and the compiler for the C programming script in the EvoNet modeling system included an -O3 variable that was intended for optimizing compile time and memory usage, the adapted grid search method was computationally taxing with respect to time, taking 2.5 days to complete the algorithm. In general, the model conditions have a large combinatory parameter search space.

The grid search optimization method is known as an exhaustive method that has a high computational time burden, especially in large search spaces.

(2) The simulated annealing algorithm, *optim_sa()*, was more efficient on time.

The specific results of the aim 3 optimization routine (found in Chapter 5) do not reflect research found in other settings. Although there are platforms such as Euresist that function as a HIVDR database and prediction tool, the tools use machine learning techniques like logistic regression models to classify treatment successes and failures [68–70]. Then, uses a Bayesian network to calculate the probabilities of treatment success from the clinical information which does not include genotype data [68–70]. Whereas the approach of aim 3 of this thesis research has the goal of switching treatment for individuals that may fail treatment based on clinical and genetic attributes like viral load, set point viral load, drug adherence, drug concentration, and CD4 count, for

example, using an adapted grid search method and simulated annealing methodology. Our model also aimed to find policies that reduce HIVDR through determining which groups to prioritize, based on the threshold values for select parameters in the model. Therefore, although tools like Euresist include clinical results needed to help with individual HIV treatments and patient care, Euresist does not capture the biological, genetic, epidemiological, treatment, and social network data needed understand and analyze the evolution of HIV progression and link within-host data to population-based outcomes. As such, our optimization routine differs from the methodology used to determine treatment switching.

6.5 Future Research

Multi-scale modeling for infectious disease will continue to enhance and provide insight into the fields of biomedical health, public health, and medicine. However, I recommend that future modeling efforts incorporate information from fields like pharmacogenomics that may enhance patient clinical care and work towards a personalized medicine approach. In this dissertation, I outlined how adding parameters from the fields of pharmacogenomics and pharmacokinetics could improve upon the predictions of HIVDR, HIV transmission, and drug concentrations within-host and between-host. In addition, other techniques were introduced in this thesis such as an adapted grid search method and simulated annealing for the purpose of finding the optimal parameter space for lowering drug resistance in the population and determining which individuals would be switched to second-line treatment.

In this thesis, I have demonstrated how enhancing an existing model structure may be beneficial to understanding and finding optimal results for a given medical and/or public health issue such as HIVDR. Thus, future work may include: (1) fitting the model to specific populations and countries of interest (2) a detailed parameterization of the monetary cost of first line and second-line antiretroviral treatment, (3) an assessment of the guidelines for switching treatment from public health officials that implement treatment policies, and (4) incorporate a new drug regimen into the model. The following briefly discusses the future direction of this thesis work with a more detail:

(1) Fitting the model to specific populations and countries of interest

Fitting the model to specific populations and countries of interest may include adding epidemiologic data such as prevalence, incidence, percent eligible for ART, percent treated, and percent resistant in the population, to the model. In addition, model fitting for these two countries may involve collecting data from multiple sources such as the World Health Organization, national and local surveys, and additional information collected from published literature within a specified range of time.

(2) A detailed parameterization of the monetary cost of first line and second-line antiretroviral treatment

Econometric data about the cost of HIV medication has varied across low-, middle, and high-income countries[51,52]. With the wealthiest, or high-income countries, charged higher costs for the same HIV medications than lower income countries[51,52]. Investigating these wide ranges in pill cost for certain countries may be beneficial for predicting expected lifetime costs, quality-adjusted life-month, quality-adjusted life years and in the model for aim 3 of this thesis (Chapter 5), a better estimate for the penalty or "weight" on the total number of pills. These weights may also represent the differentiation between patients that have health insurance to cover costs and those that do not. A more detailed look into econometric data of HIV medications may lead to a more accurate optimization routine for switching treatment as described in aim 3 (Chapter 5).

(3) An assessment of the guidelines for switching treatment from public health officials that implement treatment policies

Currently, the guidelines for switching treatment have been modified for the older efavirenz-based regimen from a 600-mg to 400-mg dosage. In our model, we did not modify the dosage to 400-mg if a person was a slow metabolizer or had drug concentrations above the therapeutic level (i.e., between 1000-4000 ng/mL)[232–234]. Additionally, WHO recommendations for first- and second-line therapy has changed since we started modeling the TDF-3TC-EFV regimen and the current recommendation strongly recommends using this regimen under certain circumstances such as if a country is below a certain pre-treatment drug resistance threshold (i.e., below 10%) or for certain demographic populations[52,230].

(4) Incorporating new drug regimens into the model.

As described previously future work item (3), the WHO has updated recommendations for first- and second line treatment to include dolutegravir (DTG) primarily[230]. Although effective, EFV-based medication has been associated with several adverse drug events such as increased levels of drug toxicity and adverse neurological effects, that ultimately led to a recommended reduction in drug dose to 400 mg. However, individuals taking EFV, even at a lower dose were more likely to accumulate drug resistance mutations since EFV had a lower barrier to resistance, whereas the newer drug DTG has a high barrier to resistance and as such may yield lower frequencies of drug resistance mutations, within-host[230,233]. Thus, fitting the model to a new drug regimen that includes DTG may provide a more accurate and updated account of the drug resistance levels in present day. Additionally, there are approximately 30 HIV-based drugs approved by the Food and Drug Administration including the newest HIV drugs as the injectables, cabotegravir and rilpivirine[235,236]. Given this information, the EvoNet model was developed and structured to include different drug regimens through the capability of the modeling system to allow users to modify the values of pharmacokinetic parameters for the selected drug regimen. Thus, incorporating new and updated drug regimens in the EvoNet modeling system is a strong recommendation for future works.

In all, the research conducted in this thesis should continue to explore and develop methodologies such as fitting the model to econometric or new drug classes and the development of learning-based algorithms. These efforts may assist researchers, clinicians, and other health professionals with health decision-making and health policy modeling for the purpose of preventing, reducing the transmission of, and eradicating infectious diseases such as malaria, COVID-19, and as studied in this thesis, HIV/AIDS.

References

1. Dym, C. L. Principles of Mathematical Modeling. *Principles of Mathematical Modeling* 1–303 (2004) doi:10.1016/B978-0-12-226551-8.X5000-5.

2. McMillon, D., Simon, C. P. & Morenoff, J. Modeling the underlying dynamics of the spread of crime. *PLoS ONE* **9**, (2014).

3. Vanagas, G., Krilavičius, T. & Man, K. L. Mathematical Modeling and Models for Optimal Decision-Making in Health Care. *Computational and Mathematical Methods in Medicine* **2019**, (2019).

4. Zipfel, P. H. Modeling and simulation of aerospace vehicle dynamics. 551 (2000).

5. Ibeanusi, V. M., Jackson, E., Coffen, J. & Jeilanisi, Y. Assessing bioremediation of acid mine drainage in coal mining sites using a predictive neural network-based decision support system (NNDSS). *Journal of Bioremediation and Biodegradation* **3**, (2012).

