7-28-2020 2:00 PM

Multi-Scale Evolution of Virulence of HIV-1

David W. Dick, *The University of Western Ontario*

Supervisor: Lindi Wahl, *The University of Western Ontario*
: Art Poon, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Applied Mathematics

Abstract

HIV-1 is a rapidly replicating retrovirus that faces two distinct fitness landscapes: within-host HIV-1 faces viral competition for host cells and for escape from the immune system, and between hosts HIV-1 faces a transmission bottleneck in which the majority of new infections are started by a single virus strain. Possibly as a result of these conflicting selective pressures, the rate of evolution of HIV-1 tends to be greater within-host than between hosts.

A current hypothesis for this difference in evolutionary rates is that the HIV-1 latent reservoir acts to archive virus for later transmission. We offer a related but complimentary hypothesis: while some of the viruses' life history traits are under selective pressure within-host, traits that are responsible for the efficiency of transmission to a new host are not under direct selection within-host and thus are subject to drift. Combined with the necessity of transmission through an extremely severe, competitive bottleneck, this results in the preferential transmission of founder-like viral lineages.

As further evidence of the conflict between transmission fitness and within-host fitness, experimental evidence demonstrates that subtypes A and D are 100-fold more fit than subtype C in *in vitro* fitness competitions, yet subtype C dominates the global spread of new infections. It is unclear whether this discrepancy is caused by differences in within- and between-host fitness, or primarily reflects differences in *in vitro* versus *in vivo* fitness measures. To address this question, data from a four-year, 8000 participant study in Uganda and Zimbabwe were analyzed for evidence of *in vivo* fitness differences between subtypes A, C and D. Analyzing this dataset along with simulated participant data, we conclude that either more frequent data sampling, or an even larger study, would be necessary to capture the early within-host dynamics sufficiently for a comparison across subtypes.

Similar to subtypes A and D, subtype B is estimated to have an eight- to ten-fold *in vitro* fitness advantage over subtype C. Since frequent data collection over the early course of infection is necessary to quantify *in vivo* viral fitness, another approach to this question is to use data collected for simian/human immunodeficiency virus (SHIV). We develop a non-linear mixed-effects model for a meta-analysis of 143 non-human primates from over 20 sources to study *in vivo* fitness differences between SHIV subtypes B and C. Results suggest that subtype C has a lower replicative fitness but higher burst size than subtype B.

Keywords: HIV, SHIV, multi-scale, modeling, simulation, within-host, between-host, subtype, evolution

Summary for lay audiences

Human immunodeficiency virus (HIV) is a rapidly evolving virus that faces two distinct environments. Within a single infected individual ("within host"), HIV is engaged in a race with the immune system, evolving to avoid the host defences. When transmitted from one infected individual to another ("between hosts"), HIV faces an extreme transmission bottleneck. A faster evolutionary rate is observed within-host than between-hosts. We hypothesize that the ability of the virus to transmit across the extreme transmission bottleneck is an important factor for this difference in the evolutionary rates, and that when a virus mutates within-host the ability for the virus to transmit is usually reduced. This leads to the transmission of viruses that are more like the virus that established the infection than the average virus within-host at the time of a transmission.

Further evidence of this conflict between the virus' ability to out compete other viral strains within-host and the ability of the virus to transmit is seen in the discrepancy between laboratory experiments, where HIV subtypes A and D replicate more quickly than subtype C, and the global epidemic, which is increasingly dominated by HIV subtype C. To address this conflict we consider data from a four year, 8000 participant study in Uganda and Zimbabwe where differences between subtypes in the study participants were considered. To detect the ability of one viral subtype to out-compete another subtype, samples from early in a participant's infection are required. By simulating participant's appointment timing and infection time courses, we were able to determine that the data-set did not contain the information needed to detect differences in within-host replication rates between subtypes.

In order to detect possible differences between HIV subtypes we consider the constructed virus, simian/human immunodeficiency virus (SHIV), which infects macaques and shares the subtype differences of HIV. Developing a statistical model and gathering data from over 143 non-human primates infected with SHIV from over 20 previous studies, we were able to detect distinct replication advantages for both subtype C and D. Subtype C has a lower ability to infect new cells, within-host, but generates more new virus from each cell it infects than subtype D.

Contents

List of Figures

List of Tables

List of Appendices

Chapter 1

Introduction

1.1 Human immunodeficiency virus (HIV)

Human immunodeficiency virus (HIV) is a positive sense single stranded retrovirus that primarily replicates in CD4+ immune cells [Barre-Sinoussi et al., 2013, Joseph et al., 2015]. HIV-1 faces two distinct fitness landscapes over two distinct time scales: within host HIV-1 faces selection pressure from the hosts' immune system and viral competition for the hosts' immune cells, and between hosts HIV-1 faces a transmission bottleneck, after which the majority of new infections are initially homogeneous. This homogeneous initial population of virus, seen shortly after transmission, diversifies into a viral quasi-species seen in active infection [Joseph et al., 2015, van Dorp et al., 2014].

The differentiation of HIV-1 into a genetically diverse quasi-species is facilitated by the error prone RNA replication process, resulting in a high mutation rate, along with a high replication rate within-host [Cuevas and Sanjuán, 2015, Fraser et al., 2014]. This HIV-1 mutation potential allows the virus to increase in within-host fitness as it adapts to the host and evolves to escape immune pressures [Carlson et al., 2016, Payne et al., 2014, Immonen et al., 2012].

Currently available antiretroviral therapy (ART) is effective at suppressing viral load in the plasma below detectable levels and halting viral evolution, but after treatment interruption the

virus typically rebounds to pre-treatment levels [Conway et al., 2019, Davenport et al., 2019]. There were 37.9 million people living with HIV at the end of 2018, with an estimated 62% on antiretroviral therapy, and 53% of people living with HIV have achieved suppression of the virus.

1.2 HIV-1 diversity: groups and subtypes

HIV is a zoonotic virus and the origins of the current HIV-1 epidemic are estimated to have been established in the human population around the 1920s near Kinshasa in the present day Democratic Republic of Congo (DRC) [Bbosa et al., 2019, Tebit and Arts, 2011]. Early in the epidemic two genetically diverse lentaviruses, HIV-1 and HIV-2, were prevalent but geographically separated and established through multiple zoonotic transmissions between simian immunodeficiency virus in primates and humans [Bbosa et al., 2019]. But since the introduction of HIV-2, its prevalence has reduced to near extinction [Ariën et al., 2005], while the current diversity of HIV-1 is high near its origin in the DRC near Kinshasa (formerly Leopoldville) [Bbosa et al., 2019]. The accuracy of dating the origin of the current epidemic and estimating the many cross-species jumps of HIV depends on an estimate of the rate of between-host evolution over the epidemic and the prevalence of recombination for this virus [Olabode et al., 2019].

Divergent evolution combined with the multiple zoonotic jumps has lead to the diversity of HIV-1 that we now see in the human population [Ball et al., 2003]. Separate introductions of HIV-1 in the human population led to the spread of divergent groups of HIV-1: groups M (main), O (outlier), group N and group P viruses [Bbosa et al., 2019]. Groups N, O and P have not globally spread and the majority of the epidemic is from group M viruses. Group M viruses are subdivided into ten different subtypes or clades (labeled subtypes A through J) based on the diversity of the *eve* gene [Bbosa et al., 2019, Ball et al., 2003]. The genetic diversity of HIV-1 is not completely captured by the HIV-1 subtype classification and there are at least ninety

circulating recombinant forms [Olabode et al., 2019]. The genetic distance across subtypes is estimated to be between 25 and 35% and the genetic distance within subtypes is estimated to be up to 20% [Bbosa et al., 2019]. This extreme genetic diversity between HIV-1 subtypes leads to the question whether HIV-1 genotypes are associated with measurable biological or phenotypic traits [Tebit and Arts, 2011]. Some differences in phenotypic traits have been observed, with greater rates of treatment failure and drug resistance in subtype D than subtype A [Kyeyune et al., 2013].

The genetically distinct subtypes have geographically defined distributions with subtype C and inter-subtype recombinant forms found in Africa and subtype B prevalent in North America and Europe [Bbosa et al., 2019]. The domain of subtype C has expanded since 1990 and its spread accounts for approximately 53% of the 27.2 million new infections between 1990 and 2015 [Venner et al., 2016].

1.2.1 Simian/human immunodeficiency virus (SHIV)

Animal models of HIV have been developed by combining genes from HIV and simian immunodeficiency virus (SIV) to create the novel pathogen simian/human immunodeficiency virus (SHIV). SHIV infects non-human primates and has been developed to facilitate the study of HIV transmission, pathogenesis, and the development of HIV vaccines [U et al., 2001]. SHIV variants typically have the *env*, *tat*, *rev*, and *vpu* genes of HIV while the remaining genes are from SIV [Sui et al., 2014]. SHIV variants with the envelope from subtypes A, B, C, and E have been created [Sui et al., 2014]. SHIV created from the SIVmac239 backbone expresses the chemokine receptor CXCR or CCR5, or both the CXCR4 and CCR5 receptors, to allow cell entry. These viruses cause greater depletion of naïve T cells during the acute phase of infection, leading to more rapid disease progression than is seen in either a SIV or HIV infection [Sui et al., 2014].

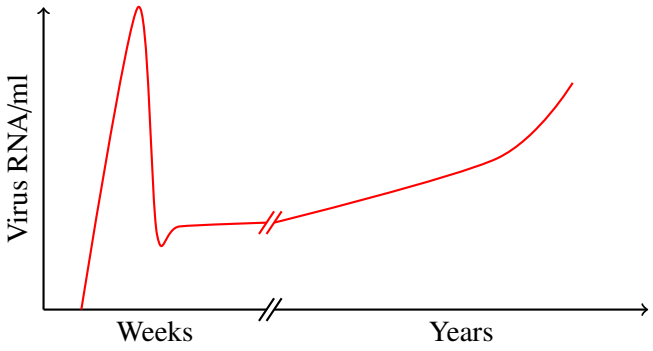


Figure 1.1: Generalized sketch of HIV disease progression

1.3 Within-host dynamics: disease progression

Typical within-host progression of HIV and the times to detectable viral load and viral antibodies have been divided into the eclipse phase and five Fiebig stages [Shaw and Hunter, 2012]. A sketch of the typical HIV-1 viral load dynamics as detected in the peripheral blood is seen in Figure 1.1. After initial infection there is an eclipse phase, that lasts between 5 to 14 days, during which the HIV-1 viral load is typically below detectable levels [Shaw and Hunter, 2012, Konrad et al., 2016]. During the rapid growth of the within-host viral load to a viral peak around 21 days post infection, viral RNA is detectable in Fiebig stage I and viral p21 antigen is detectable in Fiebig stage II [Shaw and Hunter, 2012]. Near the peak viral load, virus-specific antibodies become detectable first by recombinant protein-based, enzyme-linked immunosorbant assay (ELISA) in Fiebig stage III [Shaw and Hunter, 2012]. The duration of Fiebig stages I, II, and III were estimated to be 5.0, 5.3, and 3.2 days, respectively [Konrad et al., 2016]. As the viral load drops to a relative steady state, the set-point viral load (SPVL), antibodies become detectable by Western immunoblotting with an indeterminate banding pattern, stage IV, and a diagnostic banding pattern stage V [Shaw and Hunter, 2012]. In addition to the circulating virus there is extensive compartmentalization of HIV-1 [Chen et al., 2018, Boritz et al., 2016, Gianella et al., 2016, Feder et al., 2017].

Effective antiretroviral therapy reduces HIV-1 viral replication and disease progression, lowering virus in the blood to below detectable levels. As mentioned above, after treatment

interruption, viral replication rebounds to pre-therapy levels [Conway et al., 2019, Davenport et al., 2019]; HIV-1 forms a latent reservoir of infected, long-lived immune cells that allows for this viral rebound [Kuo and Lichterfeld, 2018].

The formation of the HIV-1 latent reservoir is not well characterized. There is conflicting evidence about the formation of the the HIV-1 latent reservoir, with evidence that indicates that the latent reservoir forms rapidly early in infection [Whitney et al., 2014, Sengupta and Siliciano, 2018] or that the reservoir's formation occurs primarily upon initiation of ART treatment [Abrahams et al., 2019]. Abrahams et. al. argue cells are infected as they are entering a resting state and that 78% of virus in the latent reservoir is most similar to virus in circulation just before the initiation of therapy [Abrahams et al., 2019]. Uncertainty about the composition of the HIV latent reservoir leads to uncertainty about its formation [Poon et al., 2018]. Determining the relative size and makeup of the latent reservoir is difficult because there is an abundance of fragments of nonproductive RNA archived in the latent reservoir and quantitative viral outgrowth assays are required [Abrahams et al., 2019].

The HIV-1 latent reservoir contributes to circulating diversity by archiving virus that closely resembles founding virus. This archive of founder-like virus is sufficiently large to allow for further transmission [Immonen and Leitner, 2014]. The relative stability of the latent reservoir, which is contributed to by a slow rate of decay and clonal expansion, allows this archive of founder-like virus to persist throughout the infection [Descours et al., 2017].

1.4 Between-host dynamics: conflicting selection pressures

Between hosts HIV-1 faces a restrictive transmission bottleneck [Kariuki et al., 2017, Lythgoe et al., 2017, Joseph et al., 2015]. When transmitted between individuals, HIV-1 undergoes a severe, and multistage, bottleneck; it is possible that only a single variant establishes the circulating infection in the new host [Carlson et al., 2014a, Tully et al., 2016, Haaland et al., 2009, Li et al., 2010, Keele et al., 2008, Zimmer, 2002]. This bottleneck persists across different modes

of transmission with estimates of: 80% of heterosexual [Carlson et al., 2014a], 75% of male-to-male homosexual (MSM) [Li et al., 2010], 70% of mother-to-child [Wolinsky et al., 1992], and 40 to 80% of intravenous drug-use transmissions originate from a single variant [Kariuki et al., 2017, Tully et al., 2016]. Clinical evidence also supports preferential transmission of some viral strains [Theys et al., 2018, Lythgoe et al., 2017, Deymier et al., 2015, Carlson et al., 2014a, Redd et al., 2012, Wolinsky et al., 1992] suggesting that the bottleneck is not strictly stochastic.

The conflicting selection pressures of the within- and between-host fitness landscapes lead to a trade-off between within-host virulence and between-host infectiousness [Theys et al., 2018, Lythgoe et al., 2017]. Understanding this trade-off and the viral strains that are transmitted and initiate infection is critical in the development of vaccines and prevention strategies for HIV-1 [Lythgoe et al., 2017, Kariuki et al., 2017], and has been addressed through clinical studies in the context of heritability of virulence [Bertels et al., 2018, Blanquart et al., 2017, Tully et al., 2016, Cuevas and Sanjuán, 2015].

This fitness trade-off is also partially manifest as a mismatch in within- and between-host evolutionary rates, where the estimated between-host rate is many times lower than the rate observed within-host [Vrancken et al., 2014, Theys et al., 2018, Volz et al., 2017], a discrepancy that has important consequences for phylodynamic inference [Volz et al., 2017].

1.5 Mathematical modelling approaches

Substantial progress in understanding these conflicting between- and within-host selection pressures has also been achieved through modelling approaches.

In early work, Fraser et al. explored the evolution of set-point viral load (SPVL). Since SPVL increases transmissibility but reduces the duration of infection, modelling predicted that SPVLs should cluster around intermediate values [Fraser et al., 2007, Shirreff et al., 2011].

In related work, Lythgoe and Fraser proposed a within-host multi-strain model, which also

included a reservoir of latently infected cells. The model demonstrates that the latent reservoir is able to archive virus that closely resembles the founding virus; virus from this latent archive is then available for further transmission. Comparing a number of hypotheses, Lythgoe and Fraser argue that this *store and retrieve* mechanism is the most important factor underlying observed differences in evolutionary rates [Lythgoe and Fraser, 2012].

This suggestion is consistent with the conclusions of Immonen and Leitner, who, using patient data, phylogenetic methods, and a detailed within-host model of HIV-1 infection predicted that the latent reservoir substantially slows observed within-host evolution.

Focusing further on the evolution of virulence, Lythgoe et al. introduced a deterministic multi-scale model which links within- and between-host evolution [Lythgoe et al., 2013]. The authors conclude that slower within-host evolution leads to intermediate virulence, whereas with faster within-host dynamics evolution is short-sighted, resulting in selection for higher within-host fitness and higher virulence. In this approach, the within- and between-host scales are linked through a non-stochastic transmission event; in particular, every viral strain in the donor is transmitted to the recipient, in proportion to the circulating abundance in the donor at the time of transmission.

Doekes et al. [Doekes et al., 2017] further developed this approach, replacing the within-host model with two coupled quasi-species equations [Eigen, 1971] such that the latent reservoir is explicitly included. The presence of a latent compartment slows the within-host dynamics when considering moderate fitness changes, but has little effect when there are large fitness differences driving the within-host dynamics [Doekes et al., 2017]. In the supporting material, Doekes et al. extend this model to include mutations that have no effect on SPVL (and thus drift neutrally within host), but reduce between-host transmission. The presence of preferential transmission further slows the between-host evolution of SPVL [Doekes et al., 2017].

1.6 Outline of the thesis

An increasingly clear picture is thus emerging of the factors affecting between-host evolution in HIV-1: the trade-off between high SPVL and low duration of infection, the store-and-retrieve mechanism facilitated by the latent archive, and mutational drift of transmission fitness during in-host evolution. To date, however, the genetic bottleneck associated with disease transmission has received comparatively little attention. Since only one or a handful of viral strains are likely to found the next infection, the bottleneck not only eliminates nearly all genetic diversity, but likely imposes intense selection for strains in which traits important to disease transmission have retained their function.

In Chapter 2, we develop a multi-scale model of the epidemiology and evolution HIV-1, linking the within- and between-host scales through a stochastic, competitive transmission bottleneck. While within-host selection favours increased virulence in our model, traits that are responsible for the efficiency of transmission to a new host are not under direct selection within-host and thus are subject to drift. Combined with the necessity of transmission through an extremely severe, competitive bottleneck, this results in the preferential transmission of founder-like viral lineages that more closely resemble the founding virus than virus circulating within the host at the time of transmission. These results are presented in Chapter 3.

In Chapter 4, we examine appointment timing and viral load data from a large study of 302 women from Uganda and Zimbabwe to determine whether fitness differences in HIV-1 among subtypes are detectable in this dataset.

In Chapter 5, we develop a non-linear mixed-effects model for a meta-analysis of 143 primates from over 20 sources to consider fitness differences between infection with SHIV subtype C (SHIV 1157) and subtype B (SHIV SF162).

Chapter 6 presents some brief conclusions and suggestions for future work.

Chapter 2

Multi-scale model of the evolution of HIV-1: Methods

2.1 Introduction

We develop multi-scale simulations of the evolution of HIV-1 over epidemic time, simulating the evolution of viral traits in response to the two distinct fitness landscapes that HIV-1 faces, within- and between-host selective pressures. We couple the within- and between-host dynamics with attention to mutations that affect specific life history traits of the virus.

The within-host dynamics are governed by a host-cell limited, differential equation model of viral dynamics [Nowak and May, 2000]. The within-host model includes evolution through phenotypic mutation classes [Lythgoe et al., 2013] with increasing fitness and a latent compartment in which evolution does not occur [Doekes et al., 2017], representing the HIV latent reservoir.

For the between-host dynamics, three models of decreasing complexity are developed, as described in further detail in Sections 2.3.1 to 2.3.3. The first and most complete model of the between-host transmission dynamics of HIV-1 is used to address the role of the HIV-1 latent reservoir in the observed differences in evolutionary rate within- and between-host, and

to analyze the application of the germline hypotheses [Lythgoe et al., 2017] to the epidemic evolution of HIV-1.

The first and most complete model of between-host dynamics will be subsequently referred to as the *full model*. In the full model, the within-host dynamics are linked to a between-host stochastic model of virus evolution including transmission fitness.

The second *fixed transmission* model is established as a control for comparison against the full model. In the fixed model, the probability of each virion establishing an infection in the recipient host, *transmission fitness*, is independent of the within-host model.

In the third model of between-host transmission, the *simplified model*, transmission fitness is neglected. This model is more computationally efficient and is used to determine bounds on the expected behavior of the more complete models. This computationally efficient model is used to test the sensitivity of the simulation to parameter and specification changes, as well as to establish the most informative parameter regimes to consider with the more complex models.

2.2 Within-host model

The within-host viral dynamics are determined by a compartmental model of CD4+ T cells with n distinct phenotypic mutation classes, as shown in system (2.1), where the time dependence t has been omitted for clarity.

The model considers the densities of uninfected target cells $x(t)$, infected cells $y_i(t)$, and latently infected cells $l_i(t)$ for mutation classes $i \in \{0, 1, 2, \ldots, n-1\}$, where densities are per unit blood volume.

In this model we neglect the usual compartment modelling free virus. The dynamics of the free virus compartment closely follow the dynamics of the infected cells with most of the relevant behaviour occurring with a time scale of days. We believe it is reasonable to model the dynamics only of infected cells because we are interested in mutation dynamics in the within-host model, which take place over years of infection, and the dynamics of the between-

host model, which consider hundreds of years of evolution over an epidemic. This modelling decision will be discussed further when model parameters are considered.

The model considers the main target of HIV-1, activated CD4 T-cells as the uninfected target cells, which have density $x(t)$ at time t. Uninfected activated CD4+ cells join the population at rate λ, have a natural death rate d, and become infected at rate $x\sum_{i=1}^{n}\beta(i,\beta_{0,h})y_i$, where the replicative capacity of mutation class i, $\beta(i,\beta_{0,h})$, is given by the saturating exponential function:

$$\beta(i,\beta_{0,h};\beta_M,\beta_\lambda) = \beta_{0,h} + (\beta_M - \beta_{0,h})(1 - e^{-\beta_\lambda i}).$$

Thus the replicative fitness function $\beta(i,\beta_{0,h})$ is a monotonic increasing function of mutation class i with minimum value $\beta_{0,h}$; $\beta_{0,h}$ is the viral trait that will be inherited from the between-host model with index h for each host, as explained further in Section 2.3. This replicative fitness function also depends on the simulation parameters for maximum replicative fitness, β_M, and slope, β_λ. Note that we can simulate a fixed value of β by setting $\beta_\lambda = 0$.

We have chosen a saturating exponential function which has a greater fitness advantage between earlier mutation classes and then saturates with increasing mutation classes.