6. Krämer, A., Kretzschmar, M. & Krickeberg, K. Modern infectious disease epidemiology : concepts, methods, mathematical models, and public health. (2010).

7. Kretzschmar, M. & Wallinga, J. Mathematical Models in Infectious Disease Epidemiology. 209–221 (2009) doi:10.1007/978-0-387-93835-6_12.

8. Jenner, A. L., Aogo, R. A., Davis, C. L., Smith, A. M. & Craig, M. Leveraging Computational Modeling to Understand Infectious Diseases. *Curr Pathobiol Rep* **8**, 149–161 (2020).

9. What Is Mathematical Modeling?

10. Cain, J. W. Mathematical Models in the Sciences. doi:10.1007/978-1-4614-6436-5_561-1.

11. Modern Infectious Disease Epidemiology | SpringerLink. https://link-springer-com.offcampus.lib.washington.edu/book/10.1007%2F978-0-387-93835-6.

12. Jenner, A. L., Aogo, R. A., Davis, C. L., Smith, A. M. & Craig, M. Leveraging Computational Modeling to Understand Infectious Diseases. *Curr Pathobiol Rep* **8**, 149–161 (2020).

13. Keeling, M. J. & Eames, K. T. D. Networks and epidemic models. *Journal of the Royal Society Interface* **2**, 295 (2005).

14. Mei, S., Zarrabi, N., Lees, M. & Sloot, P. M. A. Complex agent networks: An emerging approach for modeling complex systems. *Applied Soft Computing Journal* **37**, 311–321 (2015).

15. Willem, L., Verelst, F., Bilcke, J., Hens, N. & Beutels, P. Lessons from a decade of individual-based models for infectious disease transmission: A systematic review (2006-2015). *BMC Infectious Diseases* **17**, 1–16 (2017).

16. Nelson, M. L., Dinardo, A., Hochberg, J. & Armelagos, G. J. Brief communication: Mass spectroscopic characterization of tetracycline in the skeletal remains of an ancient population from Sudanese Nubia 350–550 CE. *American Journal of Physical Anthropology* **143**, 151–154 (2010).

17. Boyer, J. M. Journal of Health and Human Experience. *Articles* **62**,.

18. Maggiano, C. Ancient antibiotics : tetracycline in human and animal bone from the Dakhleh Oasis, Egypt. *HIM 1990-2015* (2002).

19. Ancient Nubians Made Antibiotic Beer | WIRED. https://www.wired.com/2010/09/antibiotic-beer/.

20. Alexander Fleming Discovery and Development of Penicillin - Landmark - American Chemical Society. https://www.acs.org/content/acs/en/education/whatischemistry/landmarks/flemingpenicillin.html.

21. Tan, S. Y. & Tatsumura, Y. Alexander Fleming (1881–1955): Discoverer of penicillin. *Singapore Medical Journal* **56**, 366 (2015).

22. Science Diction: The Origin Of 'Antibiotic' : NPR. https://www.npr.org/2011/02/11/133686020/Science-Diction-The-Origin-Of-Antibiotic.

23. Wise, R. *et al.* Antimicrobial resistance. *British Medical Journal* vol. 317 609–610 (1998).

24. Schwartz, B. Preventing the Emergence of Antimicrobial Resistance. *JAMA* **278**, 944 (1997).

25. Brooks, A. UK strategy to cut antibiotic use outlined. *BMJ* **317**, 699 (1998).

26. Marston, H. D., Dixon, D. M., Knisely, J. M., Palmore, T. N. & Fauci, A. S. Antimicrobial resistance. *JAMA - Journal of the American Medical Association* **316**, 1193–1204 (2016).

27. Antimicrobial resistance. https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance.

28. Global Statistics | HIV.gov. https://www.hiv.gov/hiv-basics/overview/data-and-trends/global-statistics.

29. Chapman Lambert, C., Marrazzo, J., Amico, K. R., Mugavero, M. J. & Elopre, L. PrEParing Women to Prevent HIV: An Integrated Theoretical Framework to PrEP Black Women in the United States. *Journal of the Association of Nurses in AIDS Care* **29**, 835–848 (2018).

30. Editors, G., Cáceres, C. F., Mayer, K. H., Baggaley, R. & Reilly, K. R. O. PrEP State-of-the-Art and Research Agenda. *J Int AIDS Soc* **18**, (2015).

31. Celum, C., Hallett, T. B. & Baeten, J. M. HIV-1 prevention with ART and PrEP: Mathematical modeling insights into resistance, effectiveness, and public health impact. *Journal of Infectious Diseases* vol. 208 189–191 (2013).

32. Lima, V. D. *et al.* Can the combination of TasP and PrEP eliminate HIV among MSM in British Columbia, Canada? *Epidemics* **35**, 100461 (2021).

33. *WHO HIV Drug Resistance Report 2017*. http://www.who.int/hiv/pub/drugresistance/hivdr-report-2017/en/ (2017).

34. Naik, S. & Das, B. R. HIV : Current research new WHO guidelines: Implications on therapeutics and monitoring of HIV infections. *HIV: Current Research* **1**, 1–4 (2016).

35. World Health Organization *et al.* National Comprehensive Guidelines for HIV Testing and Counselling. *World Health Organization* 14–15 (2016).

36. FDA approved changes to the DOVATO (dolutegravir/lamivudine) product labeling | FDA. https://www.fda.gov/drugs/human-immunodeficiency-virus-hiv/fda-approved-changes-dovato-dolutegravirlamivudine-product-labeling.

37. FDA approves first two-drug regimen for certain patients with HIV | FDA. https://www.fda.gov/news-events/press-announcements/fda-approves-first-two-drug-regimen-certain-patients-hiv.

38. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection (2016). https://www.who.int/publications/i/item/9789241549684.

39. World Health Organization. Consolidated guidelines on HIV prevention, testing, treatment, service delivery and monitoring : recommendations for a public health approach. 548.

40. Sabin, C. A. Do people with HIV infection have a normal life expectancy in the era of combination antiretroviral therapy? *BMC Medicine* vol. 11 251 (2013).

41. Wandeler, G., Johnson, L. F. & Egger, M. Trends in life expectancy of HIV-positive adults on antiretroviral therapy across the globe: Comparisons with general population. *Current Opinion in HIV and AIDS* vol. 11 492–500 (2016).

42. Nsanzimana, S. *et al.* Life expectancy among HIV-positive patients in Rwanda: A retrospective observational cohort study. *The Lancet Global Health* **3**, e169–e177 (2015).

43. Trickey, A. *et al.* Survival of HIV-positive patients starting antiretroviral therapy between 1996 and 2013: a collaborative analysis of cohort studies. *The Lancet HIV* **4**, e349–e356 (2017).