Thus we are modelling replicative capacity with the assumption that there exists a maximum replicative fitness that the virus could attain. Importantly, during simulations this theoretic maximum fitness is not reached when realistic parameter values are used. Beyond considering a *reductio ad absurdum* argument for the necessity of an upper limit on replicative fitness, an upper limit on replicative fitness is biologically reasonable considering that the most prevalent virus strain elicits an immune response, thus limiting maximum fitness achievable [Immonen et al., 2015].

We conjecture that the choice of a saturating exponential function is not required. As long as the final mutation class, and thus the maximum fitness modelled, is not reached any reasonable monotonically increasing function could be used to maintain the relationship between increasing mutation classes and increasing fitness. The exploration of non-monotonic fitness functions is left for future work.

HIV-1 is vulnerable to mutation upon replication; we model the probability of mutation in two steps. First, newly infected cells have a probability of mutation μ. So, newly infected cells carry the same viral traits as the infecting cell with probability $1-\mu$, or mutate to a new phenotypic mutation class with probability μ. Second, when a mutation occurs, upon replication, the mutation class the newly infected cell joins is determined by the mutation kernel

$$K(\zeta) = \begin{cases} \lambda_\mu e^{-\lambda_\mu \zeta} & \zeta > 0 \\ 0 & \zeta \leq 0. \end{cases}$$

where the constant λ_μ is a simulation parameter. Thus, cells infected by non-mutated virus join y_i at a rate $x(1-\mu)\beta(i,\beta_{0,h})y_i$ when the new infection is from a class i infected cell, and cells infected by mutated virus join y_i at a rate $x\mu\sum_{j=1}^{n}\beta(j,\beta_{0,h})K(i-j)y_j$ when the new infection is from a mutation class j cell where $j \neq i$.

Modelling cell death, infected CD4 T cells are removed with a removal rate function that is specific to their phenotypic mutation class:

$$\delta(i,\delta_M;\delta_0,\delta_\lambda) = \delta_0 + (\delta_M - \delta_0)(e^{-\delta_\lambda i}).$$

Similar to the replicative fitness function, the infected cell removal function $\delta(i,\delta_M;\delta_0,\delta_\lambda)$ is a function of mutation class i and maximum death rate δ_M, parameters which change within and between hosts during a simulated epidemic, as well as depending on simulation parameters for a minimum removal rate δ_0 and slope δ_λ.

We consider a viral reservoir of infected CD4 T cells, which has been well established, and are not modelling the possibility of non-CD4 T cell reservoirs of infectious virus despite recent evidence [Kuo and Lichterfeld, 2018]. The latent reservoir of resting CD4 T cells is assumed to have a low mutation rate because any replication of latent cells that does occur is through the DNA replication process of cellular proliferation.

The dynamics of the viral latent reservoir are not fully characterised. It is usual to describe the latent reservoir as forming early and being very long lived. But a consensus on the specifics of entry, activation, clearance, and clonal expansion of reservoir cells has not been reached [Abrahams et al., 2019, Sengupta and Siliciano, 2018].

In particular, it is believed that CD4 T-cells may be infected as they are transitioning from the active to the memory state and in this way HIV provirus is inserted into the cellular genome [Abrahams et al., 2019]. Models of within-host dynamics that include a latent reservoir often model entry into the reservoir as new infections take place. That is, entry into latent reservoir is included with the xy_i interaction terms [Doekes et al., 2017]. Given the uncertainty surrounding these dynamics, we have made the simplifying assumption of a constant rate of entry from the infected cell compartment to the latent reservoir, as well as a constant activation rate from the latent reservoir to the infected cell pool. We have also chosen to exclude a reservoir-specific removal term such that all removal of virus takes place through the infected cell compartment. Mutation class i infected cells enter the mutation class i latent reservoir (l_i) at rate γ, and are activated from l_i at rate α. Latently infected cells replicate through cellular proliferation at rate ρ. As mentioned above, the DNA replication process of cellular proliferation has a relatively low mutation rate compared with the active replication of HIV viral particles, and we therefore assume error-free viral replication through cellular proliferation [Doekes et al., 2017, Immonen and Leitner, 2014].

Taken together, these assumptions yield the following within-host model:

$$\begin{aligned}
\dot{x} &= \lambda - dx - x\sum_{i}^{n}\beta(i,\beta_{0,h})y_i \\
\dot{y}_i &= x(1-\mu)\beta(i,\beta_{0,h})y_i + x\mu\sum_{j=i}^{n}\beta(j,\beta_{0,h})K(i-j)y_j + (\alpha-\gamma)l_i - \delta(i,\delta_M)y_i \\
\dot{l}_i &= \gamma y_i + (\rho-\alpha)l_i,
\end{aligned} \tag{2.1}$$

for mutation classes $i \in \{0, 1, 2, \ldots, n-1\}$.

Given fixed values of γ and α with $\alpha > \gamma$, the parameter ρ can be used to set the relative size of the latent reservoir. In particular, we set

$$\rho = \begin{cases} s(\alpha - \gamma) & \alpha - \gamma \geq 0 \\ 0 & \alpha - \gamma < 0 \end{cases}$$

where $0 \leq s \leq 1$ is a simulation parameter which determines the size of the latent reservoir ($\sum_i l_i$) relative to the total number of actively infected cells ($\sum_i y_i$).

2.3 Between-host Dynamics

2.3.1 Full model

The full model is our most complete simulation of the between-host transmission dynamics during an HIV-1 epidemic. We develop a stochastic simulation of the transmission of HIV-1 including an extreme bottleneck, where new infections are founded by a single virion. The phenotypic mutation class of that virion, i, is used to determine the initial conditions, $\beta_{0,h}$ and δ_0, in the newly infected host, h; thus these two viral traits are carried by the founding virion to the new host.These infected hosts form a transmission tree over epidemic time where each vertex is an infected individual and the transmission events are directed edges between vertices.

In reality, the HIV-1 transmission bottleneck is multistage [Joseph et al., 2015, Fraser et al., 2014], involving virus sampled from a subset of the virus circulating within the host [Boritz et al., 2016, Gianella et al., 2016]. There is evidence that the circulating HIV-1 virus is compartmentalized and that the virus available for transmission will depend on the compartment from which the virus is transmitted. There is also evidence of preferential transmission of virus [Deymier et al., 2015, Carlson et al., 2014a, Champredon et al., 2013, Joseph et al., 2015]. An example of this preferential transmissions mechanism is the increased glycosylation of some virus which effects the ability of the virus to transit the mucosal barrier in the case of sexual

transmission [Ping et al., 2013].

After virus is transmitted to the new host it is then subject to a local fitness competition before the establishment of a systemic infection [Kariuki et al., 2017, Joseph et al., 2015, Fraser et al., 2014]. Evidence suggests that as a result of the transmission bottleneck and this local fitness competition, the majority of new infections are established by a single variant [Carlson et al., 2014b, Tully et al., 2016, Haaland et al., 2009, Li et al., 2010, Keele et al., 2008]. Thus, we assume that each new infection is established by a single virion in our model, and our simulation of the bottleneck models the distinct stages of the bottleneck as the stochastic processes described in the sections below. We address the idea of extending the model to include the transmission of multiple variants that simultaneously establish a systemic infection in Chapter 6.

In the full model, the between-host transmission of HIV-1 is separated into four distinct stages or components: (1) the probability of a *potential transmission*, which can be thought of as a contact probability; (2) given that a contact occurs, the stochastic sampling of virions to form the bottleneck inoculum; (3) the local competition that occurs between the virions in the bottleneck inoculum, determining which single virion will have the chance to establish the new infection; and (4) the probability, ϕ, that a new infection is established by that virion.

Potential transmission

A potential transmission simulates the probability of contact between an infected and an uninfected individual. This potential transmission is dependent on the contact structure [Park and Bolker, 2017] and can be thought of as a sexual contact for any sexual risk group or an instance of needle sharing.

This framework presents two challenges: the types of transmission risks individuals will face in their infected lifetime will vary[1], and the number of potential contacts for individuals of different risk groups have a large variation. Rather than combining the different contact types

[1] e.g. A single individual may have sexual contacts from different risk groups.

an individual from a particular risk group would have by forming a mean risk, we make the simplifying choice to simulate only epidemics of individuals that all belong to the same risk group. This is preferable to averaging over very different risk groups, which would obscure the interpretation of potential transmissions.

The number of potential transmissions c_j for a particular individual is Poisson distributed with mean $\mu_c t_T$, where μ_c and t_T are simulation parameters giving the mean number of potential transmission events per year, and total number of infectious years for an infected individual, respectively. Although the maximum time an individual is infectious can vary between hosts and could be reduced by a highly virulent virus or by the introduction of antiretroviral therapy, we make a strong simplifying assumption and set t_T as a constant for all individuals in the epidemic.

The times of potential transmissions, t_d, are drawn from a uniform distribution from 0 to t_T where d is an index for each potential transmission for an individual. This is equivalent to having a constant probability of contact in each time interval.

Transmission fitness ϕ

Transmission fitness, ϕ, in our simulation is the conditional probability that a single virion establishes a new infection, given that there is a potential transmission and that this virion is the "winner" of the local competition among virions in the bottleneck inoculum. Thus, the unconditional probability that a single virion establishes a new infection also depends on the size of the bottleneck inoculum and the structure of the local competition, both of which will be discussed in the following sections. A calculation of this unconditioned probability as a function of ϕ will be considered after the relevant structure of the between-host model has been detailed.

Bottleneck Inoculum

At each new potential transmission, a number of virions is generated and this ensemble of virions forms the bottleneck inoculum. To avoid bottleneck inocula that contain zero virions, the number of virions is determined as the maximum of one and a single draw from a Poisson distribution. The simulation parameter for bottleneck size sets the mean value of this Poisson distribution.

For a potential transmission event occurring at time t during an individual infection, the virions included in the bottleneck inoculum are sampled from the distribution of phenotypic mutation classes corresponding to the solution of the within-host model of the transmitting individual at that time. In other words, we use the within-host model to generate the distribution of infected cells of mutation class i at time t, and assume that this mirrors the distribution of potentially transmitted virions in each mutation class.

Calculation of transmission fitness ϕ

An important feature of our approach is that we do not assume that traits affecting transmission fitness are shielded from mutation while the virus is replicated within host. Each virion in the bottleneck inoculum is therefore assigned a transmission fitness, ϕ, but the distribution from which ϕ is chosen depends on the the mutation class of the virion. In general, virions in higher mutation classes have sustained several mutations relative to the founding viral strain, and these mutations are likely to have some effect on other traits, including transmission fitness. In addition, virions in higher mutation classes are likely to have experienced more "mutational pressure", that is, they are more likely to be part of within-host viral lineages that have been circulating within the host for long times, and have not been shielded from mutation in the latent reservoir. We argue that transmission fitness is unlikely to remain completely unaffected by this mutational pressure.

The distribution of fitness effects of new mutations on the transmission fitness of HIV-1 is not well characterized [Theys et al., 2018]. As an example, virus from chronically infected

$\phi_{0,h}$	transmission fitness of the founding virus
ϕ_{m}	transmission fitness of m^{th} virion in the bottleneck inoculum
Φ	random variable describing the change in transmission fitness, a single draw from $\mathbb{G}$
$\mathbb{G}$	The generalized flipped inverse gamma distribution

Table 2.1: Definition of terms involved in the calculation of ϕ.

individuals was found to be more highly glycosylated relative to virus circulating in acute infections. This suggests that mutations that confer a fitness advantage within-host may be deleterious for transmission fitness [Ping et al., 2013]. Generally, the distribution of fitness effects of mutations is complex and multi-modal, and single mutations can be lethal for RNA viruses [Eyre-Walker and Keightley, 2007].

We make the assumption that a mutation that is beneficial to a particular within-host trait (such as reducing the infected cell death rate) should not necessarily have correlated effects on an unrelated trait, in this case transmission fitness. Thus we expect that mutations that change the within-host fitness (and therefore mutation class) might have similar effects on transmission fitness as randomly chosen mutations. To implement this assumption the value of ϕ_m is calculated for virions $1, 2, \ldots, m$, in the bottleneck inoculum using the flipped inverse gamma distribution described below. The flipped inverse gamma distribution has been previously introduced in a number of other settings as a model of the fitness effects of a randomly chosen mutation, reflecting the fact that most random mutations are slightly deleterious [Doekes et al., 2017, Zanini et al., 2017, Eyre-Walker and Keightley, 2007, Eyre-Walker et al., 2006].

Thus, the transmission fitness for the m^{th} virion in the bottleneck inoculum, ϕ_m, is given by

$$\phi_m = \phi_{0,h} + (i - j)\Phi$$

where j is the mutation class of the virion of interest in the bottleneck inoculum, $\phi_{0,h}$ is the transmission fitness of the mutation class i virion, from host $h - 1$, that founded the within-host infection, and Φ is a random variable drawn from the generalized flipped inverse gamma

distribution:

$$\mathbb{G}(x) = \begin{cases} \exp\left(\frac{\bar{\beta}}{x-\bar{\mu}}\bar{\gamma}\right)\bar{\gamma}\frac{\bar{\beta}}{x-\bar{\mu}}^{\bar{\alpha}\bar{\gamma}+1} & x > \bar{\mu} \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

which depends on the simulation shape parameters $\bar{\alpha}, \bar{\phi}$ and simulation scale parameter $\bar{\gamma}$, and location parameter $\bar{\mu}$.

We note that a greater difference in mutation class, $i - j$, will, on average, increase the change in transmission fitness between the founding and transmitting virion. If virion m in the bottleneck inoculum has not mutated, $i = j$ and ϕ_m will be equal to $\phi_{0,h}$. See Figure 2.1 for example probability density functions of the change in transmission fitness, $(i - j)\Phi$, for changes in mutation class $i - j = 1$, 2, and 3.

Establishment viral competition

Once the bottleneck inoculum is transmitted to the new host it is then subject to a local fitness competition before the most fit variant establishes a systemic infection. We model this local establishment competition by selecting the virion in the bottleneck inoculum with the highest ϕ. The ϕ for this virion is then tested against a random number to determine if an infection is ultimately established. If a new infection is established the mutation class from the successful virion becomes the initial mutation class of the new infection, and the values of β, δ and ϕ associated with this virion in the donor become $\beta_{0,h}$, δ_0 and $\phi_{0,h}$ respectively in the recipient.

Sampling the transmission trees

Because transmission trees are initiated with a single infected individual, there is a high probability that the epidemic will go extinct in the first few steps. In cases where successful transmission is established (non-extinction), if the between-host fitness is sufficiently high the transmission tree undergoes a period of exponential growth, which typically slows as the virus evolves and ϕ decreases. For computational efficiency we do not record and simulate every infection

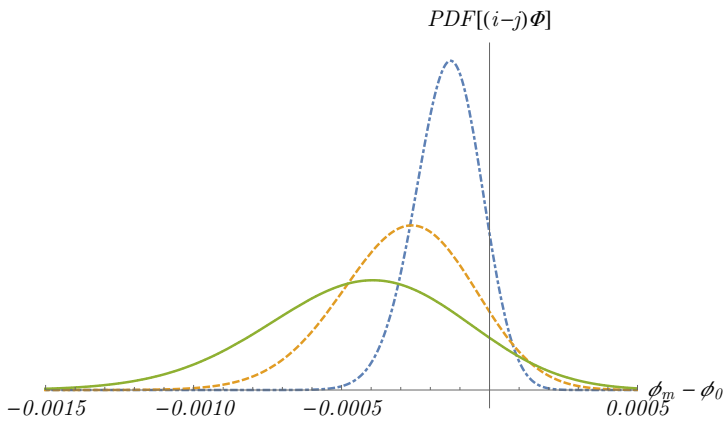


Figure 2.1: Example probability density functions (pdfs) for change in transmission fitness, $(i - j)\Phi$, for mutation classes with a change of $i - j = 1$ (dotdash), 2 (dashed), and 3 (solid) which show the increase in the variance of the distribution without changing the probability that a mutation is beneficial. The plots are generated from the trial *phiDistTest* experiment 1 and the parameter regime is available in Appendix A.1.

during the exponential growth of the epidemic. Instead, we sample lineages such that although the underlying epidemic would be expanding exponentially, we record, on average, only one new infected individual per infected individual. This sampling rate begins near one (all infections in the lineage are recorded at the start of the epidemic) and is adjusted by considering the number of new infections generated by the previous $n \leq 25$ infected individuals.

2.3.2 Fixed transmission model

The fixed transmission model is used as an experimental control. In this model, within-host mutation is uncoupled from the ability of the virus to transmit. In particular, the mean transmission fitness of the founding virus is attributed to all the virions in the bottleneck inoculum regardless of their phenotypic mutation class or class that established the infection in the transmitting host. In this case, a single virion is chosen at random from the bottleneck inoculum to establish the new infection.

Holding transmission fitness constant across the epidemic removes any transmission fitness effects, and within-host fitness is allowed to evolve without between-host transmission fitness constraints. That is, the virus is transmitted through a restrictive but random bottleneck.

2.3.3 Simplified model

In the model we will describe as the "simplified model", the complex transmission dynamics are replaced by a single chain of deterministic transmissions. Each infected individual generates exactly one new infection, the time of which is drawn uniformly from the donor's infectious lifetime. At the time of transmission a single virion is randomly sampled from the within-host model to establish the new infection. The values of β and δ associated with this virion in the donor become $\beta_{0,h}$ and δ_0 respectively in the recipient.

2.4 Experimental design

Each model simulation, containing both the within- and between-host components, is initiated with a single infected individual and generates one transmission tree. The number of transmissions and time span of this generated transmission tree will vary for each simulation. See Figure 2.2 for an example of a transmission tree. In the transmission trees, vertices represent infections and the edges represent transmission events.

Each run the simulation is repeated many times; a single repetition of the simulation is referred to as a *replicate*. For a given set of simulation parameters and initial conditions each replicate is independent and produces a single transmission tree. It is possible that this transmission tree contains only the initial vertex.

Each set of simulation parameters and initial conditions are referred to as an *experiment*. Each replicate is initiated from an experiment. One completed simulation experiment will contain many replicates. Experiments with different parameters and initial conditions are collected in *trials*. Each experiment in a trial differs in a single simulation parameter. Trial definitions can be found in Appendix A.1.

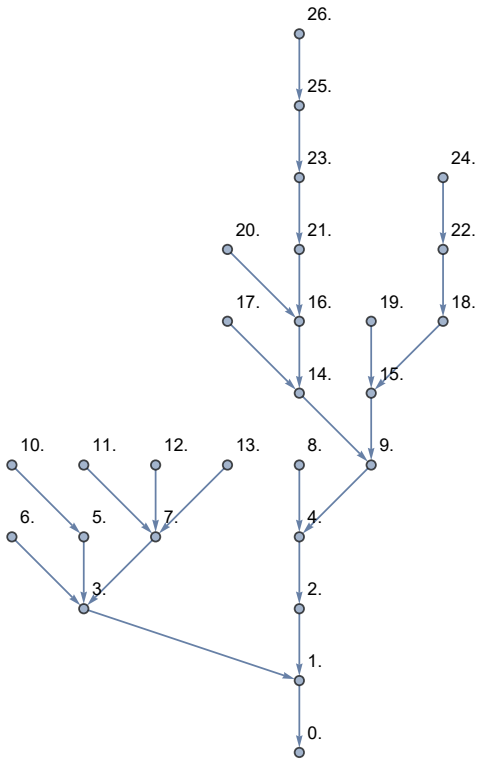


Figure 2.2: Example of a short transmission tree generated from one replicate. The tree is not time scaled. Numbers provide an index of infected individuals; the epidemic began with patient 0. Arrows point from the recipient toward the donor to illustrate that lineages are generated from the *leaves* to the *root* as discussed in Section 2.5.1 and that the ancestor node is used to generate the tree summary statistic, 2.5.2. An example lineage is the transmission chain through {10,5,3,1,0}.

2.5 Analysis

The result of each experiment is a collection of highly variable transmission trees. As the transmission trees evolve some lineages grow exponentially while others go extinct. Simulated transmission trees, started from a single vertex, can go extinct immediately or contain thousands of vertices. It is common for entire branches of the transmission tree to go extinct as the result of stochastic variation or as the result of mutations that may lower transmission fitness unsustainably. Despite a large number of transmissions over a large number of trees, significant variation remains. The stochasticity of the evolution and size of transmission trees presents challenges for analysis.

2.5.1 Random Lineage Summary Statistics

A single random lineage from each replicate is used to generate summary statistics. For each simulated transmission tree one vertex without descendants, a *leaf*, is chosen randomly and only the transmissions that belong to the path from this leaf to the initial vertex, the *root* are used in generating the summary statistics. See Figure 2.2 for an example lineage.

The use of only one lineage per replicate addresses the problem of repeated counting of transmissions near the root that would occur if all lineages in each tree were used. Repeated counting is a problem since as lineages approach the root each node is a member of more lineages and at the extreme we see that every lineage includes the root. The use of only one lineage per replicate, however, discards much potentially useful information and requires many replicates to be generated since the *lineage space* in each tree can be large.

The selection of a random lineage also introduces bias into the sample, as successful lineages will have had more transmissions and are therefore more likely to be sampled. This sampling bias is toward more successful viral strains; surviving lineages are over-represented, but these are not necessarily more evolved lineages.

2.5.2 Tree Summary Statistics

To generate tree summary statistics, the root is positioned at the centre of the tree and the total epidemic length is divided into concentric bins; for each graph edge that crosses the concentric bin line, the ancestor node is determined. The parameters recorded for this ancestor node are the initial conditions for the within-host model that established the ancestor's node's infection; it is these parameters that are used to calculate the summary statistics. Since the number of transmissions used to generate this summary statistic differ across replicates, a weighted average accounting for this difference must be considered when combining the results from multiple replicates.

The use of the ancestor node's parameters is a required simplification that does not account for within-host evolution that has occurred between the time the ancestor node was infected and the time at which the graph edge was sampled. This simplification is required since the establishment of a new infection is a stochastic process and the evolution at the time of the sample, when the concentric bin line crossed the graph edge, does not predict the parameters of the virion that was ultimately transmitted.

Summary statistics determined in this way have the advantage that nodes that result in multiple transmissions will be counted multiple times; this advantage is distinct from the disadvantage of counting the statistics from every lineage discussed in the preceding section, 2.5.1.

In contrast, summary statistics determined in this way have the disadvantage of averaging all transmissions, such that transmissions that result in extinct lineages are included along with the surviving lineages. This can result in the calculated mean value "jumping" in value when a lineage, or branch, does go extinct. For this reason, the choice of bin width can affect the resultant statistic. Finally, statistics determined in this way never consider the leaves, since only parameters from ancestor nodes are recorded.