44. Wandeler, G., Johnson, L. F. & Egger, M. Trends in life expectancy of HIV-positive adults on ART across the globe: comparisons with general population. *Curr Opin HIV AIDS* **11**, 492 (2016).

45. Mansky, L. M. & Temin, H. M. *Lower In Vivo Mutation Rate of Human Immunodeficiency Virus Type 1 than That Predicted from the Fidelity of Purified Reverse Transcriptase †*. *JOURNAL OF VIROLOGY* vol. 69 http://jvi.asm.org/ (1995).

46. Suñé, C., Brennan, L., Stover, D. R. & Klimkait, T. Effect of polymorphisms on the replicative capacity of protease inhibitor-resistant HIV-1 variants under drug pressure. *Clinical Microbiology and Infection* **10**, 119–126 (2004).

47. Quinones-Mateu, M., Ball, S., Marozsan, A., Torre, V. & Albright, J. A dual infection/competition assay shows a correlation between ex vivo human immunodeficiency virus type 1 fitness and disease progression. *J Virol* **74**, (2000).

48. Kouyos, R. D. *et al.* Assessing Predicted HIV-1 Replicative Capacity in a Clinical Setting. *PLOS Pathogens* **7**, e1002321 (2011).

49. Rhee, S.-Y. *et al.* Open Forum Infectious Diseases Virological Failure and Acquired Genotypic Resistance Associated With Contemporary Antiretroviral Treatment Regimens. doi:10.1093/ofid/ofaa316.

50. Bonsall, D. *et al.* A comprehensive genomics solution for HIV surveillance and clinical monitoring in low-income settings. *Journal of Clinical Microbiology* **58**, 382–402 (2020).

51. Opinion | H.I.V. Drugs Cost $75 in Africa, $39,000 in the U.S. Does It Matter? - The New York Times. https://www.nytimes.com/2018/09/18/opinion/pricing-hiv-drugs-america.html.

52. Belay, Y. B., Ali, E. E., Chung, K. Y., Gebretekle, G. B. & Sander, B. Cost-Utility Analysis of Dolutegravir- Versus Efavirenz-Based Regimens as a First-Line Treatment in Adult HIV/AIDS Patients in Ethiopia. *PharmacoEconomics Open* **5**, 655 (2021).

53. Chang, S. *et al.* Comparison of susceptibility of HIV-1 variants to antiretroviral drugs by genotypic and recombinant virus phenotypic analyses. *International Journal of Infectious Diseases* **37**, 86–92 (2015).

54. Ghosh, A. K., Osswald, H. L. & Prato, G. Recent Progress in the Development of HIV-1 Protease Inhibitors for the Treatment of HIV/AIDS. *Journal of Medicinal Chemistry* vol. 59 5172–5208 (2016).

55. Ridky, T. & Leis, J. *Development of Drug Resistance to HIV-1 Protease Inhibitorsâˆ—*. (1995) doi:10.1074/jbc.270.50.29621.

56. Thompson, J. A. *et al.* Evolution of Protease Inhibitor Resistance in Human Immunodeficiency Virus Type 1 Infected Patients Failing Protease Inhibitor Monotherapy as Second-line Therapy in Low-income Countries: An Observational Analysis Within the EARNEST Randomized Trial. *Clinical Infectious Diseases* **68**, 1184–1192 (2019).

57. Rhee, S. Y. *et al.* HIV-1 protease mutations and protease inhibitor cross-resistance. *Antimicrobial Agents and Chemotherapy* **54**, 4253–4261 (2010).

58. Shafer, R. W. *et al.* HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance. *Aids* (2007) doi:10.1097/QAD.0b013e328011e691.

59. Götte, M., Arion, D., Parniak, M. A. & Wainberg, M. A. The M184V Mutation in the Reverse Transcriptase of Human Immunodeficiency Virus Type 1 Impairs Rescue of Chain-Terminated DNA Synthesis. *Journal of Virology* **74**, 3579–3585 (2000).

60. Marcelin, A.-G. *Resistance to nucleoside reverse transcriptase inhibitors*. *Antiretroviral Resistance in Clinical Practice* (Mediscript, 2006).

61. Meyer, P. R., Matsuura, S. E., Mohsin Mian, A., So, A. G. & Scott, W. A. A mechanism of AZT resistance: An increase in nucleotide-dependent primer unblocking by mutant HIV-1 reverse transcriptase. *Molecular Cell* **4**, 35–43 (1999).

62. Weikl, T. R. & Hemmateenejad, B. Accessory mutations balance the marginal stability of the HIV-1 protease in drug resistance. *Proteins: Structure, Function, and Bioinformatics* **88**, 476–484 (2020).

63. Kamal Guha, S., Soumendra, N. & Haldar, K. *DRUG RESISTANCE IN HIV*.

64. Anstett, K., Brenner, B., Mesplede, T. & Wainberg, M. A. HIV drug resistance against strand transfer integrase inhibitors. *Retrovirology* vol. 14 36 (2017).

65. el Bouzidi, K. *et al.* High prevalence of integrase mutation L74I in West African HIV-1 subtypes prior to integrase inhibitor treatment. *J Antimicrob Chemother* **75**, 1575–1579 (2020).

66. Seatla, K. K. *et al.* Limited hiv-1 subtype c nef 3'ppt variation in combination antiretroviral therapy naïve and experienced people living with hiv in botswana. *Pathogens* **10**, (2021).

67. Kuiken, C., Korber, B. & Shafer, R. W. HIV Sequence Databases. *AIDS Rev* **5**, 52 (2003).

68. EuResist Network | Research in HIV | HIV resistance database. https://www.euresist.org/.

69. Zazzi, M. *et al.* Predicting response to antiretroviral treatment by machine learning: The euresist project. *Intervirology* **55**, 123–127 (2012).

70. Zazzi, M. *et al.* Prediction of response to antiretroviral therapy by human experts and by the EuResist data-driven expert system (the EVE study).[Erratum appears in HIV Med. 2012 Apr;13(4):253]. *HIV Medicine* (2011) doi:https://dx.doi.org/10.1111/j.1468-1293.2010.00871.x.

71. Tackling entrenched inequalities to end epidemics.

72. UNAIDS. Commitments To End Aids By 2030 Fast-Track Commitments To End Aids By 2030. *UNAIDS* 8 (2014).

73. Unaids. 90-90-90: An ambitious treatment target to help end the AIDS epidemic. (2020).

74. Global HIV Programme. https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/hiv/treatment/hiv-drug-resistance/laboratory-network.

75. World Health Organization (WHO). WHO HIVResNet HIV drug resistance laboratory operational framework, second edition. 180–199 (2020).