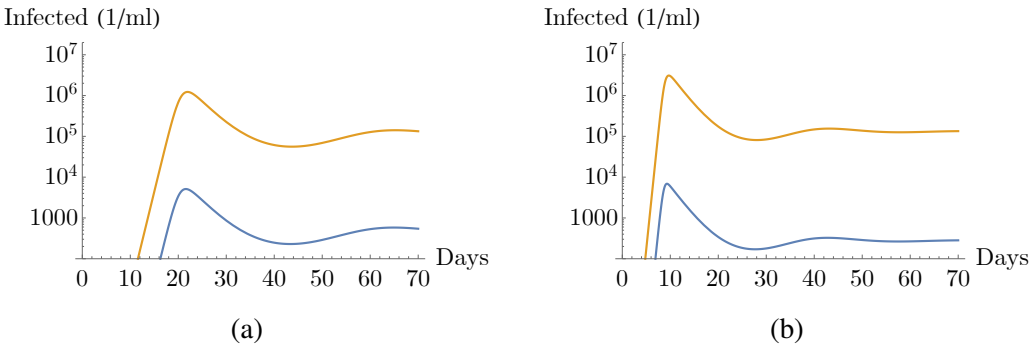


Figure 2.3: Panel (a) shows the solution to a within-host model using the best-fit parameters for patient seven from an early HIV study from Stafford et al. Panel (b) displays the solution to the same within-host model using the average estimated parameter values from the same study [Stafford et al., 2000]. The yellow line represents the viral load and the blue line represents the number of infected cells.

2.6 Model Parameters

A major drawback of a complex simulation model is the number of model parameters and the model's sensitivity to those parameters. Choosing the default parameters for our model we faced several challenges: individuals infected with HIV-1 present large variability in viral trajectories; choosing ensemble parameters from various studies and individuals can lead to unrealistic within-host viral trajectories which do not share the qualitative features of the individual infection time courses that were used to estimate these ensemble parameters; and within-host models in the literature are variable and are largely purpose-specific, so the translation of parameters between different models is required.

We have chosen to draw our initial simulation parameters from as few study participants and studies as possible. For our within-host parameters we use best-fit parameters for the participant that most closely matched the average parameters in [Stafford et al., 2000]; a solution to a within-host model using these best-fit parameters for participant seven, or using the average estimated parameter values from that study, are compared in Figure 2.3.

For each simulation trial the parameters shared by all experiments are fixed, while the parameters of interest for that trial are varied in each experiment. For full details of the parameters used for the trials considered in Chapter 3, see the full trial definitions in A.1.

parameter	description	units	value	reference
λ	Growth rate, uninfected immune cells	cells·(ml·day)$^{-1}$	$x_0^2(5.7 \times 10^{-5})$	[Stafford et al., 2000]
$\mathbb{M}(\beta(i;\beta_{0,h},\beta_M))$	mean of the β kernel	ml·(cells·day)$^{-1}$	0.0741	[Stafford et al., 2000]
$\mathbb{M}(\delta(i;\delta_0,\delta_M))$	mean of the δ kernel	ml·(cells·day)$^{-1}$	0.0741	[Stafford et al., 2000]
d	death rate, uninfected immune cells	day^{-1}	5.7^{-5}	[Stafford et al., 2000]
$\mathbb{M}[K(\zeta)]$	mean of the mutation kernel		5×10^{-5}	[Lythgoe et al., 2013]

Table 2.2: Shared within-host simulation parameters for most experiments. For full details for each experiment, see appendix A.1.

Chapter 3

Multi-scale Model: Results

3.1 Within-host Model

The typical time course of a single within-host infection, using parameters from the trial *Primary*, experiment 1, is shown in Figure 3.1. Here, the infection begins with a single infected cell in mutation class 0, and progresses for five years. For visual clarity, only actively infected cell classes are shown (colours), the sum of the actively infected cell densities is shown in black. Vertical bars illustrate the times at which a transmission event occurred, causing an infection in a new host.

Figure 3.1 illustrated only the actively infected cells, y_i. The density of infected cells in the latent reservoir is determined by the entrance rate into the latent compartment γ, activation rate from the latent compartment α, and the rate of cellular proliferation of the immune cells in the latent compartment ρ. In our simulation, the parameter ρ is used to maintain the relative size of the latent reservoir such that the total density of latently infected cells is equal to the density of actively infected cells. Figure 3.2 shows a comparison of the size of the latent reservoir with the actively infected compartment, using the cellular proliferation rate ρ to maintain equality. The results shown are using parameters from *Grade* experiments 2, 5, 8, 11 and 14. Trial definition details can be found in Appendix A.1.

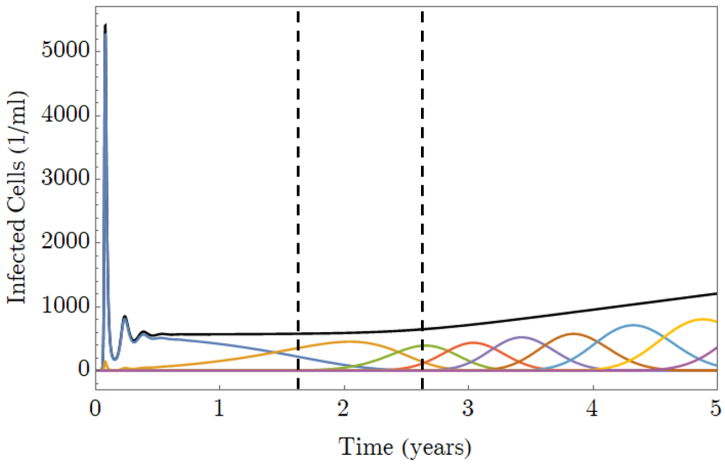


Figure 3.1: The typical time course of a single within-host infection, using parameters from the trial *Primary*, experiment 1. Actively infected cell classes are shown (colours), the sum of the actively infected cell densities is shown in black. Vertical bars illustrate the times at which a transmission event occurred, causing an infection in a new host.

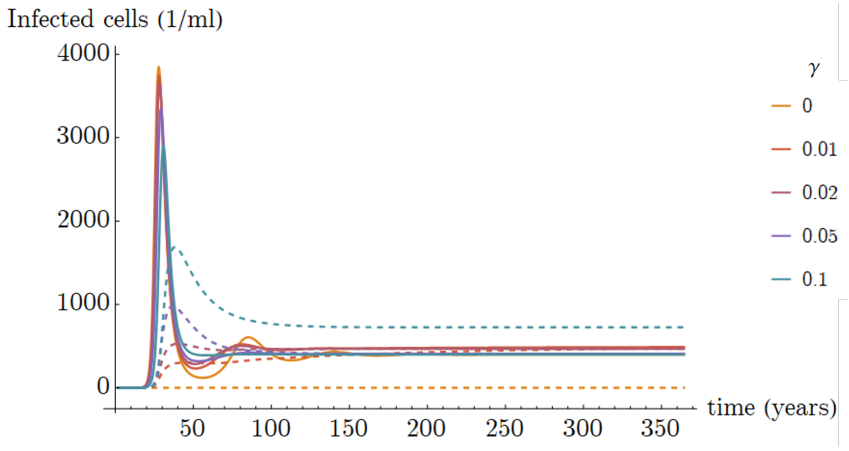


Figure 3.2: Comparison of the relative density of infected cells in the latent reservoir with variation of entry rate γ and cellular proliferation rate ρ set to maintain equilibrium. The results shown are using parameters from *Grade* experiments 2, 5, 8, 11 and 14. Trial definition details can be found in Appendix A.1. The solid lines show the actively infected cells and the dashed lines the latently infected cells. In experiment 14 with $\gamma = 0.1$, the dashed turquoise line, is above the actively infected cells because even with $\rho = 0$ the number of latent cells at equilibrium is higher than the number of infected cells at equilibrium. The dashed orange line from experiment 2, has $\gamma = 0$ and the latent reservoir remains at zero.

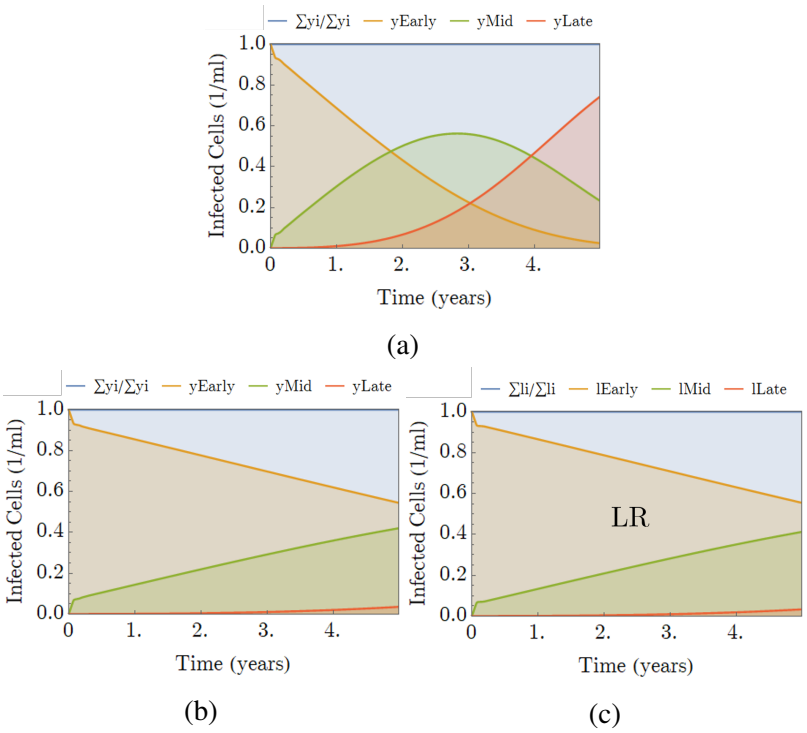


Figure 3.3: Solutions of the within-host model showing the relative frequency of phenotypic mutation classes, with mutation classes 1 through 4 shown as early mutation classes, 5 through 8 as middle mutation classes, and classes greater than 8 combined as the late mutation classes. The top panel, (a), contains a solution to the initial within-host model for the relative frequency of infected cell mutation classes from Experiment 271 in the trial *Blueberry* where $\gamma = 0$. The left panel, (b), contains the relative frequency of infected cell mutation classes y_i and the right panel, (c), the relative frequency of the latent reservoir l_i mutation classes from Experiment 271 in the trial *Blueberry* where $\gamma = 0.02, \alpha = 0.045$ and $\rho = 0.035$. The trial definition is available in Appendix A.1.

An example of the progression through mutation classes during the course of a single within-host infection is shown in Figure 3.3, both with and without the latent reservoir. To better visualize this progression, relative cell densities are plotted, with mutation classes 1 through 4 shown as early mutation classes, 5 through 8 as middle mutation classes, and classes greater than 8 combined as the late mutation classes. In panel (a) there is no latent reservoir, whereas in panels (b) and (c), with a γ value of 0.02, a ρ value of 0.035 and an α value of 0.045, the latent reservoir compartment is equal in size to the actively infected compartment and the mutation classes in the latent reservoir lag very slightly behind the mutations in the actively infected compartment.

3.2 Multi-scale Model

We are interested in the evolution of both transmission fitness and within-host fitness. As explained previously, within-host fitness increases through evolutionary changes in δ, the death rate of infected cells.

To facilitate comparisons, simulations are organized into trials, experiments, and replicates. The analysis in this chapter focuses on experiments from the *Primary* trial and all experiments are from this trial unless otherwise noted. Parameters are varied between experiments to allow the comparison of effects, see 3.1. Each experiment is then repeated in a number of replicates, with the same parameter values, where each replicate generates a transmission tree.

The mean value of δ calculated from each transmission tree, which is generated in each replicate, is used to quantify changes in δ for each experiment. The time series of each mean value of δ generated from an experiment over the epidemic is referred to as a trajectory. Figure 3.4 shows the resulting trajectories from the transmission trees that survived until the end of the simulation at 200 years. Each transmission tree is summarized into a mean trajectory using the tree summary method and the distributions of trajectories at 200 years, the final time in the trial *Primary*, are used to compare experiments. The definition of the experimental parameters for

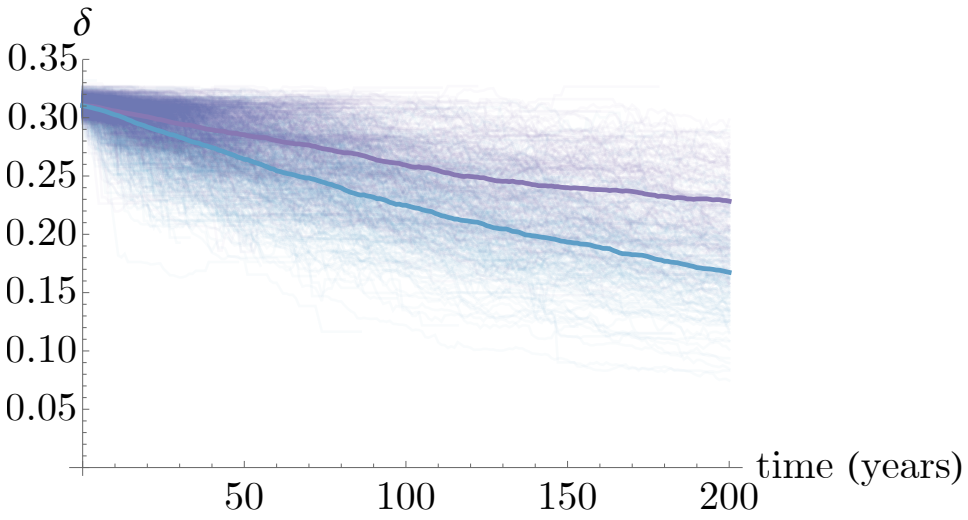
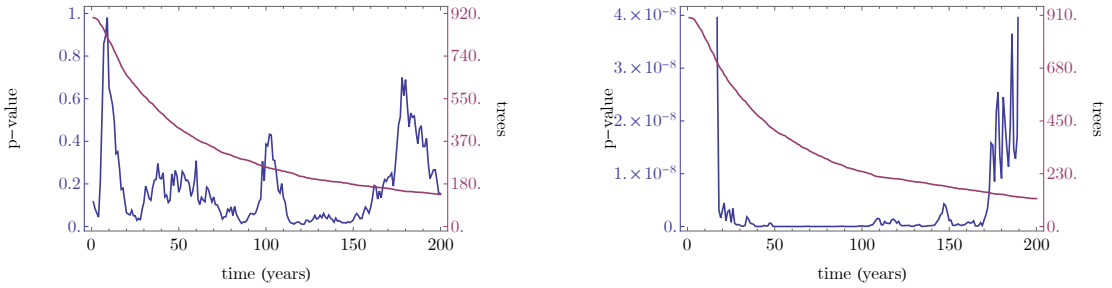


Figure 3.4: Each thin line represents the mean δ value for one replicate calculated with the tree summary method, that is, the mean δ value averaged over a single transmission tree from experiment 5 (thin purple lines) or 7 (thin blue lines). The thick lines represent the mean of all the replicates for experiments 5 (purple) and 7 (blue).

the *Primary* trial are found in Table 3.1 and the full definition of the trial is found in Appendix A.

Figure 3.4 shows the cloud of individual evolutionary trajectories that result from the tree summary statistics for each replicate. Each replicate of 1000 for the trial *Primary* was then summed to create the trial average (thick lines in the figure). In the sections to follow, we compare the distributions of averaged tree statistics to determine significance across trials, comparing mean tree statistics over time and also at the final point of the simulation.

The distribution of parameter averages across replicates vary with time; all trees in an experiment will start at the same point and then possibly diverge as the epidemic progresses over time. But the statistical power of this divergence is balanced by the loss of replicates as transmission trees go extinct. In Figure 3.5 we see that for longer final times, fewer transmission trees are available (purple lines/axes), and that the p-value for significance between two experiments fluctuates, depending on the final time chosen. Nonetheless, as shown in these typical examples, significant results remain very significant over the time course of the epidemic (panel b), while insignificant results vary greatly as transmission trees go extinct but



(a) (b)

Figure 3.5: P-value for the difference in sample medians, compared pairwise over the length of the epidemic, for experiments $3, 4$ in panel (a) and $9, 13$ in panel (b) from a two-tailed Mann-Whitney test.

never reach a level of significance (panel a). We have chosen to look at the distribution of transmission trees at the end of the simulation to allow the rate of extinctions over time to slow.

The effects of transmission fitness on the evolution of δ are included in the simulation through, ϕ, transmission fitness. To analyze the impact of including transmission fitness in the simulation a control case for each parameter set, the *fixed model*, is also simulated. Recall, from the previous chapter, that transmission fitness effects are included in the *full model* of the simulation by linking the evolution of ϕ with the number of within-host mutation classes that a virion in the transmission inoculum has passed through. In the *fixed model* this linkage is broken and the distribution that ϕ is drawn from is fixed; in the trial *Primary* when a *fixed model* is simulated that distribution is fixed to the distribution that arises when mutation class 0 is transmitted to the next infected host. This results in limited random evolution of ϕ independent from the the evolution of δ. In each experiment which model is simulated, *the full model* vs. the fixed model is recorded in the control parameter. e.g. in the *Primary* trial, experiment 1 is simulated under the *full model* and experiment 2 has the same parameters as experiment 1, but is simulated under the *fixed model*.

The inclusion of between-host fitness effects, the full model, prevents δ from evolving as quickly as it does when uncoupled from the evolution ϕ. The results of the trial *Primary* show this key result in Figure 3.6. Here, we can compare δ evolution between experiments

#	experiment	control	μ(BN)	latent	γ	λ_δ
1	Primary1	Full	1	–	0	0.0003
2	Primary2	Full	1	LR	0.02	0.0003
3	Primary3	Fixed	1	–	0	0.0003
4	Primary4	Fixed	1	LR	0.02	0.0003
5	Primary5	Full	2	–	0	0.0003
6	Primary6	Full	2	LR	0.02	0.0003
7	Primary7	Fixed	2	–	0	0.0003
8	Primary8	Fixed	2	LR	0.02	0.0003
9	Primary9	Full	1	–	0	0.003
10	Primary10	Full	1	LR	0.02	0.003
11	Primary11	Fixed	1	–	0	0.003
12	Primary12	Fixed	1	LR	0.02	0.003

Table 3.1: Experimental parameter definitions for the *Primary* trial. The control parameter indicates if the *full model* or *fixed model* is simulated. Column headings refer to: $\mu(BN)$ the bottleneck inoculum size; latent and γ both indicate if a latent reservoir compartment was simulated; λ_δ is the parameter determining the shape of the fitness kernel.

1 and 3 (full versus fixed model, without latent reservoir) and between experiments 2 and 4 (full versus fixed model, with the latent reservoir); both comparisons yield highly significant differences in the final distribution of δ (see Table 3.2). We also see that the mean transmitted mutation class is reduced with the inclusion of the linked random mutation effects on ϕ (the full model), as seen in Figure 3.8a and Table 3.2. Both of these results match our intuition that allowing transmission fitness to evolve without selection during within-host evolution increases the probability of transmitting a less evolved phenotype, which will slow the between-host evolution of δ.

In both the Full and Fixed models there is not a significant difference between final distributions of δ for experiments with a latent reservoir, $2, 4$, and without a latent reservoir, $1, 3$. There is apparent separation of the mean evolutionary paths in both cases, and the differences are stronger across the epidemic and in the final distribution of δ in the Full model, between experiments $1, 2$. But the final δ distributions in both comparisons between $1, 2$ and $3, 4$ were not sufficiently different to reach statistical significance, see Table 3.2. The effect of the latent reservoir is, however, strong enough to cause a significant reduction in the average transmitted mutation class with the inclusion of the latent reservoir. This agrees with the understanding

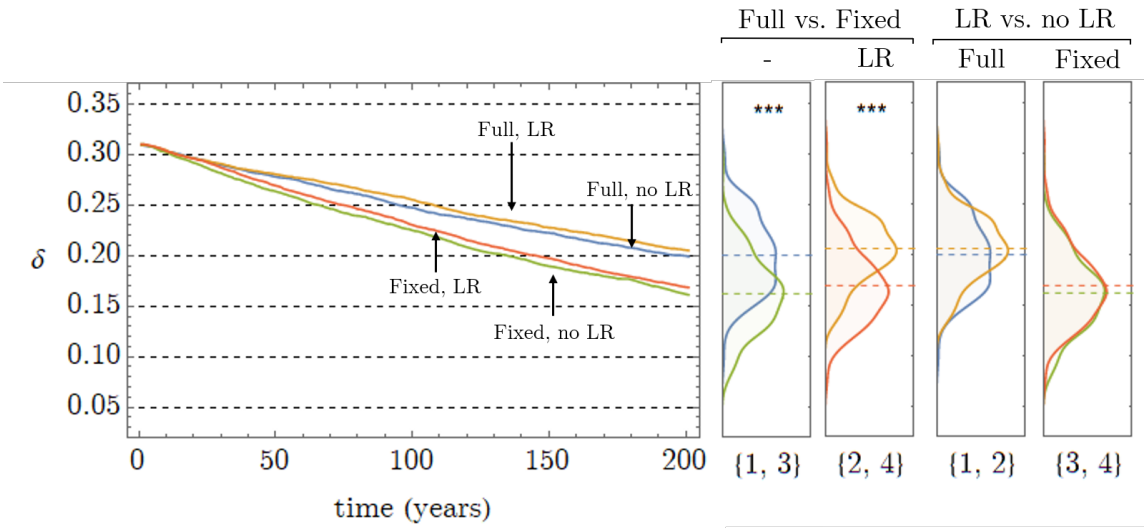


Figure 3.6: Effect of the latent reservoir, γ, and the effect of mutation class dependent selection of transmission fitness, ϕ, on the evolution of δ. The left-hand side shows the evolution of δ over 200 years of the simulation from experiments $1, 2, 3$, and 4. The top two lines in the left panel are the simulation results for experiments 1 (blue), 2 (yellow), under the full model, and the lower two lines show results from experiments 3 (green), and 4 (red), under the fixed model. The first two distributions on the right compare the fixed and full models, i.e. effect of mutation class dependent selection of ϕ, and the third and fourth distributions compare the effect of the HIV-1 latent reservoir. Refer to Table 3.2 for comparison details. Significance at the 1% (***), 5% (**), 10% (*) levels is the same as in Table 3.2.

that with the inclusion of a latent reservoir, some virus is shielded from mutation, and archived, which slows within-host evolution and that this effect would be amplified by the inclusion of a fitness cost of mutation. The marginal amplification of the latent reservoir effect can be seen in the differences between the average transmitted mutation class in figure 3.8a which shows a significant difference between experiments 1 and 2, experiments 3 and 4, and all comparisons between the Full and Fixed models, see Table 3.2. Although amplified by the inclusion of a fitness cost of within-host mutation the effect of the latent reservoir is small in comparison with direct effects of mutations that are deleterious to transmission.

We also considered the effect of changes to the size of the transmission bottleneck, as described in detail in Chapter 2. In the full model, Figure 3.7, we see that the evolution of δ is slowed by an increase in bottleneck inoculum size. Changes to the bottleneck inoculum size in *ceteris paribus* comparisons have a significant difference in cases with the latent reservoir, $1, 5$

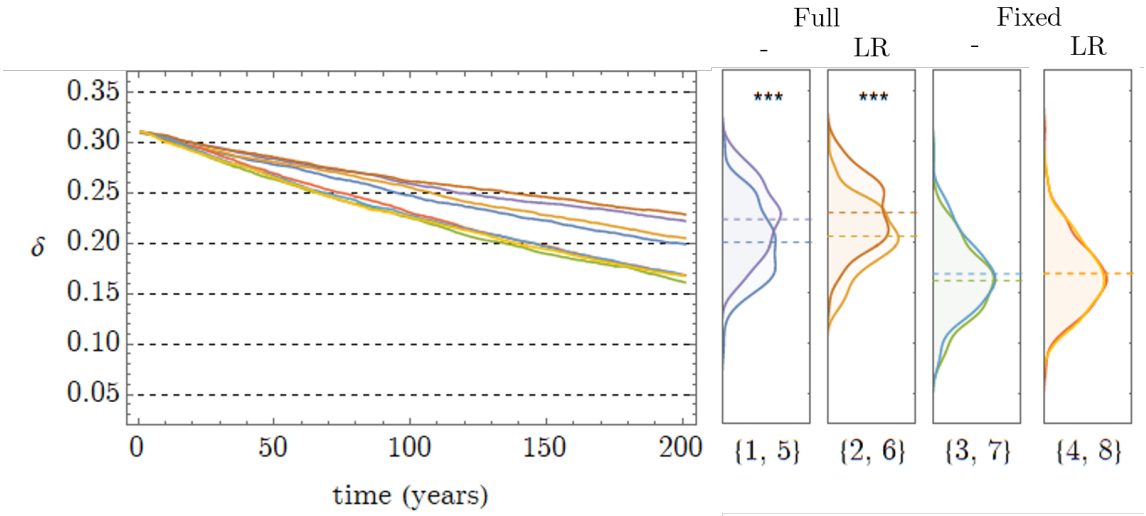
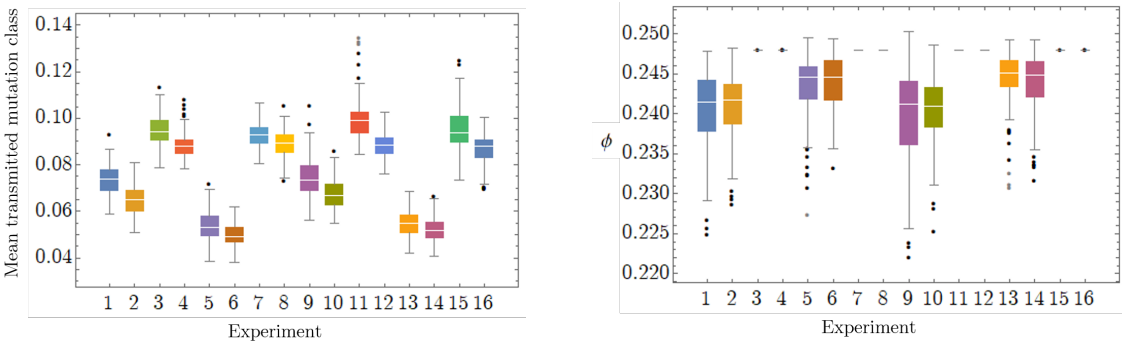


Figure 3.7: Effect of bottleneck inoculum on virus evolution. The left-hand side shows the evolution of the mean of δ over all replicates for 200 years of the simulation from experiments $1, 2, 3, 4, 5, 6, 7$, and 8. The top two lines in the left panel are the simulation results for the full model, bottleneck inoculum 2, with and without the latent reservoir (experiments 5 (purple), 6 (brown)), the middle two lines show the same results for a bottleneck inoculum of 1 (experiments 1 (blue), 2 (yellow)). We see that a larger transmission bottleneck slows within-host evolution. The bottom group of four lines show results from the fixed model, in which the bottleneck size has no effect; results from experiment 3 (green), and 4 (red) with a bottleneck inoculum of 1 and 7 (light blue), 8 (light yellow) with a bottleneck inoculum of 2 are shown. Refer to table 3.2 for comparison details. Significance at the 1% (***), 5% (**), 10% (*) levels is the same as in Table 3.2.

and without the latent reservoir $2, 6$. This accords with our intuition that when the bottleneck inoculum is larger, there is a higher chance of sampling a virion with a low mutation class and correspondingly higher transmission fitness. Since the virion with the maximum ϕ wins the local fitness competition and can potentially transmit, larger bottlenecks slow the rate of δ evolution in the full model.

Differences between experiments with different mean bottleneck inoculum sizes are not seen when the fixed model is simulated, comparisons between experiments 3 and 7 or experiments 4 and 8. This agrees with our expectation that the within-host mutation does not affect the distribution from which the transmitted ϕ is drawn in the fixed model.

The bottleneck size does not affect our previous conclusions about the presence or absence of a latent reservoir: the p-value comparing experiments with and without a latent reservoir is



(a) (b)

Figure 3.8: Mutation class and ϕ results for experiments from *Primary*. Panel (a) shows the mean transmitted mutation class averaged over the epidemic and panel (b) shows mean ϕ at 200 years. For analysis of the significance of the differences between select experiments see Table 3.3.

0.17 with mean bottleneck inoculum size of 1 (experiments 1 and 2), and is 0.16 with a mean inoculum size of 2 (experiments 5 and 6).

We have tested an extreme case for the size of the bottleneck inoculum; the inoculum is possibly orders of magnitude larger than one or two virions, but the effect of the distribution of the transmitted ϕ is largest with this small size and quickly diminishes. Simulations of the maximum ϕ value obtained for a given bottleneck size reveal that a doubling of the bottleneck inoculum from 1 to 2 virions has a much more pronounced effect on the distribution of the transmitted value of ϕ than a doubling from 5 to 10.

3.3 Sensitivity Analysis

To analyze the sensitivity of the simulation we considered 10% changes to the following parameters: the latent reservoir entry rate, latent reservoir activation rate, the shape of the mutation kernel, the shape of the fitness kernel, the initial transmission fitness, and the size of the latent reservoir, see Table 3.4. These parameters were chosen as those which had the most uncertainty in their quantitative values. We found no significant changes to the final median value of δ except a single case, in which a 10% increase to α resulted in a change with a p-value of

comparison #		LR	BN	control	δ sig.	mutation sig.	ϕ sig.
LR vs no LR	1, 2	-	1	Full	-	***	-
Full vs Fixed	1, 3	LR	1	-	***	***	***
LR vs no LR	3, 4	-	1	Fixed	-	***	-
Full vs Fixed	2, 4	no LR	1	-	***	***	***
LR vs no LR	5, 6	-	2	Full	-	***	-
Full vs Fixed	5, 7	LR	2	-	***	***	***
LR vs no LR	7, 8	-	2	Full	-	***	-
Full vs Fixed	6, 8	no LR	2	-	***	***	***
BN 1 vs BN 2	1, 5	LR	-	Full	***	***	***
BN 1 vs BN 2	2, 6	LR	-	Full	***	***	***
BN 1 vs BN 2	3, 7	no LR	-	Full	-	***	-
BN 1 vs BN 2	4, 8	no LR	-	Full	-	-	-

Table 3.2: Results where the compared experiments are given in the second column. See Table 3.1 for the experiment definitions. Significance includes the Bonferroni correction for the 12 simultaneous tests and is given for difference in sample medians determined from a two-tailed Mann-Whitney test at 1% (***), 5% (**), 10% (*) significance level. See results Table 3.3 for p-values.

0.046. Therefore our results are quite robust to changes in this set of parameters. We further considered qualitative changes to the shape of the final distribution of δ, see Figure 3.9.

Changing the entry rate γ while holding the size of the latent reservoir fixed yielded a more complex picture; while 10% changes in γ do not have a significant effect in the final median value of δ, both the larger and smaller values of γ produced multi-modal distributions of δ, where the central peak of the distribution has much less variance and there are two peaks on each side flanking the central peak. We suggest that this effect is possibly caused by transmission trees that have large components that have yet to mutate (the higher sub-peak), or have yet to go extinct (the lower sub-peak). Simulating the epidemic for longer would presumably result in a smoother δ distribution with less variance, but would be computationally prohibitive.

3.3.1 Conclusions

With the inclusion of random mutations affecting transmission, tied to the *time under mutation* pressure (i.e. the number of mutation classes a virion has transited), evolution of within-host fitness is slowed over the epidemic. Also, the evolution of transmission fitness within-host has

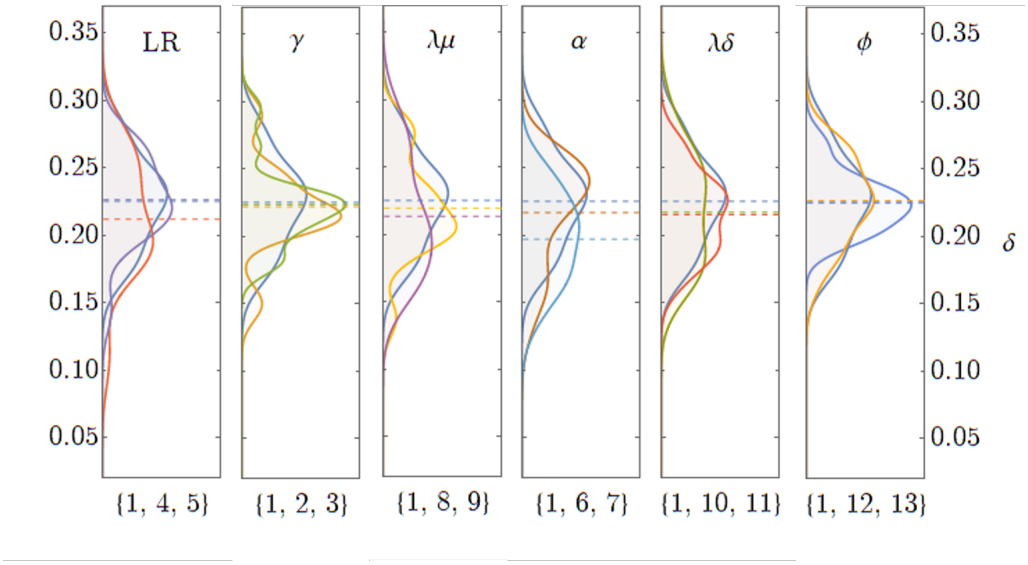


Figure 3.9: From left to right is the simulations sensitivity to: latent reservoir size, latent reservoir size entry $\Delta\gamma$, shape of the mutation kernel λ_μ, latent reservoir activation $\Delta\alpha$, shape of the fitness kernel, $\Delta\lambda_\delta$, and initial transmission fitness $\Delta\phi$, comparing 10% changes to the parameters to the effects those changes have on the final δ distributions. For a table of results see 3.4

a greater impact on the slowing of the evolution of HIV-1 between-hosts than the impact from the inclusion of a latent reservoir. Although we do not see significant changes in the evolution of δ with the inclusion of a latent reservoir, we do see a change in the average mutation class of transmitted virions, Figure 3.8a. There is an increase in the probability that a less evolved virion will be selected for the possibility of transmission.

We used conservative assumptions for the shape and strength of the mutation kernel, assuming that mutations affecting transmission fitness are random and only correlated with mutations affecting within-host fitness through shared time under mutational pressure. If we instead assumed that beneficial mutations to within-host fitness necessarily result in detrimental changes to between-host fitness, we would likely see an increase in the separation between experiments with and without a latent reservoir.

The simulation is computationally expensive and produces a large number of transmissions; within each experiment, transmission trees were significantly variable. This variability remained even after 1000 replicates. Producing reliable summary statistics proved challenging

and controlling the exponential growth of the transmission trees introduced uncertainty into the results. Given the number of assumptions required to build the model and the number of parameters and modelling choices, we have confidence in the direction and significance of changes to the final distributions of δ, but do not have confidence in the quantitative magnitude of these changes. In particular, the choice of the minimum attainable δ, controlled by the fitness kernel, determined the magnitude of changes we see to δ. The minimum attainable δ, which we set equal to the clearance rate of healthy CD4 cells, d, was never reached in simulation but did determine the magnitude of both within- and between-host evolution observed.

Considering relative changes to δ as a measure of within-host fitness change, we find that with the inclusion of random within-host mutations affecting transmission fitness, an HIV-1 latent reservoir is not required and see a reduction in the evolution of within-host fitness over the epidemic.

					δ		avg(transmitted mutation class)		ϕ	
comparison #		LR	BN	control	p-value	sig.	p-value	sig.	p-value	sig.
LR vs no LR	1, 2	-	1	Full	0.1742	-	6.0080×10^{-22}	***	0.8822	-
Full vs Fixed	1, 3	LR	1	-	8.7170×10^{-14}	***	3.3420×10^{-59}	***	8.2010×10^{-46}	***
LR vs no LR	3, 4	-	1	Fixed	0.1546	-	1.5930×10^{-18}	***	0.5715	-
Full vs Fixed	2, 4	no LR	1	-	7.1180×10^{-14}	***	8.1100×10^{-60}	***	2.5010×10^{-38}	***
LR vs no LR	5, 6	-	2	Full	0.1633	-	5.5140×10^{-8}	***	0.4510	-
Full vs Fixed	5, 7	LR	2	-	1.2980×10^{-23}	***	6.5680×10^{-60}	***	2.4390×10^{-36}	***
LR vs no LR	7, 8	-	2	Full	0.9042	-	1.8050×10^{-7}	***	0.4957	-
Full vs Fixed	6, 8	no LR	2	-	1.1610×10^{-27}	***	1.6850×10^{-63}	***	1.7510×10^{-35}	***
BN 1 vs BN 2	1, 5	LR	-	Full	1.5000×10^{-6}	***	5.1070×10^{-55}	***	6.5440×10^{-8}	***
BN 1 vs BN 2	2, 6	LR	-	Full	4.5720×10^{-7}	***	6.3910×10^{-49}	***	1.5170×10^{-9}	***
BN 1 vs BN 2	3, 7	no LR	-	Full	0.1588	-	0.0005	*	0.5525	-
BN 1 vs BN 2	4, 8	no LR	-	Full	0.9129	-	0.8164	-	0.6000	-
LR vs no LR	9, 10	-	1	Full	0.3502	-	3.1570×10^{-9}	***	0.8187	-
Full vs Fixed	9, 11	LR	1	-	5.1970×10^{-15}	***	9.0360×10^{-52}	***	9.9510×10^{-35}	***
LR vs no LR	11, 12	-	1	Full	0.0756	-	2.1390×10^{-25}	***	0.6638	-
Full vs Fixed	10, 12	no LR	1	-	6.0210×10^{-20}	***	2.5720×10^{-52}	***	4.7820×10^{-49}	***
LR vs no LR	13, 14	-	2	Full	0.3834	-	0.0011	*	0.2424	-
Full vs Fixed	13, 15	-	2	-	1.1580×10^{-26}	***	6.9480×10^{-60}	***	6.4690×10^{-32}	***
LR vs no LR	15, 16	-	2	Full	0.2766	-	2.1250×10^{-17}	***	0.2709	-
Full vs Fixed	14, 16	-	2	-	3.3850×10^{-25}	***	2.2290×10^{-62}	***	3.9450×10^{-36}	***
BN 1 vs BN 2	9, 13	LR	-	Full	2.8010×10^{-6}	***	2.0270×10^{-51}	***	1.9480×10^{-13}	***
BN 1 vs BN 2	10, 14	LR	-	Full	0.0008	*	1.7930×10^{-57}	***	7.6730×10^{-12}	***
BN 1 vs BN 2	11, 15	no LR	-	Full	0.2270	-	0.0083	-	0.2186	-
BN 1 vs BN 2	12, 16	no LR	-	Full	0.5055	-	0.2421	-	0.4465	-

Table 3.3: Results where the compared experiments are given in the second column. See table 3.1 for the experiment definitions. Significance includes the Bonferroni correction for the 24 simultaneous tests and is given for difference in sample medians determined from a two-tailed Mann-Whitney test at 1% (***), 5% (**), 10% (*) significance level.

effect	parameter	value	% Δ value	mean	sd	% Δmean	% Δmedian	% Δsd	p-value
LR entry	(γ, ρ)	$(0.020, 0.035)$	0.0	0.23	0.033	0.0	0.0	0.0	1.0
		$(0.018, 0.037)$	-10	0.21	0.033	-5.9	-11	35	0.47
		$(0.022, 0.033)$	10	0.23	0.044	0.47	-4.8	0.44	0.87
LR activation	α	0.055	0.0	0.23	0.033	0.0	0.0	0.0	1.0
		0.0495	-10	0.22	0.039	-3.7	0.32	18	0.72
		0.0605	10	0.20	0.033	13	-13	2.2	0.046
mutation kernel	λ_μ	3.55	0.0	0.23	0.033	0.0	0.0	1	
		3.195	-10	0.22	0.039	-2.5	-8.4	20	0.71
		3.905	10	0.21	0.043	-5.3	-11	31	0.33
fitness kernel	λ_δ	0.003	0.0	0.23	0.033	0.0	0.0	0.0	1
		0.0027	-10	0.22	0.039	-2.5	8.4	20	0.71
		0.0033	10	0.21	0.043	-5.3	11	31	0.33
transmission fitness	ϕ	0.003	0.0	0.23	0.033	0.	0.	0.	1
		0.0027	-10	0.22	0.021	-0.14	-3.1	-36	0.85
		0.0033	10	0.23	0.031	0.55	-2.4	-6.1	0.94
reservoir size	(γ, ρ)	$(0.02, 0.035)$	0.0	0.23	0.033	0.0	0.0	0.0	1
		$(0.0145, 0.037)$	-10	0.21	0.044	-5.9	-11	35	0.47
		$(0.0255, 0.033)$	10	0.23	0.033	0.47	-4.8	0.44	0.87

Table 3.4: Simulation sensitivity to a ±10% change in parameters. The resultant change is at 150 years from the start of the simulation for 100 replicates. The parameter *reservoir size* represents a latent reservoir size relative to the actively infected virus, where the latent reservoir is equal to, 10% less, and 10% more than the actively infected virus. P-Value is given for difference in sample medians and is determined from a two-tailed Mann-Whitney test. The simulation variation is high so we suggest the difference is replicate significance overstates the significance of the difference between simulations with only 100 replicates.

Chapter 4

HIV subtype fitness differences *in vivo*

4.1 Introduction

There is evidence of phenotypic differences in replicative fitness, disease progression, and epidemiology among HIV-1 subtypes [Bertels et al., 2017, Venner et al., 2016, Kyeyune et al., 2013]. Previous direct *in vitro* growth competitions have shown subtypes A and D to be 100-fold more fit than subtype C [Morrison et al., 2007]. *In vitro* studies have also shown that subtype C has slower cell fusion times than subtypes A and D [Venner et al., 2016].

In the same dataset, the mean loss of CD4 cell concentration per week differed between subtypes, and a difference was also seen between Uganda, with mainly subtypes A and D, and Zimbabwe, with mainly subtype C, where the loss of CD4 cells was slower in both Zimbabwean woman and for subtype C [Venner et al., 2016].

How these *in vitro* fitness differences between subtypes manifest in HIV infected individuals is not well understood. Our goal is to look for evidence of differences in replicitive fitness and cell fusion times between subtypes *in vivo* that are been seen *in vitro* [Venner et al., 2016, Abraha et al., 2009, Ariën et al., 2005, Ball et al., 2003].

	Uganda	Zimbabwe	
A	84	3	**87**
C	6	174	**180**
D	34	1	**35**
	124	**178**	**302**

Table 4.1: Summary of study participants' region and HIV-1 subtype.

4.2 Study Population

Over four years, from November 1999 to January 2004, over 8000 woman seeking reproductive and general healthcare services were screened at three sites in Uganda and four sites in Zimbabwe. Of the woman screened, 2235 from Uganda and 2296 from Zimbabwe were included in the study population. Study participants were aged 18-35 years, sexually active, not HIV infected, and had not used injection drugs nor had a blood transfusion in the three months prior to enrolment [Morrison et al., 2007].

Follow-up screening visits for study participants were scheduled every 12 weeks, for 15 to 24 months. At the follow-up screening visits, study participants donated blood samples and serum was tested for HIV using enzyme- linked immunosorbent assays (ELISA) for HIV-1 infection and infection was confirmed with a Western blot test. During the study period, 302 of the study participants contracted HIV and are included in our dataset. Participants who contracted HIV-1 subtypes A, C, and D are represented in the data with the division between subtype and region detailed in Table 4.1.

When a study participant tested antibody positive (AB+), HIV polymerase chain reaction (PCR) was performed serially on samples from their previous visits. The date of HIV acquisition in the dataset was defined as the date of the first positive PCR result [Morrison et al., 2007]. The results of the PCR test for HIV-1 viral RNA was recorded in the dataset as RNA+ or RNA- for each sample. Participants who tested antibody positive for an HIV-1 infection were scheduled to return for additional appointments at two week intervals twice, at four week intervals twice, and then every three months.

There are two possible scenarios upon discovery of an infection at a follow-up visit: a participant's sample from a previous screening visit tests RNA+ or RNA-. If a sample from a participant's previous screening visit tests RNA+, then the true date of infection was at least 83 days prior to detection and our data does not contain the initial viral growth phase.

If a participant's sample from the previous screening visit tests RNA-, then the participant became infected in the intervening 83 days between the follow-up visit in which the participant had an AB+ test and the previous screening visit. This window is further narrowed with the assumption that the immune system takes 7 to 14 days for detectable antibodies to develop with negative results from the ELISA for up to 14 days post infection [Alexander, 2016]. Thus, the true date of HIV acquisition falls within 83 to 7 days prior to detection. We can hope to accurately capture the viral growth phase only if the time of first detection was before the viral peak, which we assume to occur within the first 27 days post infection [Karlsson and Mittler, 2000]. Therefore the participant would have to become infected within 20 days of detection.

The two detection routes are summarized in Figure 4.1 where the red indicates the possible time of first infection.