76. GLOBAL REPORT ON EARLY WARNING INDICATORS OF HIV DRUG RESISTANCE HIV DRUG RESISTANCE TECHNICAL REPORT. (2016).

77. Fokam, J. *et al.* Declining trends in early warning indicators for HIV drug resistance in Cameroon from 2008-2010: Lessons and challenges for low-resource settings. *BMC Public Health* **13**, 1–10 (2013).

78. Getaneh, Y. *et al.* HIV drug resistance early warning indicators in Ethiopia: Variability at regional and health facility levels and trend over time. *International Journal of Infectious Diseases* **95**, 90–97 (2020).

79. Rhee, S.-Y., Tzou, P. L. & Shafer, R. W. Temporal Trends in HIV-1 Mutations Used for the Surveillance of Transmitted Drug Resistance. *Viruses* **13**, 879 (2021).

80. Crowell, T. A. *et al.* Pretreatment and Acquired Antiretroviral Drug Resistance Among Persons Living With HIV in Four African Countries. *Clinical Infectious Diseases* (2020) doi:10.1093/cid/ciaa1161.

81. Poppe, L. K. *et al.* HIV drug resistance in infants increases with changing prevention of mother-to-child transmission regimens. *AIDS* **31**, 1885–1889 (2017).

82. Yang, W. L. *et al.* Assessing the paradox between transmitted and acquired HIV type 1 drug resistance mutations in the Swiss HIV Cohort Study from 1998 to 2012. *Journal of Infectious Diseases* **212**, 28–38 (2015).

83. Aldous, A. M., Castel, A. D. & Parenti, D. M. Prevalence and trends in transmitted and acquired antiretroviral drug resistance, Washington, DC, 1999-2014. *BMC Research Notes* vol. 10 474 (2017).

84. Sethi, A. K., Celentano, D. D., Gange, S. J., Moore, R. D. & Gallant, J. E. Association between Adherence to Antiretroviral Therapy and Human Immunodeficiency Virus Drug Resistance. *Clinical Infectious Diseases* **37**, 1112–1118 (2003).

85. Abaasa, A. M. *et al.* Good adherence to HAART and improved survival in a community HIV/AIDS treatment and care programme: The experience of the AIDS Support Organization (TASO), Kampala, Uganda. *BMC Health Services Research* **8**, 1–10 (2008).

86. Schrager, L., Saah, A. & Jacobson, L. Acquired immune deficiency syndrome occurring within 5 years of infection with human immunodeficiency virus type–1: The multicenter aids cohort study. *Journal of Acquired Immune Deficiency Syndromes* **5**, 490–496 (1992).

87. Nachega, J. B. *et al.* Adherence to nonnucleoside reverse transcriptase inhibitor-based HIV therapy and virologic outcomes. *Annals of Internal Medicine* **146**, 564–573 (2007).

88. P, M. & BC, M. HIV drug resistance mutations following poor adherence in HIV-infected patient: a case report. *Clin Case Rep* **3**, 353–356 (2015).

89. Nachega, J. B. *et al.* HIV treatment adherence, drug resistance, virologic failure: evolving concepts. *Infect Disord Drug Targets* **11**, 167–74 (2011).

90. Oyugi, J. H. *et al.* Treatment interruptions predict resistance in HIV-positive individuals purchasing fixed-dose combination antiretroviral therapy in Kampala, Uganda. *AIDS* **21**, 965–971 (2007).

91. A, M., ZP, Y. & AR, N. N. HIV-infected patients' adherence to highly active antiretroviral therapy: a phenomenological study. *Nurs Health Sci* **12**, 464–469 (2010).

92. Katz, I. T. *et al.* Impact of HIV-related stigma on treatment adherence: systematic review and meta-synthesis. *J Int AIDS Soc* **16**, (2013).

93. Mellins, C. A. *et al.* Adherence to antiretroviral medications and medical care in HIV-infected adults diagnosed with mental and substance abuse disorders. *AIDS Care* **21**, 168 (2009).

94. Bamberger, J. D. *et al.* Helping the urban poor stay with antiretroviral HIV drug therapy. *American Journal of Public Health* **90**, 699 (2000).

95. Adewuya, A. O. *et al.* The Effect of Psychological Distress on Medication Adherence in Persons With HIV Infection in Nigeria. *Psychosomatics* **51**, 68–73 (2010).

96. S, M. *et al.* An audit of viral load in one clinical population to describe features of viraemic patients on antiretroviral therapy. *HIV Med* **9**, 208–213 (2008).

97. A “Mini-Pillbox” Might Someday Deliver A Week’s Worth Of Anti-HIV Drugs : Goats and Soda : NPR. https://www.npr.org/sections/goatsandsoda/2018/01/09/576587339/swallowing-a-mini-pillbox-could-change-the-way-hiv-drugs-are-delivered.

98. Bionghi, N. *et al.* Pilot evaluation of a second-generation electronic pill box for adherence to Bedaquiline and antiretroviral therapy in drug-resistant TB/HIV co-infected patients in KwaZulu-Natal, South Africa. *BMC Infectious Diseases* **18**, 1–9 (2018).

99. Bangsberg, D. R. *et al.* High levels of adherence do not prevent accumulation of HIV drug resistance mutations. *AIDS* **17**, 1925–1932 (2003).

100. G, V. *et al.* Drug resistance outcomes of long-term ART with tenofovir disoproxil fumarate in the absence of virological monitoring. *J Antimicrob Chemother* **73**, 3148–3157 (2018).

101. Gardner, E. M., Burman, W. J., Steiner, J. F., Anderson, P. L. & Bangsberg, D. R. Antiretroviral medication adherence and the development of class-specific antiretroviral resistance. *AIDS* vol. 23 1035–1046 (2009).

102. Bangsberg, D. R. *et al.* High levels of adherence do not prevent accumulation of HIV drug resistance mutations. *AIDS* **17**, 1925–1932 (2003).

103. Bukenya, D., Mayanja, B. N., Nakamanya, S., Muhumuza, R. & Seeley, J. What causes non-adherence among some individuals on long term antiretroviral therapy? Experiences of individuals with poor viral suppression in Uganda. *AIDS Research and Therapy* **16**, 1–9 (2019).

104. Paredes, R. *et al.* Pre-existing minority drug-resistant HIV-1 variants, adherence, and risk of antiretroviral treatment failure. *Journal of Infectious Diseases* **201**, 662–671 (2010).

105. Cassels, S., Clark, S. J. & Morris, M. Mathematical models for HIV transmission dynamics: Tools for social and behavioral science research. in *Journal of Acquired Immune Deficiency Syndromes* vol. 47 S34 (NIH Public Access, 2008).

106. Berge, T., Lubuma, M.-S., Moremedi, G. M., Morris, N. & Kondera-Shava, R. Journal of Biological Dynamics A simple mathematical model for Ebola in Africa) A simple mathematical model for Ebola A simple mathematical model for Ebola in Africa. *Journal of Biological Dynamics* **11**, 42–74 (2017).