4.2.1 Within-host model

In this chapter we use a standard three compartment within-host model of HIV-1 [Nowak and May, 2000]. A schematic of this three compartment model is found in Figure 4.2. This within-host model considers a well stirred environment and models a systemic HIV infection in the blood. The modeled compartments are: target cells, T, which are activated CD4+ T-cells that can become infected; infected cells, I, that are infected CD4+ T-cells which produce free virus; and free virus, V, which infects target cells. Target cells have a constant entry rate, λ, a natural death rate, d, and become infected when they interact with free virus at rate β and join I. Infected cells are lost with death rate δ. Free virus is produced from I with burst size p and

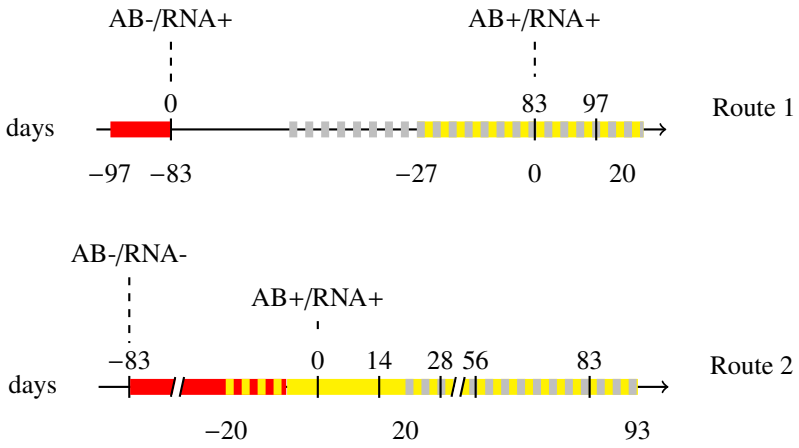


Figure 4.1: The date of participant samples are marked with vertical lines. The days of potential infection are indicated in red. The days in which an infection would have to have occurred so our data potentially contains the viral growth phase is marked in solid yellow. Only infections acquired in the area marked in alternating red and yellow have the potential to contain the viral growth phase. Thus in our dataset only 33 participants are expected to have been infected 7 to 20 days prior to detection, and therefore have the time to set point information captured. The grey bars represent the time past which peak viremia has likely already occurred in our data.

free virus is cleared with clearance rate c. This yields the system

$$\begin{aligned} \frac{\mathrm{d}T}{\mathrm{d}t} &= \lambda - dT - \beta TV \\ \frac{\mathrm{d}I}{\mathrm{d}t} &= \beta TV - \delta I \\ \frac{\mathrm{d}V}{\mathrm{d}t} &= pI - cV. \end{aligned} \tag{4.1}$$

4.2.2 Time to setpoint viral load

This standard host cell limited HIV model, (4.1), is well-studied; it has been shown that a viral phenotype with higher replicative fitness will more quickly reach a higher peak viremia. Having infected all the circulating target cells, the free virus in this model will also decay to a stable setpoint viral load more quickly than a phenotype with a lower replicative fitness [Nowak and May, 2000].

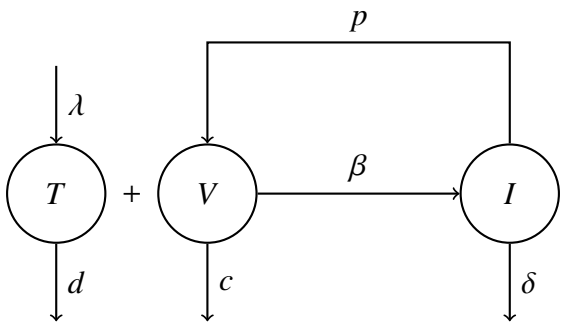


Figure 4.2: Schematic diagram of a standard within-host HIV model. The modelled compartments are: target cells, T; infected cells, I; and free virus, V.

To understand the parameters affecting replicative fitness, consider system (4.1), and let $T = \frac{\lambda}{c}x, I = \frac{\lambda}{c}, V = \frac{c}{\beta}v$, and $t = \frac{1}{c}\tau$. This yields the non-dimensional system

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}\tau} &= 1 - Ax - xv \\ \frac{\mathrm{d}y}{\mathrm{d}\tau} &= xv - By \\ \frac{\mathrm{d}v}{\mathrm{d}\tau} &= Cy - v \end{aligned} \tag{4.2}$$

where $A = d/c, B = \delta/c, C = \lambda\beta p/c^3$. Examining parameter C, we see that the growth rate of target cells λ, the interaction infection rate, β, and burst size, p, all increase the growth rate of the virus, and the free viral clearance rate c decreases the growth rate of the virus.

To detect phenotypic differences in replicative fitness between HIV-1 subtypes, we use time to setpoint viral load as a proxy for replicative fitness. We follow the method of [Morrison et al., 2010] and attempt to use Locally Weighted Scatterplot Smoothing (lowess) smoothing to detect the time to viral load setpoint in our data. The participant data is considered together because each individual participant has too few data points to consider a local regression that could capture the time to the viral load setpoint. The first positive test for RNA is used as the time of initiation of infection for each participant. Included in the appendix are Figures B.6 and B.7 which show the results of lowess smoothing using a sliding window of 20% of the timecourse, and including all data in our dataset.

Identifying the time to the viral load setpoint in the data with lowess smoothing requires a

decision rule because the begining of the stable viral load is not well defined. In the following analysis, we identified the first minimum after the highest rate of descent and use this as a proxy for the time to setpoint viral load.

The choice of the percent of the time course to include in the local smoothing has a strong effect on the time to setpoint viral load that is identified. In Figure 4.3 we see the lowess smoothing results for randomly chosen 90% boot strap samples, using smoothing windows that include 10 to 100 percent of the timecourse from the first 300 days for each participant. These results, summarized in Figure 4.3, show that the choice of lowess smoothing percentage has a strong effect on the observed time to setpoint. In Figure 4.4 the relative order among subtypes of time to setpoint also changes with the choice of smoothing percentage. With these results we conclude that lowess smoothing is not a reliable method for comparing time to setpoint viral load or the setpoint viral load between subtypes in our data.

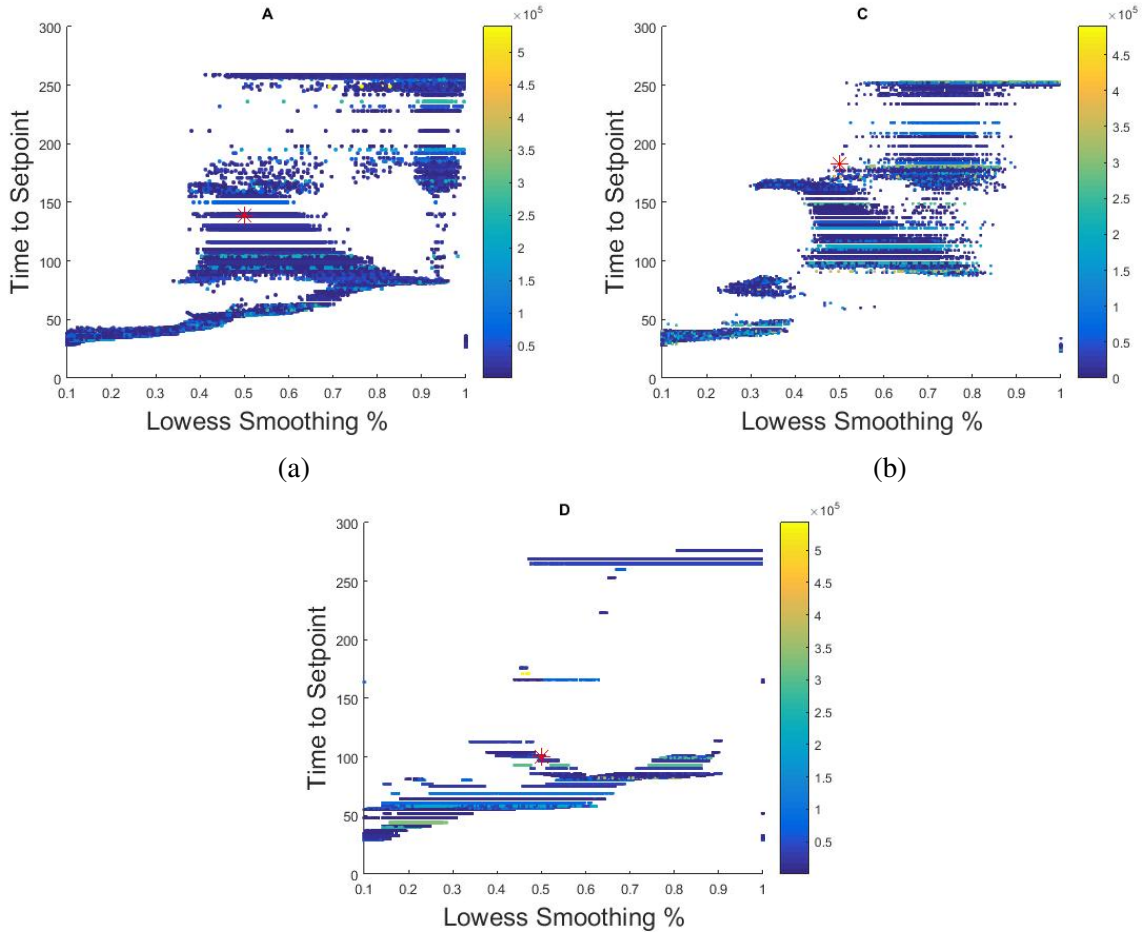


Figure 4.3: Bootstrap samples using 90% of the data for the first 300 days were generated for subtypes A, C, and D shown on panels (a), (b), and (c) respectively. Each dot's location with respect to the vertical axis represents the estimated time to setpoint using the first minimum after the highest rate of descent method, and each dot's location with respect to the horizontal axis represents the percent of the first 300 days of data used in the local smoothing window. The colour of the dots represents the plasma viral RNA/ml at setpoint. The red '*' represents the results from Morrison et al. 2010 [Morrison et al., 2010]. The horizontal location of '*' was arbitrarily chosen since Morrison et al. 2010 did not restrict the data to the first 300 days and a direct comparison is not available.

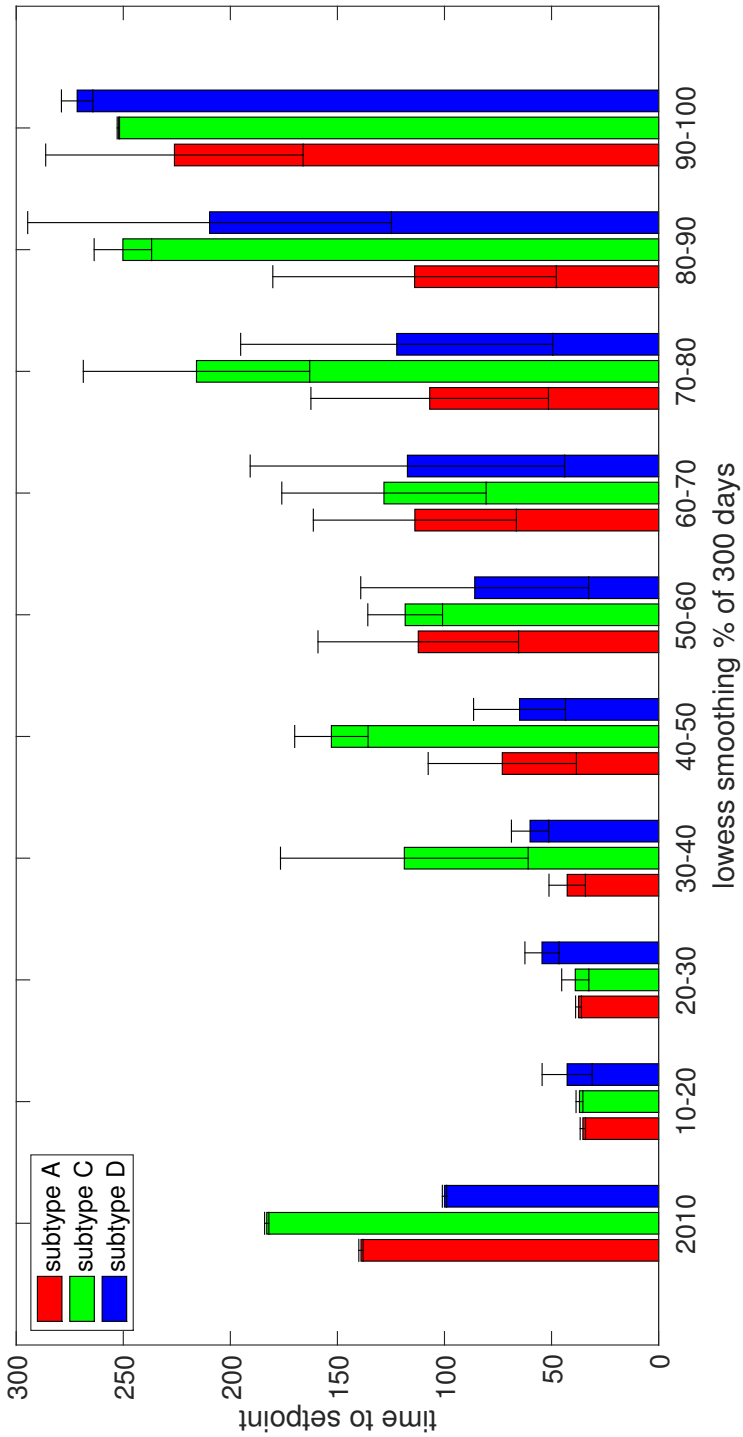


Figure 4.4: The first column contains the time to set point estimates from Morrison et al. 2010 where the data were not restricted to the first 300 days. The remaining columns contain time to setpoint estimates, found using the first minimum after the highest rate of descent decision rule, for the robust lowess smoothing of repeated 90% samples of the data for the first 300 days with randomly chosen smoothing percentages from 10 to 100 sorted into 10% wide bins.

4.2.3 Best Fit Parameters

A major barrier to finding the individual participant's best fit parameters for a within-host model, as well as aggregating the data, is each participant's unknown date of infection. In the previous section the participant data were aggregated using first date of an RNA+ sample as the initiation of infection and all data were aligned at time zero with this initial date for each participant. This crude approximation is not feasible with the limited number of appointments for participants in our dataset and would drastically affect the parameter values found as well as the ability to find best fit parameters. The model dynamics would not fit the data if the date of infection was off drastically and the model was at near equilibrium.

To overcome this difficulty, we fit the time of infection as an additional parameter in the model. Best fit parameters were found for the non-dimensional system (4.2) using an iterative approach and *fminsearch* in *Matlab*. The best fit parameters from each previous iteration were used to scale the viral load from the participants to fit in the next iteration. The system was initiated from the best fit parameters found for the dimensional system. This method was necessary because the units for time and viral load had to be scaled into non-dimensional form, which is not exact unless the exact parameter values are known. Figure 4.5 contains some of the best fits that were found, but adequate model fits were not found for a large number of participants. Too few acceptable sets of best fit parameters for individual participants were found to analyze the differences in replicative fitness between subtypes. We propose that the participants where best fit parameters were not found did not contain the early dynamics of the virus, the dynamics that our within-host model was modelling. In the next section we explore the probability that our data contained these early dynamics given the appointment schedule followed by each participant and whether this appointment timing allowed for high quality fits with acceptable uncertainty so conclusions about replicative fitness can be safely drawn.

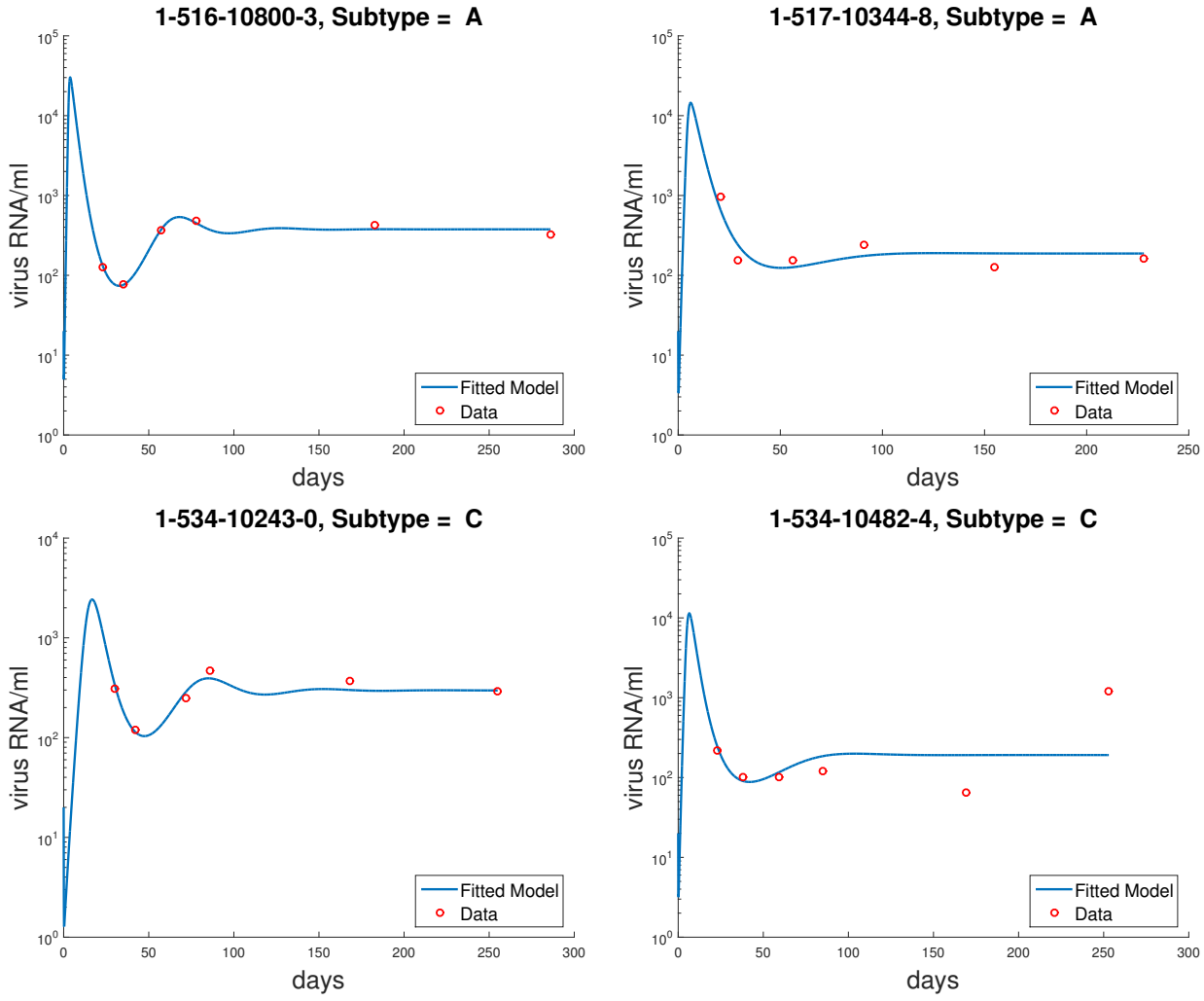


Figure 4.5: participant data, red circles, shown with the solution to 4.1 using the individual best fit parameters found. Insufficient participants have acceptably good fits to analyze the differences in replicative fitness between subtypes.

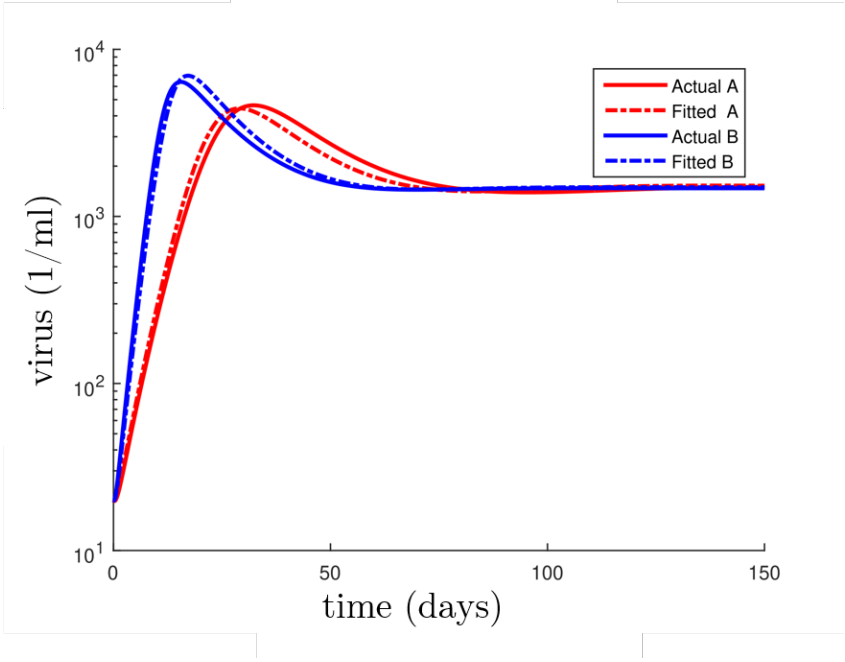


Figure 4.6: The solid lines are the simulated participant types and the dashed lines are the mean of the fitted parameters for the 236 simulated participants with simulated realistic appointment timing and the true time of infection known. The two types of simulated participants share the parameters $\lambda = 12, d = 0.04, p = 20, c = 0.1$. and participant type **A** has parameters $\beta = 3 \times 10^{-5}, \delta = 0.848$ and participant type **B** has parameters $\beta = 8.1 \times 10^{-5}, \delta = 1.22$. Thus the β parameters have a 92% difference and the δ parameters have a 36% difference.

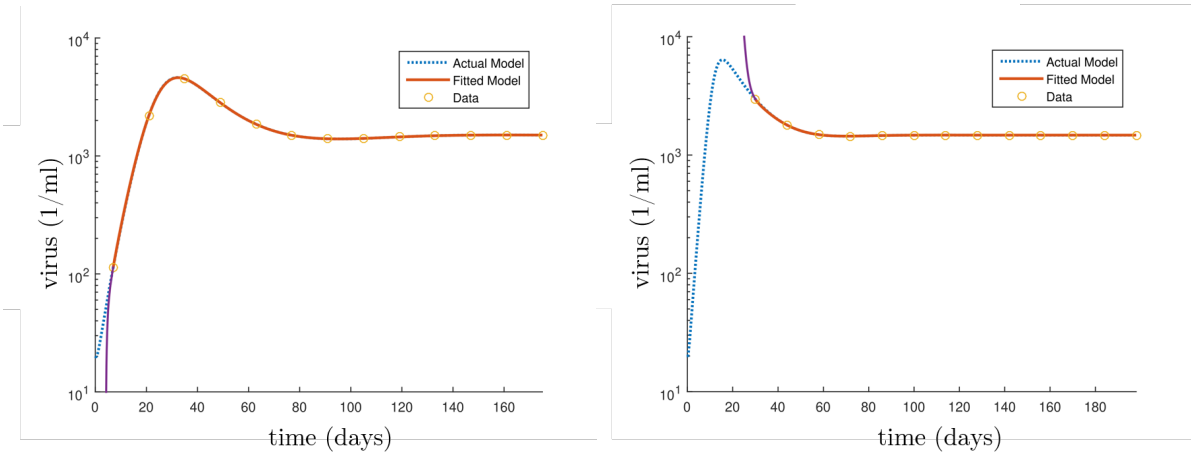
4.2.4 Simulated Participant Data

Participants' appointment timing has an effect both on the best fit parameters found, and also on the probability that the dataset contains viral load data that might allow for an accurate detection of the replication rate. To quantify the uncertainty in the best-fit parameters and to elucidate the probability that the dataset contains the necessary viral load data, we simulated data with three appointment timing regimes and two phenotypically different HIV types. Figure 4.6 shows the two simulated HIV types with different replicative fitness and death rate of infected CD4 cells.

Three screening visit appointment timing regimes are considered: ideal, fixed, and randomly sampled realistic. The ideal regime has the screening visit appointment schedule fixed with appointments every two weeks for 168 days. The date of the initiation of the virus is chosen randomly from 0 to 83 days before the first screening appointment that is simulated. This regime has 7 more screening appointments than were planned in the study. Figure 4.7 shows data simulated with ideal appointment timing, for two simulated participants of type A and B, along with the solution to the model with best-fit parameters. The number of days between the screening appointment at which the virus was detected and the date of initiation of the infection is the *shift* required to fit the simulated participant data with the correct date of initiation of infection.

The fixed regime has the screening visit appointment schedule fixed with appointments at $14, 28, 56, 84, 168$ days and the date of the initiation of infection is chosen randomly from 0 to 83 days prior to the first visit at which the infection was detected. This regime would correspond to the appointment timing in the clinical study, if appointments could be scheduled exactly as planned and no appointments were missed.

In reality, appointments in the clinical study did not occur at precise times, and adherence to the study regime was imperfect. In the randomly sampled realistic regime, as in the fixed regime, the infection is initiated on date randomly chosen uniformly from 0 to 83 days prior the first viral load sample. But, in the randomly sampled realistic regime, realistic timing for



(a) (b)

Figure 4.7: Simulated participants with ideal appointment timing, where the blue dashed line is the actual model, the orange line is the fitted model where there are data-points, and the purple line shows the fitted model prior to the first appointment. Panel (a) shows a participant with a type A simulated infection where the first data point is 7 days post infection. In this case where the time of infection and the first data point are closely aligned the models with best-fit parameters closely match the viral load trajectory but misclassify the time of infection by two days. Panel (b) shows a participant with simulated type B infection and ideal appointment timing, where the first data point is 30 days after the time of infection. In this case the solution to the model with the best-fit parameters also closely matches the data points, but the estimate of the early infection dynamics and the estimate of the time of infection do not match the simulated values.

the participants' follow-up screening appointments is also simulated to match the appointment timing found in the dataset. Histograms of the waiting times seen between appointments are included in the Appendix in Figure B.1.

To match the observed appointment timing distribution, each participant's appointment timing schedule is simulated as a renewal process. To create a distribution of appointments to match the data, future appointments are normally distributed around the next scheduled difference in the appointment schedule regardless of the time of the current appointment. That is, if the participant is visiting for their first follow-up, that would originally have been scheduled around 14 days from the discovery of infection, the next appointment is randomly chosen from a normal distribution with mean 14 days from the current appointment. To match the actual distribution of appointment times, participants also had to randomly miss appointments. The

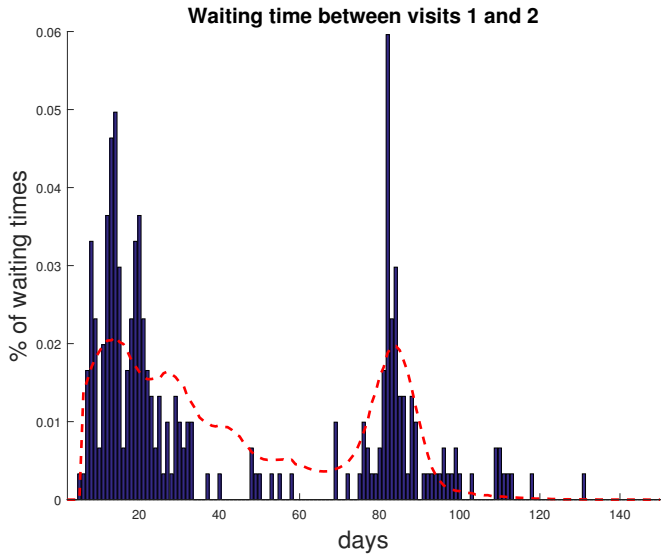


Figure 4.8: The simulated appointment timing distribution generated with a renewal process is indicated with the red dashed line overlayed with the blue histogram of the appointment timing seen in our dataset.

probability of a participant missing an appointment is .4, .3, .2, .1 for the first, second, third, and fourth appointment respectively. Also 40% of participants all miss the frequent appointments and return only at 84 days. An example of the generated distribution is seen in Figure 4.8.

The two simulated HIV phenotypes had different replication rates, β, and infected cell death rates, δ. These changes were chosen to allow for realistically different initial growth rates while still maintaining similar set-point viral loads. Figure 4.6 shows the two phenotypes' within-host dynamics as well as the results of the mean best fit parameters from 236 simulated participants with realistic appointment timing. Note that to obtain these fits, we assume that the infection was detected via route two (Figure 4.1) and that the time of infection is known. This figure indicates that with a known time of infection, and a large sample of participants with realistic and unbiased randomly sampled appointment timing, the two HIV phenotypes can be distinguished.

A major challenge in determining the best fit parameters for our dataset is that the time of infection of the participants is unknown and also needs to be estimated. Figure 4.7 demonstrates this problem under ideal appointment timing, which includes a follow-up sample every

7 days after the virus is detected. In this case if the virus is detected close to the time of infection a reasonable fit to the data was found and the estimated date of infection was close to the true date. As the date of infection moves further from the date of first detection of the virus, even with some of the early dynamics captured and fitting parameters for the reduced dimension model (4.2), the estimates of the time of infection and the best-fit parameters did not match the simulated true parameters and time of infection. Furthermore, a solution for the best-fit parameters was not always possible to find. The unknown time of infection prevented any confidence in the estimated parameter values.

To understand the limitations of finding the best fit parameters, even if the time of infection was known, we found best fit parameters for the 236 simulated participants using the true time of infection. The two simulated phenotype β parameters differed by 92% while the δ parameters had a difference of 36%. This difference affects only the reduced dimensional parameters B and C; the reduced dimensional parameter $A = d/c$ was unchanged, since d and c were identical between phenotypes.

The number of participants that contain the time to peak viral load information is determined by the shift from the day of initiation of infection and the day of the screening visit at which the virus was first detected. In Figure 4.9 the simulated participants best-fit parameters, for the reduced dimension system 4.2, were found with the sample points shifted to the correct number of days from the date of initiation of infection.

Considering this best case scenario, in which the true date of infection is known, we find that the reduced dimensional parameters are correctly identified as significantly different for the ideal appointment timing when all participants with a shift less than 50 days are considered. These differences lose significance, however, when all participants are considered. If only participants with a shift less than 15 days are considered, parameter A is incorrectly identified as different between phenotypes. Note that in this case, only 32 of the 236 participants in the study are considered for analysis.

The fixed appointment timing performed well only when the analysis was restricted to

participants for whom the time of infection was near the time of detection. When participants with longer shifts were included, parameters B and C did not show significant differences, and parameter A was sometimes misidentified as significantly different between subtypes.

For the random appointment timing regime, the ability to detect the difference between phenotypes for parameters B and C increased as participants with longer shifts were included in the analysis, and could be detected for sample sizes greater than 50 participants which correlates with a shift of more than 14 days. The A parameter for the reduced dimension system was never misidentified in the random appointment timing regime. Here the random appointment timing out-performed the ideal appointment timing, but failed to identify the early dynamics if the first appointment was too close to the time of infection. If the time of detection is close to the time of infection the probability that there is a follow-up appointment near the peak viral load is reduced because the initial appointment is before the viral peak.

Using the simulated participants with realistic appointment timing and an unknown time of infection, we calculate the expected number of participants of each subtype with enough information to capture the early behavior in the dataset, that is, with adequate data within the first 50 days post infection. In a dataset of 302 participants the mean expected number of each subtype are type $A = 22$, type $C = 27$, and type $D = 9$.

Figure 4.9 demonstrates that phenotypic differences between the two subtypes could in fact be detectable, if the time of infection for each participant were known. Unfortunately, Figure 4.7 demonstrates that with even a 30 day shift between infection onset and data collection, the time of first infection is not accurately determined by data fitting. Considering both our inability to determine the time of infection, the different phenotypes, and the sample size required to detect those phenotypes, we conclude that the dataset does not contain enough information to determine the phenotypic differences between subtypes, even if they do exist.

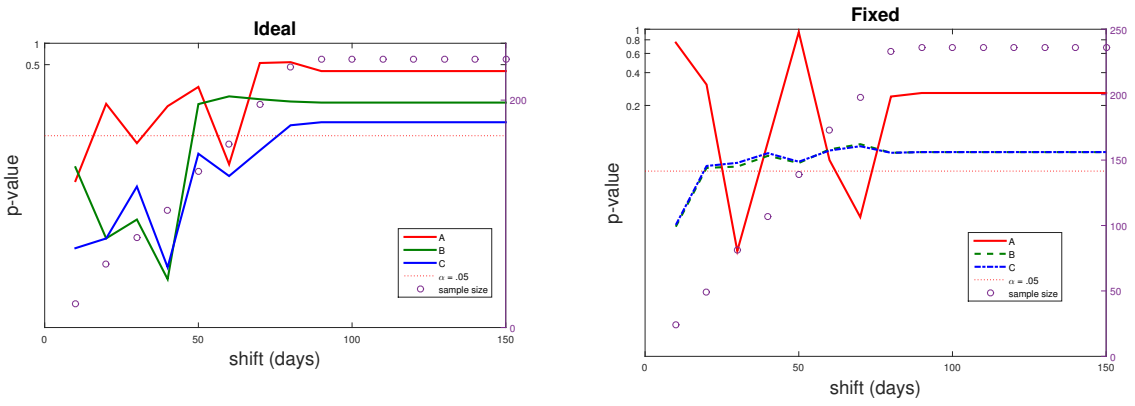


Figure 4.9: The p-value that the best-fit parameters for the two simulated participant phenotypes have unequal means with the 236 simulated participants with ideal (a), fixed (b), and random (c), appointment timing. The red, blue and green lines, corresponding to the left axes, represent the p-values for the dimensionless parameters A, B, and C respectively. The right axes and the circles show the sample size available for a given date of infection within the first number of days shown on the horizontal axes.

4.2.5 Conclusion

Fitting within-host dynamics, without the true time of infection, is a difficult problem and we were unable to find enough high quality fits to the data to determine whether there is a phenotypic difference between subtypes. Using simulated participant data, we discover that the expected number of participants for whom the early dynamics are captured in the dataset is too small to be able to determine, with confidence, a difference between subtypes.

Chapter 5

SHIV Meta Analysis

5.1 Introduction

In vitro fitness competitions between HIV-1 subtypes suggest a relative order in replicative fitness and that this fitness difference is directly related to host cell entry. During *in vitro* fitness competition, subtype B is estimated to have an eight- to ten-fold fitness advantage over subtype C [Marozsan et al., 2005]. This host cell entry fitness advantage does not translate to an advantage for the spread of subtype B in the epidemic. HIV-1 subtype C or recombinant forms containing at least one envelope gene of subtype C accounts for more the 50% of all infections worldwide. HIV subtype C displaced subtype B and the subtype A/E recombinate, CRF01_AE, in South China and is an increasing proportion of HIV-1 infections in Kenya [Abraha et al., 2009, Bbosa et al., 2019].

It appears that the transmission fitness of the virus dominates transmission in the epidemic, and replicative fitness within host plays a smaller role. The goal of this study is to examine the evidence that the *in vitro* fitness advantages of subtype B also exist *in vivo*, and therefore the differences in prevalence and growth of HIV-1 subtypes are indeed in spite of a replicative fitness advantage for subtype B. As seen in Chapter 4, to capture the initial growth of the virus within host and therefore some indication of the replicative fitness of a subtype *in vivo*,

requires high sampling at the time of infection. This has proven to be impractical, and even following approximately 3000 patients with a rigorous schedule of appointments has yielded an inadequate data set to detect differences in the initial growth phase.

To approach the question of replicative fitness of HIV-1 *in vivo* we have chosen to consider an animal model of HIV-1, simian/human immunodeficiency virus (SHIV). Given that both SHIV types use the subtype specific envelope gene from HIV-1, we expect the cell entry dynamics to be a good proxy for the *in vivo* HIV-1 dynamics [Sui et al., 2014]. We consider SHIV SF162 and SHIV 1157 clones, which use the CCR5 co-receptor and are created with the *env* gene from subtype B and subtype C respectively, performing a meta analysis of SHIV infected primates using data from 20 papers [Eugene et al., 2013, Robinson et al., 2010, Lederman et al., 2015, Peterson et al., 2014, Lakhashe et al., 2014, García et al., 2010, Sholukh et al., 2015, Polacino et al., 2008, Mumbauer et al., 2013, Harbison et al., 2014] as well as unpublished data.

Two methods are considered to detect differences in replicative fitness *in vivo*: a compartmental model of viral dynamics is used to find best fit parameters, and the interval with the rate of maximum growth is found as a more direct proxy of replicative fitness [Nowak and May, 2000, Ribeiro et al., 2010].

First, best fit parameters for a within-host model of viral dynamics are found. These parameters give insight into the efficiency of host cell entry through the replicative fitness term, β, which includes host cell entry as one of the dynamic processes it models. Fitting the within-host model also provides insight into the other life history traits of the virus and a more complete picture of the interactions between host and pathogen.

The selection of a within-host model that is identifiable but also retains the relevant dynamics is challenging. To keep the data fitting process tractable we have chosen to consider a class of within-host models that do not explicitly consider delays. The efficiency of viral attachment and time required for host cell entry are indirectly included in rate term β which also includes the probability of interactions between infected and target cells. Thus the differences in the

time required for cell attachment or cell entry, which could be what is driving the observed *in vitro* replicative fitness differences, are captured by the models.

Experimental and biological differences among primates, which include inoculum size and infection route, confound estimates of parameter values when fitting viral load to a dynamic model of acute SHIV from diverse experimental studies. Viral load data outside the exponential growth phase has a strong influence on the best fit parameters and may obscure replicative fitness differences. For this reason we limited the data considered to viral load data taken for the first year post infection.

Second, following a method similar to [Ribeiro et al., 2010] for the determination of the maximum viral growth rate, the sampling interval with the maximum increase of viral load was found. This method complements the within-host model as it is a more direct proxy of *in vivo* replicative fitness.

A simplified model of the exponential growth phase is developed and used to estimate the within-host model parameter β from the rate of maximum growth found using interval of maximum slope. A hybrid model, using a simplified model of the growth phase to detect the maximum growth rate in the first few weeks of infection, as an alternative to finding the interval of maximum slope, is left for future work.

Both methods – fitting a within-host dynamics model and finding the interval of maximum growth – share sensitivity to the sampling interval timing, with the interval of maximum growth being less comparable across subjects with different sampling intervals than the model fitting approach.

Using an animal model and SHIV as a method to detect HIV-1 subtype differences has several limitations. The primates are given viral inoculums that ensure a near 100% infectivity. This high initial dose of virus may bypass some of the early viral competition. But even the largest viral inoculums are relatively small compared to the large number of circulating virus particles during the explosive growth stage of initial infection. With simian immunodeficiency virus, depending on the challenge method, the dose of the virus inoculum significantly affects

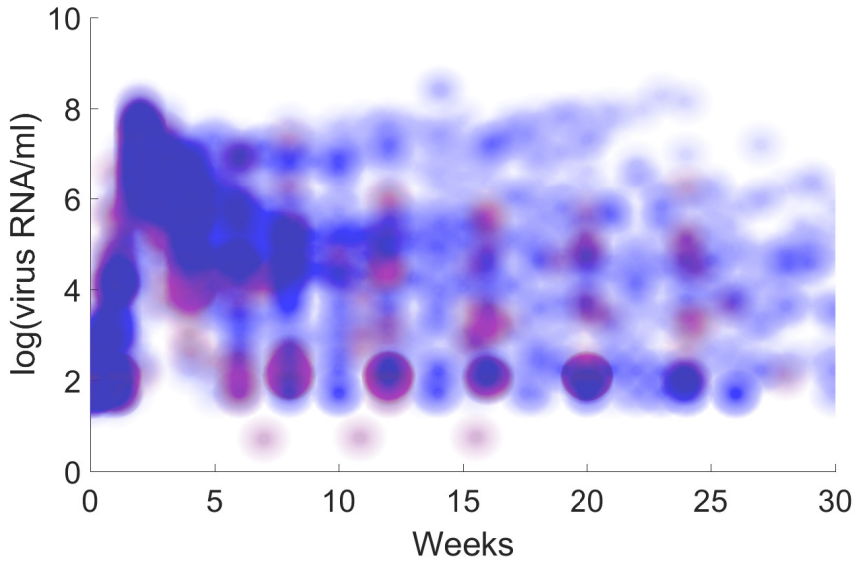


Figure 5.1: Density plot of SHIV SF162 (Subtype B) in blue and SHIV 1157 (Subtype C) in red where more data points in an area result in higher colour intensity.

the number of transmitted variants [Sui et al., 2014].

5.1.1 Dataset

Viral load time series for up to one year from 143 primates, yielding 2179 data points, were compiled from multiple studies and unpublished data. The data represent both rhesus and pig-tailed macaques infected with SHIV subtype C (SHIV 1157) and subtype B (SHIV SF162). The primates were challenged with SHIV in three ways: intrarectal, intravaginal, and intravenous. Three rhesus macaques that were given a subtype B intravaginal challenge failed to become infected and were subsequently challenged with a second higher dose; in the results to follow, the covariate *double* is used to denote this situation. See Table 5.1 for a summary of the dataset and Figure 5.1.1 for a representation of the data. In Figure 5.1.1 the wide range of within-host dynamics and set-point viral loads is clearly seen.

The primates challenged with SHIV SF162 are further divided between SHIV isolates, SF162, SF162P4, SF162P3N, and SFV162P3. The primates challenged with SHIV 1157 are divided between SHIV isolates 1157ipd3N4, 1157ipEL, and 1157i. The different iso-

	Pigtailed	Rhesus	
Subtype B	21	91	**112**
Subtype C	3	28	**31**
Intrarectal challenge	12	59	**71**
Intravaginal challenge		48	**48**
Intravenous challenge	12	12	**24**
	24	**119**	**143**

Table 5.1: Primate Data Summary

lates within each subtype are much more closely related than isolates between subtypes. This smaller within-subtype diversity could also result in fitness differences. For further details about the data and methods of challenge see Appendix C.

5.2 Best Fit Parameters

5.2.1 Within-host Dynamics Models

A series of within-host nested models are considered. The nested models are derived from two closely related compartmental models. The base within-host model considered is similar to the generalized immune response model introduced by [Nowak and May, 2000]. The most complete model extends the base model to include resting CD4+ T-cells. The addition of a resting CD4+ T-cell compartment was motivated by finding the best fit parameters for the data in this chapter as well as the experience gained from chapter 4. The addition of a resting CD4+ compartment accounts for the initial stock of CD4+ cells that cannot be initially infected but are more quickly available for infection than the CD4+ cells' reproduction rate. Thus, the addition of a resting CD4+ cell compartment accounts for the delay in activation of the cells that are initially available within the host.

The most complete model has compartments representing: resting CD4+ T-cells, z, which are not susceptible to infection; activated CD4+ T-cells, x, which can become infected; infected CD4+ T-cells, y, which produce free virus; free virus, v, which both activates an immune

response and infects activated CD4+ T-cells; and a compartment to capture a general immune response, u. The second compartment of activated CD4+ T-cells was added to better capture the dynamics of early infection and to allow the comparison of measured CD4+ T-cells with the CD4+ levels in the model. In the absence of an activated CD4+ compartment, only cells that can be infected are considered in the model.

The parameters in this model are: natural growth of naive CD4+ T-cells, λ; natural death rate of CD4+ T-cells, d_x; a function for immune activation rate of CD4+ T-cells function, $A(\alpha, z, y, v, u)$; infection and interaction rate between virus particles and activated cells, β; increased death rate of infected cells, d_y; general immune killing of infected cells, r_y; effective viral burst size, p_0; general immune clearance of virus particles, r_v; natural clearance of virus particles, c_v; virus induced activation of general immune response, α_v; and the degradation of the immune response, d_u. This yields the following 11-parameter model of within-host viral dynamics,

$$\begin{aligned}
\frac{\mathrm{d}z}{\mathrm{d}t} &= \lambda - d_x z - A(\alpha, z, y, v, u) \\
\frac{\mathrm{d}x}{\mathrm{d}t} &= A(\alpha, z, y, v, u) - d_x x - \beta x v \\
\frac{\mathrm{d}y}{\mathrm{d}t} &= \beta x v - (d_x + d_y) y - r_y y u \\
\frac{\mathrm{d}v}{\mathrm{d}t} &= p_0 y - r_v u v - c_v v \\
\frac{\mathrm{d}u}{\mathrm{d}t} &= \alpha_v v - d_u u.
\end{aligned} \tag{5.1}$$

Activation of resting CD4+ T-cells is the result of the interaction of the host's immune system and the virus or its products, but the details of this activation and how they map onto the generalized immune compartment are not well understood. The within-host model (5.1) is constructed with a general function, $A(\alpha, z, y, v, u)$, for activation which reflects the uncertainty of the appropriate form and is a function of z and the interaction with infected T-cells, free virus, and the generalized immune response. When finding the best fit parameters we model activation with a linear constant activation rate of αz as well as with a non-linear interaction

term which includes y, v, or u. e.g. $A(z, y, v, u) = \alpha z u$.