107. Nosyk, B. *et al.* Why Maximizing Quality-Adjusted Life Years, rather than Reducing HIV Incidence, Must Remain Our Objective in Addressing the HIV/AIDS Epidemic. *J Int Assoc Provid AIDS Care* **18**, (2019).

108. Nosyk, B. *et al.* Modelling the cost-effectiveness of population-level HAART expansion in British Columbia. *Lancet HIV* **2**, e393 (2015).

109. Kiweewa, F. *et al.* HIV virologic failure and its predictors among HIV-infected adults on antiretroviral therapy in the African Cohort Study. *PLoS ONE* **14**, (2019).

110. Perelson, A. S. & Ribeiro, R. M. Modeling the within-host dynamics of HIV infection. *BMC Biology* vol. 11 (2013).

111. Hill, A. L. Mathematical Models of HIV Latency. in *Current Topics in Microbiology and Immunology* vol. 417 131–156 (Springer Verlag, 2018).

112. Chun, T. W. *et al.* Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* **387**, 183–188 (1997).

113. Vynnycky, Emilia. An introduction to infectious disease modelling / Emilia Vynnycky and Richard G. White ; with an introduction by Paul E.M. Fine. 358–364 (2010).

114. Hoare, A. *et al.* Hidden Drug Resistant HIV to Emerge in the Era of Universal Treatment Access in Southeast Asia. *PLOS ONE* **5**, e10981 (2010).

115. Ming, R. X., Liu, J. M., William, W. K. & Wan, X. Stochastic modelling of infectious diseases for heterogeneous populations. *Infectious Diseases of Poverty* **5**, 107 (2016).

116. Tracy, M., Cerdá, M. & Keyes, K. M. Agent-Based Modeling in Public Health: Current Applications and Future Directions. *Annu. Rev. Public Health* **39**, 77–94 (2018).

117. Bauer, A. L., Beauchemin, C. A. A. & Perelson, A. S. Agent-based modeling of host-pathogen systems: The successes and challenges. *Information Sciences* **179**, 1379–1389 (2009).

118. Benenson, I. & Hatna, E. The Third State of the Schelling Model of Residential Dynamics. (2009).

119. DeAngelis, D. L. & Diaz, S. G. Decision-Making in Agent-Based Modeling: A Current Review and Future Prospectus. *Frontiers in Ecology and Evolution* **6**, 237 (2019).

120. Zarrabi, N., Mancini, E., Tay, J. C., Shahand, S. & Sloot, P. M. A. Modeling HIV-1 intracellular replication: Two simulation approaches. in *Procedia Computer Science* vol. 1 555–564 (Elsevier B.V., 2010).

121. Castiglione, F., Pappalardo, F., Bernaschi, M. & Motta, S. Optimization of HAART with genetic algorithms and agent-based models of HIV infection. **23**, 3350–3355 (2007).

122. Adams, J. W. *et al.* Potential drivers of HIV acquisition in African-American women related to mass incarceration: An agent-based modelling study. *BMC Public Health* **18**, 1–11 (2018).

123. Brookmeyer, R. *et al.* Combination HIV prevention among MSM in South Africa: Results from agent-based modeling. *PLoS ONE* **9**, (2014).

124. Bershteyn, A. & Eckhoff, P. A. A model of HIV drug resistance driven by heterogeneities in host immunity and adherence patterns. *BMC Systems Biology* (2013) doi:10.1186/1752-0509-7-11.

125. Prejean, J. *et al.* Estimated HIV incidence in the United States, 2006-2009. *PLoS ONE* **6**, (2011).

126. Garira, W. & Mathebula, D. Development and application of multiscale models of acute viral infections in intervention research. *Mathematical Methods in the Applied Sciences* **43**, 3280–3306 (2020).

127. M, M. & XZ, L. Linking immunological and epidemiological dynamics of HIV: the case of super-infection. *J Biol Dyn* **7**, 161–182 (2013).

128. X, S. *et al.* Early HAART Initiation May Not Reduce Actual Reproduction Number and Prevalence of MSM Infection: Perspectives from Coupled within- and between-Host Modelling Studies of Chinese MSM Populations. *PLoS One* **11**, (2016).

129. DF, C. & G, G.-R. Variable effect of co-infection on the HIV infectivity: within-host dynamics and epidemiological significance. *Theor Biol Med Model* **9**, (2012).

130. Ma, Q., Brazeau, D., Forrest, A. & Morse, G. D. Advances in pharmacogenomics of antiretrovirals: an update. *Pharmacogenomics* **8**, 1169–1178 (2007).

131. Baeten, J. M. & Grant, R. Use of antiretrovirals for HIV prevention: What do we know and what don't we know? *Current HIV/AIDS Reports* **10**, 142–151 (2013).

132. Rosenbloom, D. I. S., Hill, A. L., Rabi, S. A., Siliciano, R. F. & Nowak, M. A. Antiretroviral dynamics determines HIV evolution and predicts therapy outcome. *Nature Medicine* **18**, 1378–1385 (2012).

133. von Hentig, N. [Measurement of the plasma concentration of antiretroviral drugs in HIV therapy]. *Dtsch Med Wochenschr* **133**, 191–195 (2008).

134. Beck, I. A. *et al.* Pre-treatment HIV-drug resistance associated with virologic outcome of first-line NNRTI-antiretroviral therapy: A cohort study in Kenya. *EClinicalMedicine* **18**, (2020).

135. Bousmah, M. al Q. *et al.* Cost-Utility Analysis of a Dolutegravir-Based Versus Low-Dose Efavirenz-Based Regimen for the Initial Treatment of HIV-Infected Patients in Cameroon (NAMSAL ANRS 12313 Trial). *Pharmacoeconomics* **39**, 331–343 (2021).

136. Pillay, P., Ford, N., Shubber, Z. & Ferrand, R. A. Outcomes for Efavirenz versus Nevirapine-Containing Regimens for Treatment of HIV-1 Infection: A Systematic Review and Meta-Analysis. *PLOS ONE* **8**, e68995 (2013).

137. Kumar, D. Genomic medicine: a new frontier of medicine in the twenty first century. *Genomic Medicine* **1**, 3 (2007).

138. Altman, R. B., Flockhart, D. & Goldstein, D. B. Principles of pharmacogenetics and pharmacogenomics. *Principles of Pharmacogenetics and Pharmacogenomics* **92**, 1–269 (2012).

139. FAQ About Pharmacogenomics - National Human Genome Research Institute (NHGRI). *National Human Genome Institute* https://www.genome.gov/27530645/ (2016).

140. Natarajan, S. *et al.* Effect of CYP2C9 and VKORC1 genetic variations on warfarin dose requirements in Indian patients. *Pharmacological Reports* **65**, 1375–1382 (2013).