The units of the generalized immune response compartment do not have a biological interpretation, and therefore we are able to re-scale this compartment to reveal correlations between parameters and reduce the dimension of the system. The substitution $\frac{1}{r_v}\bar{u}$ for u yields the system,

$$\begin{aligned}
\frac{\mathrm{d}z}{\mathrm{d}t} &= \lambda - d_x z - A(\alpha, z, y, v, \frac{1}{r_v}\bar{u}) \\
\frac{\mathrm{d}x}{\mathrm{d}t} &= A(\alpha, z, y, v, \frac{1}{r_v}\bar{u}) - d_x x - \beta x v \\
\frac{\mathrm{d}y}{\mathrm{d}t} &= \beta x v - (d_x + d_y) y - \frac{r_y}{r_v} y \bar{u} \\
\frac{\mathrm{d}v}{\mathrm{d}t} &= p_0 y - \bar{u} v - c_v v \\
\frac{\mathrm{d}\bar{u}}{\mathrm{d}t} &= \alpha_v r_v v - d_u \bar{u},
\end{aligned} \tag{5.2}$$

and shows clearly the expected correlation between the immune activation rate and the two killing rates. To simplify notation, the bar was dropped from $\bar{u}$, and we define $a_v := \alpha_v r_v$.

The simpler model that is considered, system (5.3), does not have the added compartment z for resting CD4+ T-cells and instead activated CD4+ T-cells are introduced at a constant rate, λ.

$$\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}t} &= \lambda - d_x x - \beta x v \\
\frac{\mathrm{d}y}{\mathrm{d}t} &= \beta x v - (d_x + d_y) y - \frac{r_y}{r_v} y u \\
\frac{\mathrm{d}v}{\mathrm{d}t} &= p_0 y - u v - c_v v \\
\frac{\mathrm{d}u}{\mathrm{d}t} &= a_v r_v v - d_u u.
\end{aligned} \tag{5.3}$$

Model Fitted	Acronym	System	Compartments	Set to 0	$A(\alpha, z, y, v, u)$
Base Model	Base	5.3	x, y, v	u_0, a_v	-
Immune Response	IR	5.3	x, y, v, u	-	-
Resting CD4+ T-cell	Resting	5.2	z, x, y, v	u_0, a_v	-
Resting-Immune Response	RIR-Linear	5.2	z, x, y, v, u	-	αz
	RIR-u	5.2	z, x, y, v, u	-	$\alpha z u$
	RIR-y	5.2	z, x, y, v, u	-	$\alpha z y$
	RIR-v	5.2	z, x, y, v, u	-	$\alpha z v$

Table 5.2: Summary of the within-host models used to find best-fit parameters.

5.2.2 Model selection

Nine combinations of within-host model, data, and covariates were considered. An overview of the models considered can be found in Figure 5.4 and Table 5.2. The *Immune-Response* (IR) within-host model, which is based on system (5.3), was considered first and formed the basis of the initial population parameter estimation. All initial population parameters were informed by the literature where possible and chosen so that the qualitative dynamics of viral load for the numerically solved models matched the primate data. The population parameters were then refined for each statistical model iteratively by running the SAMBA algorithm (described in Section 5.2.3 below), setting the results as new initial parameters until the models were consistently convergent. To achieve consistent convergence some parameters that did not show differences between covariates of interest were fixed or restricted with prior distributions after the initial fits were found. For the three models with the lowest corrected Bayesian Information Criteria (BICc) an additional convergence assessment was carried out.

Base Model

The *Base* model is the statistical model based on the within-host system (5.3) with the parameters $u_0 := 0$ and $a_v := 0$ so that u remains at 0 and the immune response is not considered. A summary of the initial estimates and distributions of the initial estimates is found in Figure 5.3.

Parameter	Distribution	Covariate	Population IE	Units
λ	Lognormal	Species	8000	cells (ml·day)$^{-1}$
β	Lognormal	Subtype	2.92e-6	ml (virions·day)$^{-1}$
d_x	Lognormal	-	0.04	day^{-1}
d_y	Lognormal	-	0.8	day^{-1}
p_0	Lognormal	Species,Subtype	205	day^{-1}
c_v	Lognormal	-	2	day^{-1}
T_0	Lognormal	-	10000	cells
v_0	Lognormal	-	0.03	virions

Table 5.3: Base model initial parameter distributions, covariates, and initial estimates (IE).

Immune Response Model

The *Immune Response* model is the statistical model using the base within-host system (5.3) with the immune response considered. A summary of the initial estimates and the distributions of the initial estimates is found in the appendix Figure D.1.

Resting CD4+ T-cell Model

The *Resting* CD4+ T-cell model is the statistical model using the more complex within-host system (5.2) with $u_0 := 0$ and $a_v := 0$ so that the immune response is not considered. A summary of the initial estimates and distributions of the initial estimates is found in appendix Figure D.2.

Resting-Immune Response Models

The *Resting-Immune Response* (RIR) models are the statistical models using the more complex within-host model (5.1) with the immune response considered. The dynamics of activation of the CD4 T-cells are not fully understood and the most appropriate form for the activation function, $A(\alpha, z, y, v, u)$, is not clear. The early dynamics of an HIV infection are better fit if some stock of resting CD4+ cells are available for activation and further infection. We have chosen to consider four simplified activation functions and acknowledge the best model would include a more complex relation between the virus, infected cell, and immune response.

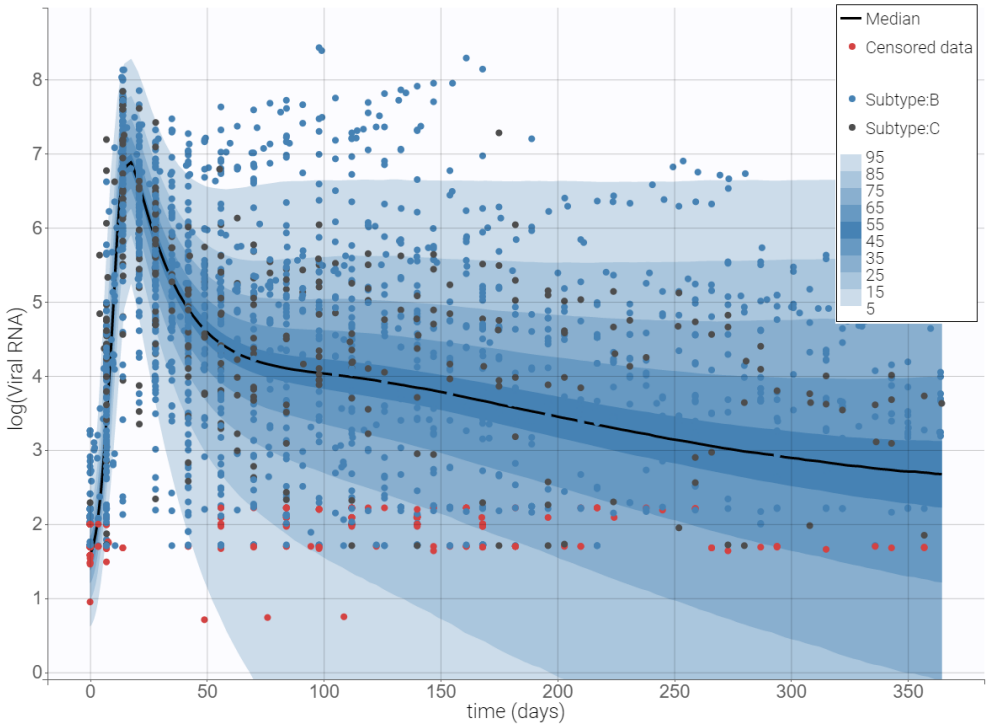


Figure 5.2: Primate viral load data superimposed over the numerical predicted distribution from 500 simulations using the parameters' posterior distributions from the RIR-Linear model.

We consider activation of CD4+ cells from the interaction of resting cells and the immune response αzu, free virus αzv or infected cells αzy as well as a linear activation of CD4+ T-cells αz; these four models are referred to as RIRu, RIRv, RIRy and RIR-Linear, respectively. The linear activation function simplifies the dynamics of the activation of CD4 T-cells and is only a reasonable approximation once the immune response reaches equilibrium, but was the only model that was successfully fit to the full dataset using the Stochastic Approximation Expectation-Maximization (SAEM) algorithm (as described below). A summary of the initial estimates and the distributions of the initial estimates is found in appendix Figure D.3.

When considering the numerical predicted distribution (Figure 5.3) from the RIR-Linear model fit, it is clear that the dynamics of some primates are not captured. From the numerical predicted distribution (Figure 5.3) thirteen primates were identified as outliers, i.e. primates whose viral loads were outside the eighty-fifth percentile beyond fifty days. To address the

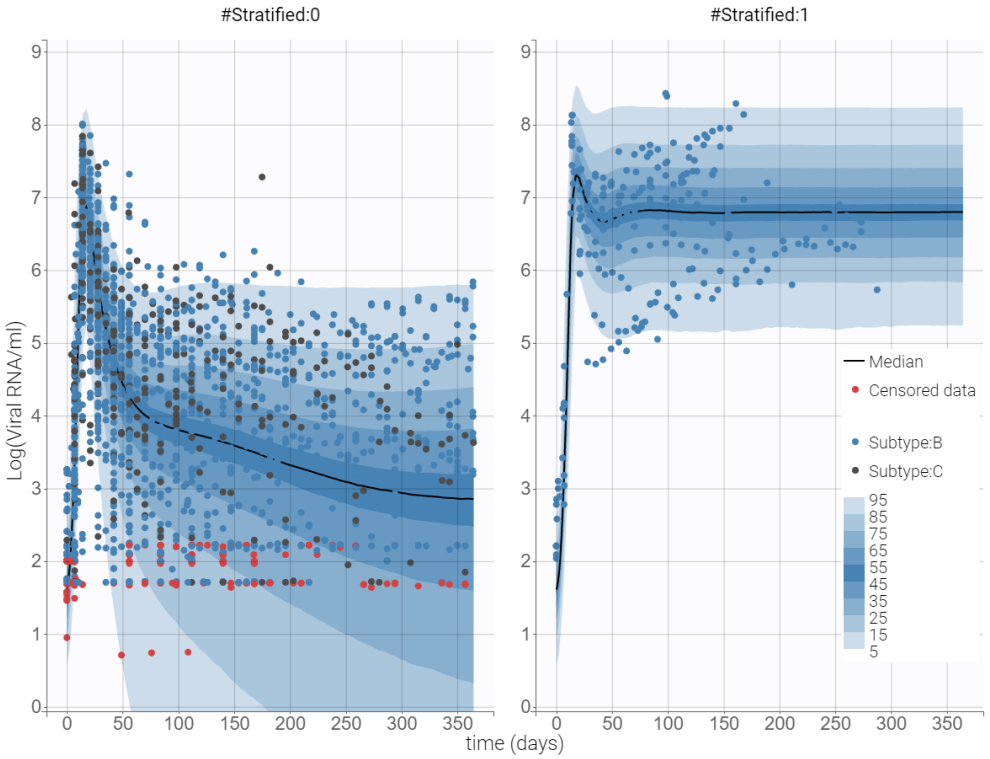


Figure 5.3: Primate viral load data superimposed over the numerical predicted distribution from 500 simulations using the parameters' posterior distributions from the RIR-Linear-Stratified statistical model fit with the stratified dataset excluding the outlying primates shown on the left and the data from the outlying primates on the right.

thirteen outliers we tried two approaches: stratifying the data or adding a latent covariate.

To stratify the dataset a numerical predicted distribution was found from the fit RIR-Linear model by simulating 500 viral load trajectories with the parameters drawn from the posterior distributions. This separated the data into two sets with 13 primates with higher sustained viral loads and 130 with the dominant dynamics. The primates with higher sustained viral loads were all pig-tailed macaques with SHIV subtype C. The RIR-linear model was then fit for the data that was stratified with a covariate for the two sets of primates with qualitatively different dynamics.

For the second approach, to identify the primates whose dynamics were not captured, the RIR-Linear model was fit with the addition of a latent variable that was a covariate for the immune response parameters r_y, a_v, and d_u.

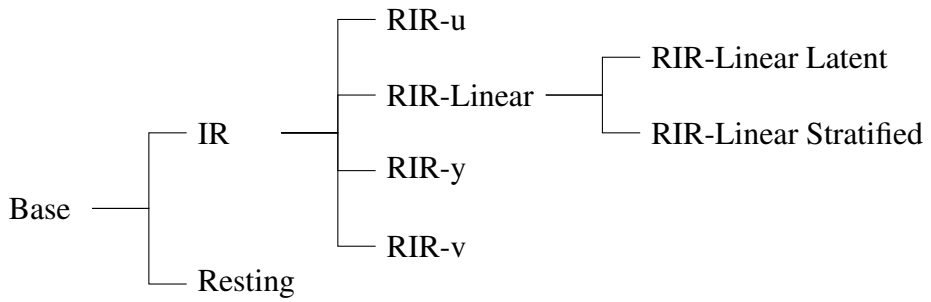


Figure 5.4: Hierarchy of statistical models.

Statistical Model	Within-host model	$A(\alpha, z, y, v, u)$	Add. Covar.	BICc
Base Model	(5.3) with $a_v, u_0 := 0$	NA	NA	8310.4
IR	(5.3)	NA	NA	4794.19
Resting	(5.1) with with $a_v, u_0 := 0$	αz	NA	7475.86
RIR-Linear	(5.1)	αz	NA	4684.32
RIRu	(5.1)	$\alpha z u$	NA	NA
RIRv	(5.1)	$\alpha z v$	NA	NA
RIRy	(5.1)	$\alpha z y$	NA	NA
RIR-Linear	(5.1)	αz	Latent	4642.45
RIR-Linear	(5.1)	αz	Stratified	4604.02

Table 5.4: Comparison of statistical models. This table shows the within-host model fit for each statistical model along with the functional form of $A(\alpha, z, y, v, u)$, any covariates added to the model and the corrected Bayesian Information Criterion. The models without a BICc failed to converge.

5.2.3 Statistical Model Selection

A non-linear mixed-effects model was developed and fit to the data for each of nine within-host dynamics models. Model details can be seen in Figure 5.4. The space of possible covariate and correlation model variations are prohibitively large to test all possibilities. To allow comparison between models, the covariate and correlation model selection was completed on the Immune-Response (IR) within-host model. There was not support for including correlations and the covariate model that had the most support in the IR model was then used when finding the best fit parameters for all nine models.

Automatic covariate and correlation model building was used to select the best statistical model. The covariate model was determined using the *Stochastic Approximation for Model Building Algorithm* (SAMBA) implemented in the monolix package. The biologically relevant

	Species	Subtype	Challenge	Dose	Double
λ	X				
β	x	X			
d_x	x				
d_y	x	x			
r_y	x	x			x
p_0	X	X			
c_v	x	x			
a_v	x	x	x		x
d_u	x				
T_0	x				
v_0			x	x	
u_0	x				x

Table 5.5: The covariates tested using the SAMBA algorithm are marked with x where the uppercase X indicated the covariate was selected by the algorithm.

covariates considered are: species, for pig-tailed and rhesus macaques; challenge for intrarectal, intravaginal and intravenous challenges; dose for the initial viral inoculum received which is comparable only within studies; and "double" for the primates that failed to become infected after the first challenge. Figure 5.5 details the biologically relevant covariates that were considered for each model parameter. Other possible covariates (which were judged to have weaker biological justification) were locked out of the analysis. Based on the results of the SAMBA algorithm the primate species was used as a covariate for λ and p_0 and subtype was used as a covariate for β and p_0. No correlations between parameters were strongly suggested by the automatic model selection.

All initial model fits were performed with the covariates found for the IR model. The automatic covariate model building was repeated using the *Conditional Sampling use for Stepwise Approach based on Correlation test* (COSSAC) algorithm for the RIR-Linear model and the RIR-Linear model with the stratified dataset after the best-fit parameters were found. This COSSAC analysis yielded the same suggested covariates as the covariates determined by the SAMBA algorithm for the IR model.

For the models with a promising Corrected Bayesian Information Criteria (BICc), after the initial run of the SAEM algorithm, the Fisher Information matrix of the estimates was

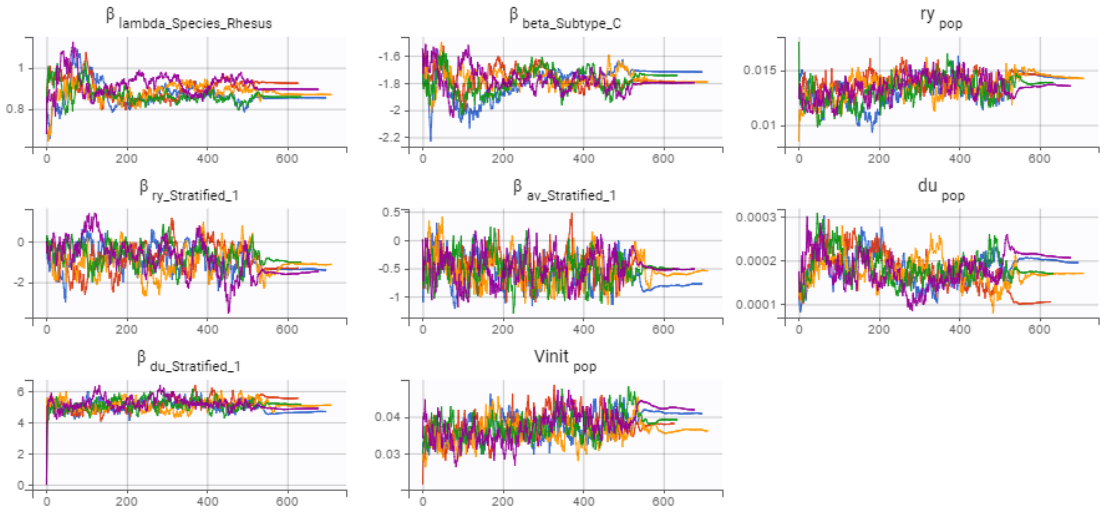


Figure 5.5: The convergence diagnostic plots for a refinement iteration of the RIR-Linear statistical model with the stratified data covariate. Estimated parameter values are shown over the iterations of the SAEM algorithm for four model runs with random initial parameters.

estimated and considered to check the uncertainty of the parameter estimates and whether the model was over determined. Any parameters with correlation values greater than 0.5, which suggests non-identifiability of the parameter, were considered and one of the parameter values was fixed. This process was repeated with the addition of sensitivity analyses of the parameters by completing five additional runs of the SAEM algorithm to check the convergence via the convergence diagnostic plots. Figure D.4 contains the RIR-Linear convergence diagnostic plot with the stratified data covariate from a refinement iteration.

5.3 Results

The RIR-Linear Stratified model achieved the lowest Corrected Baysian Information Criteria (BICc) of 4604.2. The Fisher information matrix condition number for the model is 6.4, which is a strong indication that the model is not over-specified.

The individual weighted residuals (IWRES) and normalized prediction distribution errors (NPDEs), shown in Figure 5.6, are roughly centered around zero. There is room for a better specified model, but this might require more covariates than available in our dataset and a

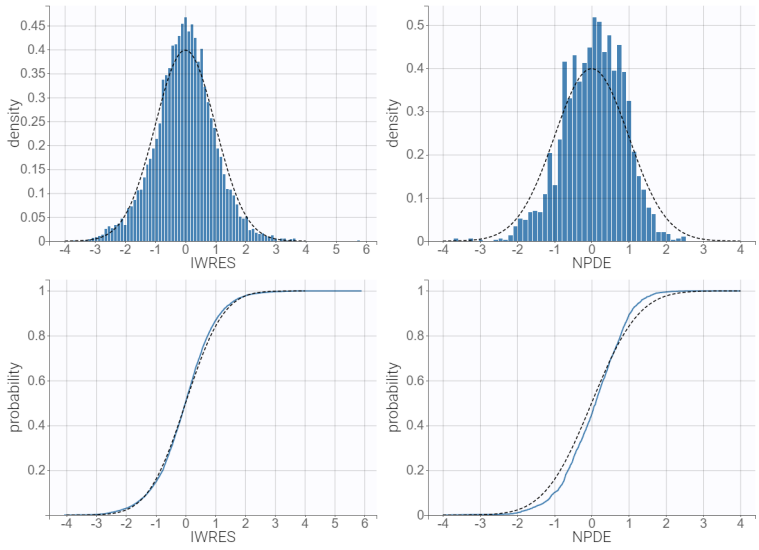


Figure 5.6: Individual weighted residuals (IWRES) and normalized prediction distribution errors (NPDEs), together with the standard Gaussian probability density function and cumulative distribution function for the RIR-Linear Stratified statistical model.

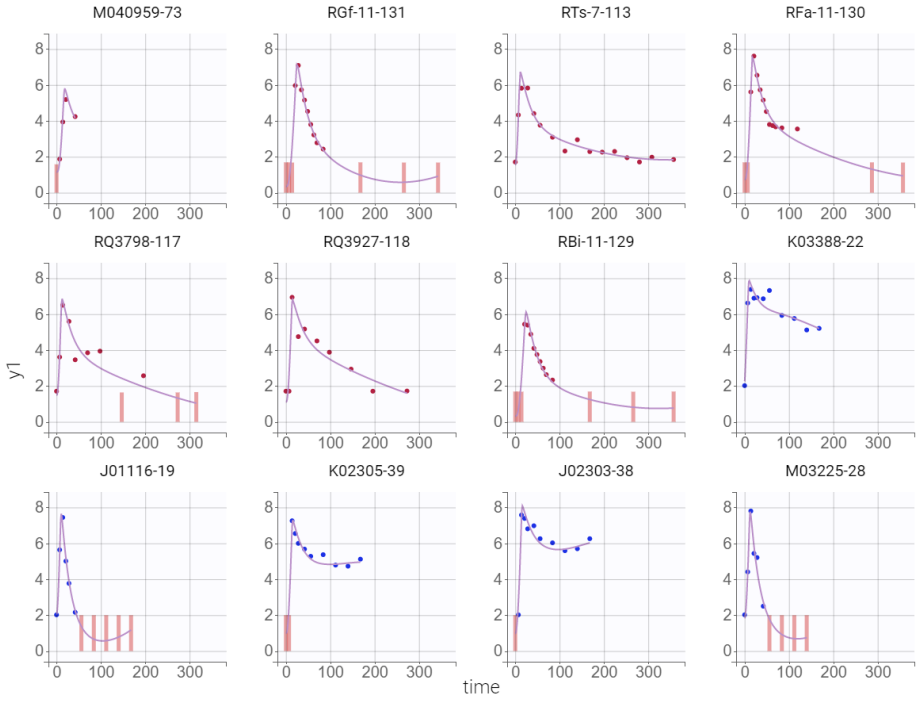


Figure 5.7: Example individual fits from RIR-Linear-Stratified. Plots of viral load over time are shown with the data points and the primates individual fits. The red dots represent the primate data for subtype C and the blue data for subtype B. Pink bars indicate viral load measurements that were below the limits of detection.

more complicated model that will be difficult to fit. A clear avenue for improvement is the more biologically realistic activation functions that we attempted to fit.