141. Dean, L. Warfarin Therapy and VKORC1 and CYP Genotype. *Medical Genetics Summaries* (2018).

142. Russell, L. E. *et al.* Pharmacogenomics in the era of next generation sequencing – from byte to bedside. *https://doi.org/10.1080/03602532.2021.1909613* **53**, 253–278 (2021).

143. Haas, D. W. Will pharmacogenomic discoveries improve HIV therapeutics? *Top HIV Med* **13**, 90–95 (2005).

144. L, G., M, W.-C. & TE, K. PharmGKB, an Integrated Resource of Pharmacogenomic Knowledge. *Curr Protoc* **1**, (2021).

145. Asensi, V., Collazos, J. & Valle-Garay, E. Can antiretroviral therapy be tailored to each human immunodeficiency virus-infected individual? Role of pharmacogenomics. *World J Virol* **4**, 169–177 (2015).

146. Borghetti, A. *et al.* SLC22A2 variants and dolutegravir levels correlate with psychiatric symptoms in persons with HIV. *J Antimicrob Chemother* **74**, 1035–1043 (2019).

147. Lakhman, S. S., Ma, Q. & Morse, G. D. Pharmacogenomics of CYP3A: considerations for HIV treatment. *Pharmacogenomics* **10**, 1323–39 (2009).

148. Ngaimisi, E. *et al.* Importance of Ethnicity, CYP2B6 and ABCB1 Genotype for Efavirenz Pharmacokinetics and Treatment Outcomes: A Parallel-Group Prospective Cohort Study in Two Sub-Saharan Africa Populations. *PLoS ONE* **8**, (2013).

149. Ngaimisi, E. *et al.* Effect of Rifampicin and CYP2B6 Genotype on Long-Term Efavirenz Autoinduction and Plasma Exposure in HIV Patients With or Without Tuberculosis. *Clinical Pharmacology & Therapeutics* **90**, 406–413 (2011).

150. Nyakutira, C. *et al.* High prevalence of the CYP2B6 516G→T(*6) variant and effect on the population pharmacokinetics of efavirenz in HIV/AIDS outpatients in Zimbabwe. *European Journal of Clinical Pharmacology* **64**, 357–365 (2008).

151. Ma, J. D., Lee, K. C. & Kuo, G. M. HLA-B*5701 testing to predict abacavir hypersensitivity. *PLoS Currents* **2**, (2010).

152. HIV Drug Resistance Testing, Genotype | Labcorp. https://www.labcorp.com/help/patient-test-info/hiv-drug-resistance-testing-genotype.

153. HIRGT - Overview: HIV-1 RNA Quantification with Reflex to HIV-1 Genotypic Drug Resistance to Protease and Reverse Transcriptase Inhibitors, Plasma. https://www.mayocliniclabs.com/test-catalog/Overview/65713.

154. HL57R - Overview: HLA-B*57:01 Genotype, Pharmacogenomics, Varies. https://www.mayocliniclabs.com/test-catalog/Overview/610054.

155. Ngaimisi, E. *et al.* Importance of Ethnicity, CYP2B6 and ABCB1 Genotype for Efavirenz Pharmacokinetics and Treatment Outcomes: A Parallel-Group Prospective Cohort Study in Two Sub-Saharan Africa Populations. *PLOS ONE* **8**, e67946 (2013).

156. Ratain, M. J. & William K. Plunkett, J. Principles of Pharmacokinetics. (2003).

157. Holford, N. & Yim, D. S. Volume of Distribution. *Translational and Clinical Pharmacology* **24**, 74–77 (2021).

158. NRTI Resistance Notes - HIV Drug Resistance Database. https://hivdb.stanford.edu/dr-summary/resistance-notes/NRTI/.

159. Perelson, A. S. & Ribeiro, R. M. Modeling the within-host dynamics of HIV infection. *BMC Biology* **11**, 1–10 (2013).

160. Fda. HIGHLIGHTS OF PRESCRIBING INFORMATION.

161. Tang, M. W., Liu, T. F. & Shafer, R. W. The HIVdb system for HIV-1 genotypic resistance interpretation. *Intervirology* **55**, 98–101 (2012).

162. Haas, D. W. *et al.* Pharmacogenetics of nevirapine-associated hepatotoxicity: an Adult AIDS Clinical Trials Group collaboration. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **43**, 783–786 (2006).

163. Haas, D. W. *et al.* Pharmacogenomics of HIV therapy: summary of a workshop sponsored by the National Institute of Allergy and Infectious Diseases. *HIV Clin Trials* **12**, 277–85 (2011).

164. Lin, R. S., Rhee, S., Shafer, R. W. & Das, A. K. Prediction of HIV Mutation Changes based on Treatment History Stanford Medical Informatics and 2 Division of Infectious Diseases. in *AMIA Symposium Proceedings* (2006).

165. Gregson, J. *et al.* Global epidemiology of drug resistance after failure of WHO recommended first-line regimens for adult HIV-1 infection: a multicentre retrospective cohort study. *Lancet Infect Dis* **16**, 565–575 (2016).

166. Gupta, R. K. *et al.* HIV-1 drug resistance before initiation or re-initiation of first-line antiretroviral therapy in low-income and middle-income countries: A systematic review and meta-regression analysis. *The Lancet Infectious Diseases* (2017) doi:10.1016/S1473-3099(17)30702-8.

167. Leger, P. *et al.* Pharmacogenetics of efavirenz discontinuation for reported central nervous system symptoms appears to differ by race. *Pharmacogenetics and Genomics* **26**, 473–480 (2016).

168. Pennings, P. S., Kryazhimskiy, S. & Wakeley, J. Loss and Recovery of Genetic Diversity in Adapting Populations of HIV. *PLOS Genetics* **10**, e1004000 (2014).

169. Maynard Smith, J. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genetics Research* **89**, 391–403 (2007).

170. WHO. WHO | HIV/AIDS. *WHO* (2016).

171. Clutter, D. S., Jordan, M. R., Bertagnolio, S. & Shafer, R. W. HIV-1 Drug Resistance and Resistance Testing. *Infect Genet Evol* **46**, 292 (2016).

172. HIV DRUG RESISTANCE REPORT 2017 TRENDS QUALITY ACTION.

173. Chimukangara, B. *et al.* Trends in Pretreatment HIV-1 Drug Resistance in Antiretroviral Therapy-naive Adults in South Africa, 2000–2016: A Pooled Sequence Analysis. *EClinicalMedicine* **9**, 26–34 (2019).

174. Gupta, S. & Neogi, U. Following the path: Increasing trends of HIV-1 drug resistance in China. *EClinicalMedicine* vol. 18 (2020).

175. Gupta, R. K. *et al.* HIV-1 drug resistance before initiation or re-initiation of first-line antiretroviral therapy in low-income and middle-income countries: a systematic review and meta-regression analysis. *The Lancet Infectious Diseases* **18**, 346–355 (2018).

176. Crowell, T. A. *et al.* Pretreatment and Acquired Antiretroviral Drug Resistance Among Persons Living With HIV in Four African Countries. *Clinical Infectious Diseases* (2020) doi:10.1093/cid/ciaa1161.

177. Coelho, L. P. O. *et al.* Prevalence of HIV-1 transmitted drug resistance and viral suppression among recently diagnosed adults in São Paulo, Brazil. *Archives of Virology* **164**, 699–706 (2019).

178. Zou, X. *et al.* Prevalence of acquired drug resistance mutations in antiretroviral-experiencing subjects from 2012 to 2017 in Hunan Province of central South China. *Virology Journal* **17**, 1–9 (2020).

179. Schultze, A. *et al.* HIV resistance testing and detected drug resistance in Europe. *AIDS* **29**, 1379–1389 (2015).

180. LM, H. *et al.* Transmission of HIV Drug Resistance and the Predicted Effect on Current First-line Regimens in Europe. *Clin Infect Dis* **62**, 655–663 (2016).

181. Pretreatment human immunodeficiency virus type 1 (HIV-1) drug resistance in transmission clusters of the Cologne-Bonn region, Germany | Elsevier Enhanced Reader. https://reader.elsevier.com/reader/sd/pii/S1198743X18306657?token=E86346FA42577D625B5F031AF351D25C3044C518A38CFB8CE403CF62FECA09A9114E8FB1E5863AB6D1D61D92411AEBE2&originRegion=us-east-1&originCreation=20210612165955.

182. Stecher, M. *et al.* Molecular Epidemiology of the HIV Epidemic in Three German Metropolitan Regions – Cologne/Bonn, Munich and Hannover, 1999–2016. *Scientific Reports 2018 8:1* **8**, 1–9 (2018).

183. Wainberg, M. A. & Han, Y. S. Will drug resistance against dolutegravir in initial therapy ever occur? *Frontiers in Pharmacology* vol. 6 (2015).

184. Levintow, S. N. *et al.* Prevalence and transmission dynamics of HIV-1 transmitted drug resistance in a southeastern cohort. *Open Forum Infectious Diseases* **5**, (2018).

185. Geographic Distribution | Statistics Overview | Statistics Center | HIV/AIDS | CDC. https://www.cdc.gov/hiv/statistics/overview/geographicdistribution.html.

186. Rich, S. N. *et al.* Sociodemographic, Ecological, and Spatiotemporal Factors Associated with Human Immunodeficiency Virus Drug Resistance in Florida: A Retrospective Analysis. *J Infect Dis* **223**, 866–875 (2021).

187. Drug resistance, strain, and cluster monitoring - King County. https://kingcounty.gov/depts/health/communicable-diseases/hiv-std/patients/epidemiology/drug-resistance-surveillance.aspx.

188. Chimbetete, C. *et al.* HIV-1 Drug Resistance and Third-Line Therapy Outcomes in Patients Failing Second-Line Therapy in Zimbabwe. *Open Forum Infectious Diseases* **5**, (2018).

189. Erick Ntambwe, K. Genetic Diversity and Antiretroviral Drug Resistance among Drug-Naïve HIV Type 1 Infected Patients attending Clinics in Kinshasa, Democratic Republic of Congo. *Journal of HIV and AIDS (ISSN 2380-5536)* **1**, (2015).

190. Wang, Y., Zhang, X. S. & Chen, L. Optimization meets systems biology. *BMC Systems Biology* **4**, S1 (2010).

191. Algorithms for Optimization - Google Books. https://www.google.com/books/edition/Algorithms_for_Optimization/uBSMDwAAQBAJ?hl=en&gbpv=1&dq=optimization+algorithm+defined+computer+science&printsec=frontcover.

192. Nagarajan, G. & Dhinesh Babu, L. D. A hybrid of whale optimization and late acceptance hill climbing based imputation to enhance classification performance in electronic health records. *Journal of Biomedical Informatics* **94**, (2019).

193. Osareh, A. & Shadgar, B. Machine learning techniques to diagnose breast cancer. *2010 5th International Symposium on Health Informatics and Bioinformatics, HIBIT 2010* 114–120 (2010) doi:10.1109/HIBIT.2010.5478895.

194. Usha Nandhini, A. & Dharmarajan, K. Analysis of Diabetic Mellitus Using Predictive Algorithm – A Literature Review. (2021) doi:10.4108/EAI.16-5-2020.2304019.

195. Navascués, M., Budroni, C. & Guryanova, Y. Disease control as an optimization problem. *PLOS ONE* **16**, e0257958 (2021).

196. Liashchynskyi, P. & Liashchynskyi, P. Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS. (2019).

197. Connecting Medical Informatics and Bio-informatics: Proceedings of MIE2005 ... - Google Books. https://books.google.com/books?hl=en&lr=&id=HXTk5rWOdG4C&oi=fnd&pg=PA193&dq=grid+search+optimization+health&ots=ZVfCWAxMbb&sig=EwB1xuwS1sROelekjtQeVERdjSA#v=onepage&q&f=false.

198. Introduction to Hill Climbing | Artificial Intelligence - GeeksforGeeks. https://www.geeksforgeeks.org/introduction-hill-climbing-artificial-intelligence/.

199. Hill Climbing Algorithm | Hill Climbing Algorithm in AI | Edureka. https://www.edureka.co/blog/hill-climbing-algorithm-ai/.

200. Ruder, S. An overview of gradient descent optimization algorithms. (2016).

201. Aarts, E. H. L., Korst, J. H. M. & van Laarhoven, P. J. M. A Quantitative Analysis of the Simulated Annealing Algorithm: A Case Study for the Traveling Salesman Problem. *Journal of Statistical Physics* **50**, (1988).

202. Aarts, E. H. L. & van Laarhoven, P. J. M. Simulated annealing: An introduction. *Stat Neerl* **43**, 31–52 (1989).

203. Henderson, D., Jacobson, S. H. & Johnson, A. W. The Theory and Practice of Simulated Annealing. *Handbook of Metaheuristics* 287–319 (2003) doi:10.1007/0-306-48056-5_10.

204. Roumeliotis, M., Yates, B., Watt, E., Frederick, A. & Meyer, T. Demonstration of simulated annealing optimization for permanent breast seed implant treatment planning. *Brachytherapy* **17**, 615–620 (2018).

205. Pennisi, M., Catanuto, R., Pappalardo, F. & Motta, S. Optimal vaccination schedules using simulated annealing. *Bioinformatics* **24**, 1740–1742 (2008).

206. Taffa, N. *et al.* Pretreatment HIV drug resistance among adults initiating ART in Namibia. *Journal of Antimicrobial Chemotherapy* **73**, 3137–3142 (2018).

207. Phillips, A. N. *et al.* Impact of HIV drug resistance on HIV/AIDS-associated mortality, new infections, and antiretroviral therapy program costs in Sub-Saharan Africa. *Journal of Infectious Diseases* **215**, 1362–1365 (2017).

208. Inzaule, S. C. *et al.* Affordable HIV drug-resistance testing for monitoring of antiretroviral therapy in sub-Saharan Africa. *The Lancet Infectious Diseases* vol. 16 e267–e275 (2016).

209. Manyana, S. *et al.* HIV-1 Drug Resistance Genotyping in Resource Limited Settings: Current and Future Perspectives in Sequencing Technologies. *Viruses* **13**, 1125 (2021).

210. Gachogo, R. W., Mwai, D. N. & Onyambu, F. G. Cost analysis of implementing HIV drug resistance testing in Kenya: A case study of a service delivery site at a tertiary level hospital in Kenya. *F1000Res* **9**, (2020).

211. Mittler, J., Goodreau, S. & Herbeck, J. Integrated bio-social models for HIV epidemiology.

212. Herbeck, J. T., Mittler, J. E., Gottlieb, G. S. & Mullins, J. I. An HIV Epidemic Model Based on Viral Load Dynamics: Value in Assessing Empirical Trends in HIV Virulence and Community Viral Load. *PLoS Computational Biology* **10**, (2014).

213. Curlin, M. E., Iyer, S. & Mittler, J. E. Optimal Timing and Duration of Induction Therapy for HIV-1 Infection. *PLoS Computational Biology* **3**, 1239–1256 (2007).

214. Herbeck, J. T. *et al.* Evolution of HIV virulence in response to widespread scale up of antiretroviral therapy: a modeling study. *Virus Evol* **2**, vew028 (2016).

215. GitHub - EvoNetHIV/EvoNetHIV: Current development version of EvoNetHIV package. https://github.com/EvoNetHIV/EvoNetHIV.

216. Computing Resources | Center for Studies in Demography and Ecology. https://csde.washington.edu/computing/resources/#Sim_Details.

217. Health, E. Program. D. of H. Protection. G. D. of P. HIV Epidemiologic Profile for Georgia, and for the 4 " Ending the HIV Epidemic " Jurisdictions June 2021. 65 (2021).

218. Mabaso, M. *et al.* Trends and correlates of HIV prevalence among adolescents in South Africa: evidence from the 2008, 2012 and 2017 South African National HIV Prevalence, Incidence and Behaviour surveys. *AIDS Research and Therapy 2021 18:1* **18**, 1–8 (2021).

219. 90–90–90: good progress, but the world is off-track for hitting the 2020 targets | UNAIDS. https://www.unaids.org/en/resources/presscentre/featurestories/2020/september/20200921_90-90-90.

220. Maseng, M. J. *et al.* Association of CYP2B6 Genetic Variation with Efavirenz and Nevirapine Drug Resistance in HIV-1 Patients from Botswana. *Pharmgenomics Pers Med* **14**, 335–347 (2021).

221. Rotger, M. *et al.* Influence of CYP2B6 polymorphism on plasma and intracellular concentrations and toxicity of efavirenz and nevirapine in HIV-infected patients. *Pharmacogenet Genomics* **15**, 1–5 (2005).

222. Xu, C., Ogburn, E. T., Guo, Y. & Desta, Z. Effects of the CYP2B6*6 allele on catalytic properties and inhibition of CYP2B6 in vitro: Implication for the mechanism of reduced efavirenz metabolism and other CYP2B6 substrates in vivo. *Drug Metabolism and Disposition* **40**, 717–725 (2012).

223. Dhoro, M. *et al.* CYP2B6*6, CYP2B6*18, Body weight and sex are predictors of efavirenz pharmacokinetics and treatment response: Population pharmacokinetic modeling in an HIV/AIDS and TB cohort in Zimbabwe. *BMC Pharmacology and Toxicology* **16**, (2015).

224. Gandhi, M. *et al.* A single-nucleotide polymorphism in CYP2B6 leads to >3-fold increases in efavirenz concentrations in plasma and hair among HIV-infected women. *J Infect Dis* **206**, 1453–1461 (2012).

225. Paganotti, G. M. *et al.* CYP2B6 poor metaboliser alleles involved in efavirenz and nevirapine metabolism: CYP2B6*9 and CYP2B6*18 distribution in HIV-exposed subjects from Dschang, Western Cameroon. *Infection, Genetics and Evolution* **35**, 122–126 (2015).

226. Mukonzo, J., Okwera, A., Nakasujja, N., Luzze, H. & Sebuwufu, D. Influence of efavirenz pharmacokinetics and pharmacogenetics on neuropsychological disorders in Ugandan HIV-positive patients with or without tuberculosis: a prospective cohort study. *BMC Infect Dis* **13**, (2013).

227. Gengiah, T. N. *et al.* The influence of tuberculosis treatment on efavirenz clearance in patients co-infected with HIV and tuberculosis. *European Journal of Clinical Pharmacology* **68**, 689–695 (2012).

228. Maseng, M. J. *et al.* Association of CYP2B6 Genetic Variation with Efavirenz and Nevirapine Drug Resistance in HIV-1 Patients from Botswana. *Pharmacogenomics and Personalized Medicine* **14**, 335 (2021).

229. Feder, A. F., Harper, K. N., Brumme, C. J. & Pennings, P. S. Understanding patterns of HIV multi-drug resistance through models of temporal and spatial drug heterogeneity. *Elife* **10**, (2021).

230. UPDATE OF RECOMMENDATIONS ON FIRST-AND SECOND-LINE ANTIRETROVIRAL REGIMENS. (2019).

231. Nwogu, J. N. *et al.* Associations between efavirenz concentrations, pharmacogenetics and neurocognitive performance in people living with HIV in Nigeria. *AIDS* **35**, 1919–1927 (2021).

232. Amin, J. *et al.* Efficacy of 400 mg efavirenz versus standard 600 mg dose in HIV-infected, antiretroviral-naive adults (ENCORE1): a randomised, double-blind, placebo-controlled, non-inferiority trial. *Lancet* **383**, 1474–1482 (2014).

233. Desta, Z. *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2B6 and Efavirenz-Containing Antiretroviral Therapy. **106**, (2019).

234. Efficacy of 400 mg Efavirenz Versus Standard 600 mg Dose in HIV/TB Co-infected Patients - Full Text View - ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT04513379.

235. FDA Approves Cabenuva and Vocabria for the Treatment of HIV-1 Infection | FDA. https://www.fda.gov/drugs/human-immunodeficiency-virus-hiv/fda-approves-cabenuva-and-vocabria-treatment-hiv-1-infection.

236. Jaeger, H. *et al.* Long-acting cabotegravir and rilpivirine dosed every 2 months in adults with HIV-1 infection (ATLAS-2M), 96-week results: a randomised, multicentre, open-label, phase 3b, non-inferiority study. *The Lancet HIV* **8**, e679–e689 (2021).