Figure 5.4 shows the viral load data superimposed over the numerical predicted distribution from 500 simulations sampled from the parameters' posterior distributions of the best-fit parameters, with the stratified dataset excluding the outlying primates shown on the left and data from the outlying primates shown on the right. With the stratified data we can see that the predicted parameters capture the majority of the data, but a few primates with relatively high viral loads are not well modeled. This further indicated that a covariate that is not available is important to fully understand the data. An attempt to use a latent covariate to capture the sustained high viral load was attempted in the RIR-Latent model, but the latent covariate was unable to detect the primates with high viral load and only isolated a single primate for which the covariate was applied. Below we speculate that the immune response is driving the changes in viral load and find some evidence of immune differences between the two stratified datasets, as seen in Figure 5.8.

Further considering the results in Figure 5.8 from the RIR-Linear Stratified model, we see there is a relative replicative fitness advantage, as represented by β, with SHIV SF162 created with the *env* gene from subtype B over SHIV 1157 created with the *env* gene from subtype C. This is consistent with the experimental results that show subtype B having an *in vitro* fitness advantage over subtype C [Marozsan et al., 2005]. The overall within-host fitness advantage is less clear with subtype C showing an increased burst size parameter p_0 over subtype B. The overall effect of these competing life history traits is not clear with the experimental results supporting the importance of host cell entry.

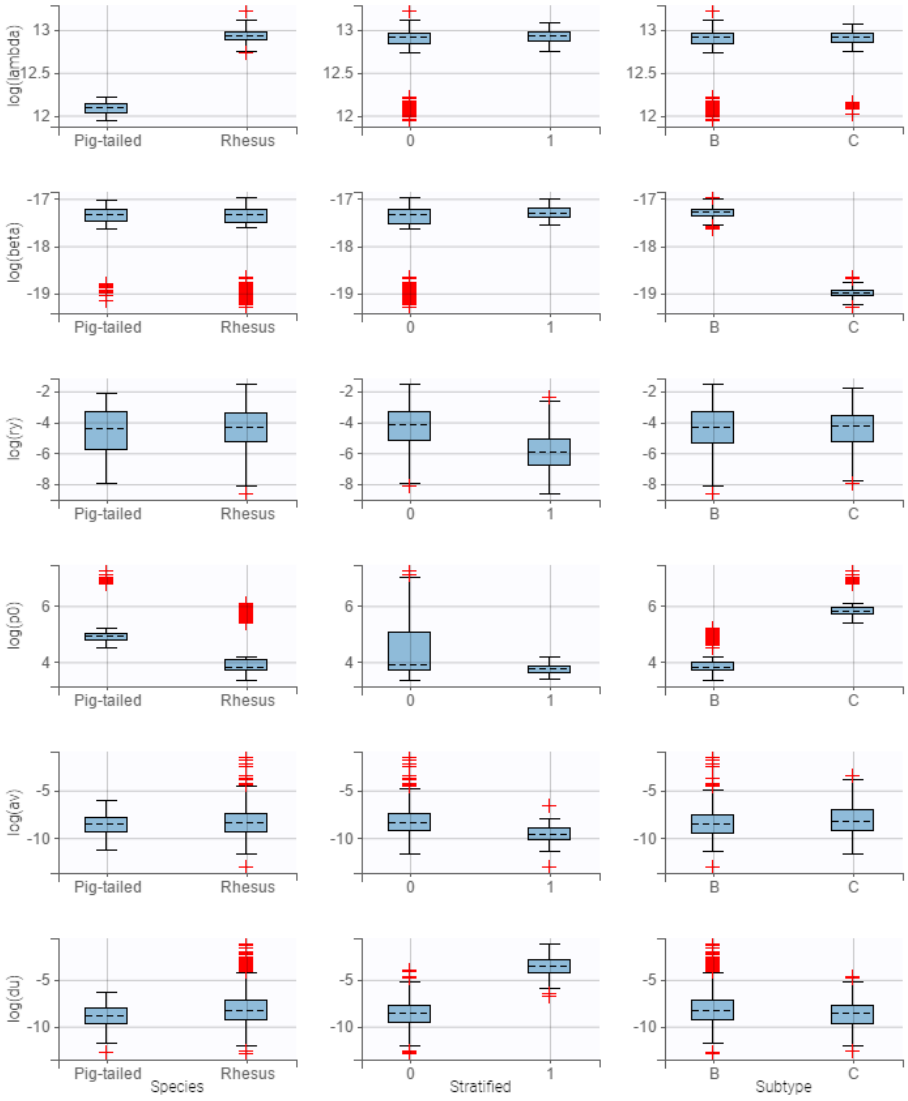


Figure 5.8: Results from the RIR-Linear Stratified statistical model showing the parameter differences between covariates. Only parameters that display differences for the covariates of interest are shown. The parameter values are displayed using log_e.

The mechanics of any particular experiment and virus might play a pivotal role in determining the relative effect of the competing life history traits of replicate fitness and viral burst size. The details of host entry and co-infection could have unexpected results that are not captured by the within-host models considered in this chapter, and therefore using these parameter estimates results to model the in-vitro competition may be misleading. Simple comparison of the relative size of the change in burst size verses replicate fitness is not a viable approach to determine the overall fitness effect.

There is also indication that the qualitative differences that lead to the stratification of the data were a result of the immune killing rate ratio $\frac{r_y}{r_v}$. In particular the outlying data, with higher sustained viral load, showed a slight bias toward removal of virus over infected cells. A stronger difference is seen between the two stratified data sets in the rate of degradation of the immune response, with the outlying data showing a faster degradation.

This result suggests that a generalized immune response is not a good proxy for the much more complex immune response in primates, and underscores the issue that parameter identifiability is a difficult problem when dealing with complex model equations. Considering an alternate reduced dimension system:

$$\begin{aligned}
\frac{\mathrm{d}z}{\mathrm{d}t} &= \lambda - d_x z - A(\alpha, z, y, v, \frac{1}{r_v}\bar{u}) \\
\frac{\mathrm{d}x}{\mathrm{d}t} &= A(\alpha, z, y, v, \frac{1}{r_v}\bar{u}) - d_x x - \beta x v \\
\frac{\mathrm{d}y}{\mathrm{d}t} &= \beta x v - (d_x + d_y) y - r_y d_u y \bar{u} \\
\frac{\mathrm{d}v}{\mathrm{d}t} &= p_0 y - r_v d_u \bar{u} v - c_v v \\
\frac{\mathrm{d}\bar{u}}{\mathrm{d}t} &= \frac{\alpha_v}{d_u} v - \bar{u},
\end{aligned} \tag{5.4}$$

for (5.1) we see the degradation of the immune response, d_u, is tied, through the scaling of the immune compartment, to both removal rates. Therefore our choice of re-scaling the

immune compartment to be proportional to the removal rate of free virus may have shifted the effect of the immune activation rate to the immune degradation rate. This highlights the difficulty with the non-identifiable parameters in complex systems.

The activation of the immune response, a_v, does not show a strong difference between the two data sets. The Wald test shows only marginal support for stratification as a covariate of a_v with a p-value of 0.142 and the distribution of standardized random effects shows that a_v is the only parameter with random effects not centered at zero. So there is evidence that a_v is not well estimated.

The best fit parameters from the RIR-Linear Stratified model also show a higher viral burst size, p_0, for pig-tailed relative to rhesus macaques that may have implications for experimental design considering mixed types of primates and also could be a valuable avenue for future research. Differences in the reproduction rate of CD4+ cells were also detected with rhesus macaques showing a higher rate of creation of target cells. No other significant parameter differences were found between the two primate species. We attempted to find best fit parameters with the additional differences in origin of the primates but did not have enough data to further find reliable fits for the four different primate origins and species categories.

All thirteen primates identified as outliers with high sustained viral loads were rhesus macaques infected with SHIV SF162 (subtype B). Considering only the stratified data set with the outlying primates removed, Figure 5.9, the primary conclusions for the differences between subtypes are maintained. The only large change is in the β estimates for rhesus macaques where the outlying parameter estimates (the outlying red plus signs seen in Figure 5.8 for β and p_0) are now included in the boxplot.

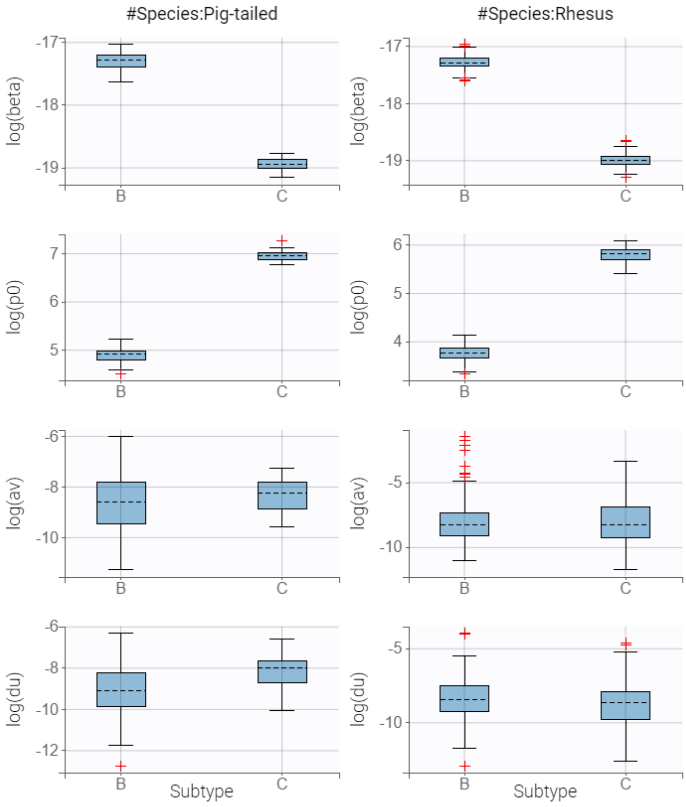


Figure 5.9: Parameter estimates for the RIR-Linear Stratified model (outliers removed) showing the parameter differences between covariates. The parameter values are displayed using log_e.

5.4 Interval of Maximum Slope

The maximum rate of increase of viral load is used as a proxy for the maximum replicative fitness of the virus in-vivo. To estimate the maximum rate of increase without fitting a model, we find the interval between samples with the maximum slope [Ribeiro et al., 2010]. Finding the interval of maximum slope is a much simpler method than fitting a mixed effects model based on a complex within-host model. A difficulty with this simpler method is that the sampling rate of the primates' viral load has a strong effect on the detected maximum slope.

To quantify the effect of the sampling rate on the maximum slope found, we simulated primate data. Twenty primate acute infections, viral load and CD4+ levels, were simulated using model 5.5 with the initial parameters set from best fit parameters from primate RM3311 found via maximum likelihood methods implemented in MATLAB. The initial parameters for viral load and activated T-cells were drawn from a log-normal distribution centered on the respective best-fit parameters.

$$\begin{aligned}
\frac{\mathrm{d}z}{\mathrm{d}t} &= \lambda - d_x z - \alpha x u \\
\frac{\mathrm{d}x}{\mathrm{d}t} &= \alpha x u - d_x x - \beta x v \\
\frac{\mathrm{d}y}{\mathrm{d}t} &= \beta x v - d_y y - r_y y u \\
\frac{\mathrm{d}v}{\mathrm{d}t} &= p_0 y - r_v u v - c_v v \\
\frac{\mathrm{d}u}{\mathrm{d}t} &= a_v v - d_u u.
\end{aligned} \tag{5.5}$$

We find the length of the sampling interval had a strong effect on the calculated maximum slope as demonstrated in Figure 5.10 that shows the results of changes to the sampling interval. Given the strong effect of sampling interval, only primates with similar sampling intervals can be meaningfully compared.

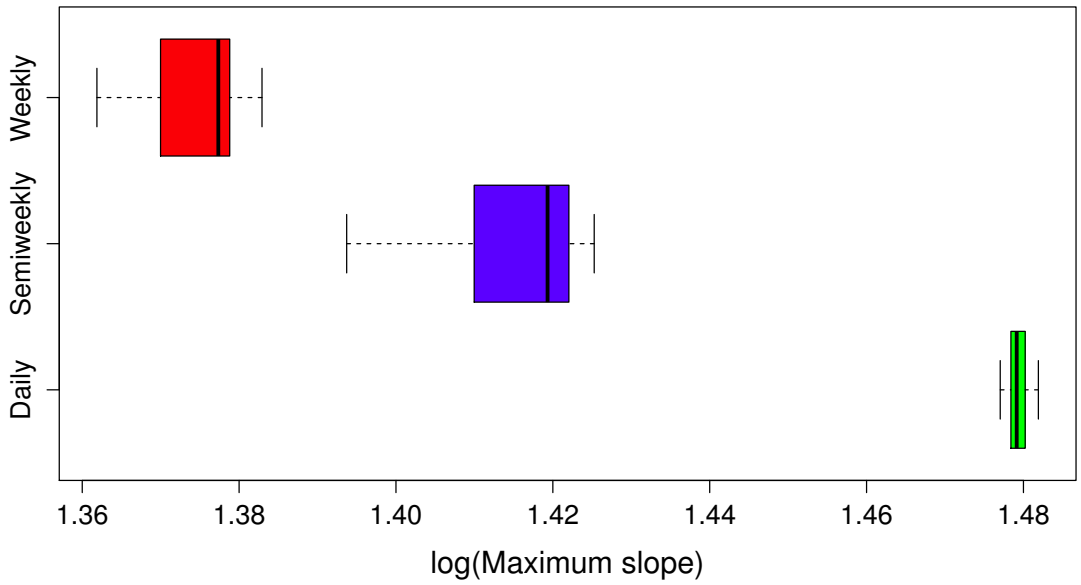


Figure 5.10: The maximum slope estimates calculated from 20 simulated infections using (5.5) model with best fit parameters from primate RM33113. The initial conditions for viral load and activated T-cells were drawn from log-normal distributions around the best-fit parameters' means.

5.4.1 Results

The interval of maximum slope was found for all primates. See Figure 5.10 for primates with a semiweekly sampling rate and Appendix C.5 for all primates.

The length of the sampling intervals varied for each primate and among primates. Primates were sorted by the length of the sampling interval with the maximum slope. Considering primates with near semiweekly sampling rates, as seen in Figure 5.12, we can see some evidence of differences in the interval of maximum slope. Primates infected with SHIV 1157, subtype C, show a higher replicative fitness by this method. This result disagrees with the results for β found in the previous section. This result is likely partially a result of the interval of maximum slope approximating both the results of replicative fitness, β, and viral burst size, p_0, which we found in the previous section to affect within-host viral fitness in opposite directions.

This result is not repeated when primates with a near weekly sampling interval are considered where the larger number of primates do not display a significant difference.

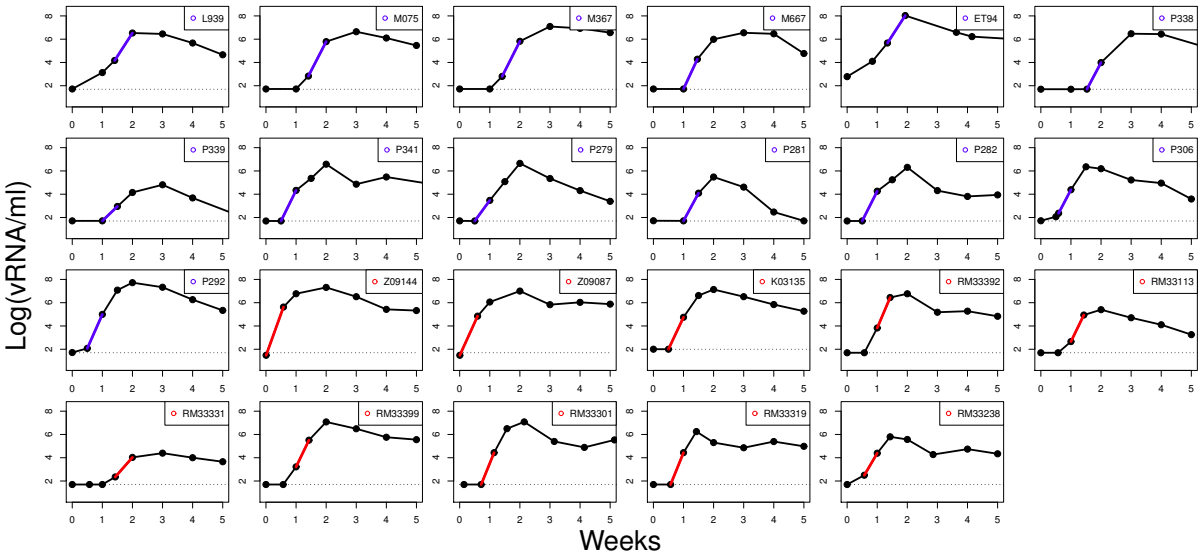


Figure 5.11: Time series of viral load for primates that were sampled semiweekly with the interval of maximum slope highlighted in blue for primates infected with SHIV SF162 (Subtype B) and highlighted in red for primates infected with SHIV 1157 (Subtype C). The detection limit of the viral load assay used is marked with a dotted line.

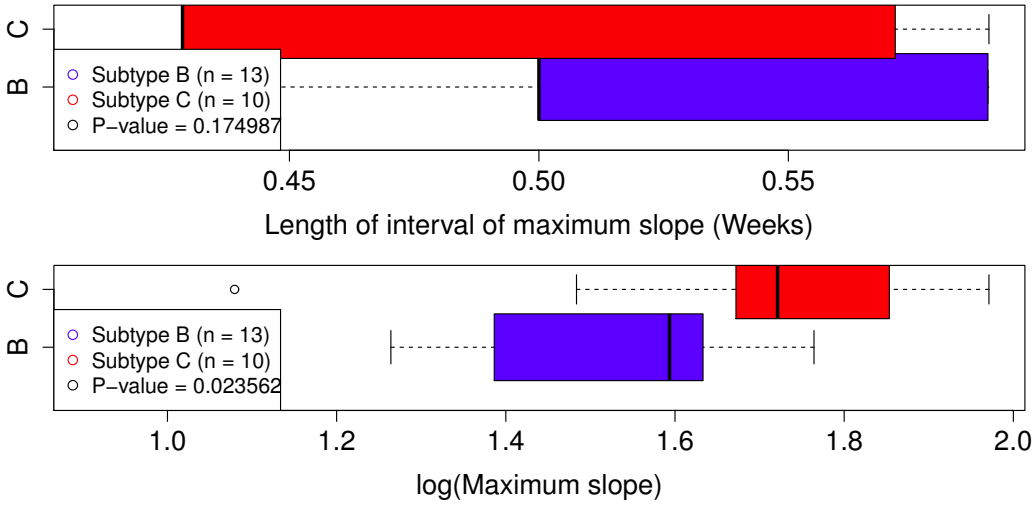


Figure 5.12: Boxplots for primates with near semiweekly sampling rates grouped by SHIV subtype, SHIV SF162 (Subtype B) in blue and SHIV 1157 (Subtype C) in red. The Wilcoxon rank-sum test for equal means yields a p-value of 0.0236.

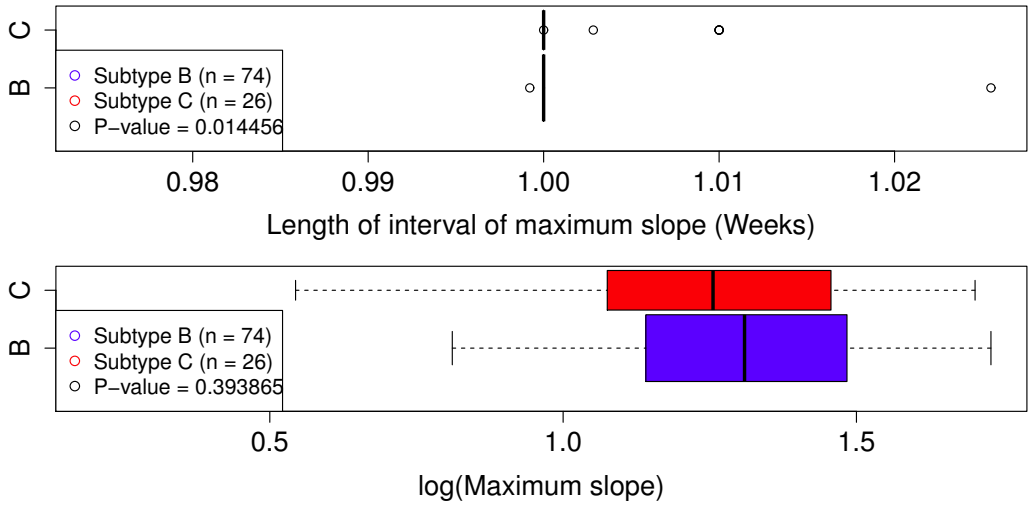


Figure 5.13: Boxplots for primates with near weekly sampling rates grouped by SHIV subtype, SHIV SF162 (Subtype B) in blue and SHIV 1157 (Subtype C) in red. The Wilcoxon rank-sum test for equal means yields a p-value of 0.3938.

	Intravaginal	Intravenous	Unknown
Intrarectal	0.493	0.185	0.007
Intravaginal	*	0.175	0.047
Intravenous	0.175	*	0.012

Table 5.6: P-value for unequal means of the interval of maximum slope of the different challenge methods using the Wilcoxon rank sum test comparing all primates.

Tables 5.6 and 5.7 show the p-value for unequal means of the interval of maximum slope between different challenge methods using the Wilcoxon rank sum test comparing all primates and primates with a weekly sampling rate respectively. Focusing on the weekly sampling rate, there is some evidence that challenge method affects the maximum slope found. This agrees with our intuition that different challenge methods will introduce varying numbers of virus particles to different initial host cell populations. This also hints at the importance of transmission method in the evolution of HIV. This difference in the replicative fitness related to challenge method found with the interval of maximum slope method presumably confounds detection of replicative fitness and indicates that mixed effects models are required to account for these effects.

	Intravaginal	Intravenous
Intrarectal	0.054	0.020
Intravaginal	*	0.63754644

Table 5.7: P-value for unequal means of the interval of maximum slope between different challenge methods using the Wilcoxon rank sum test comparing primates with a weekly sampling rate.

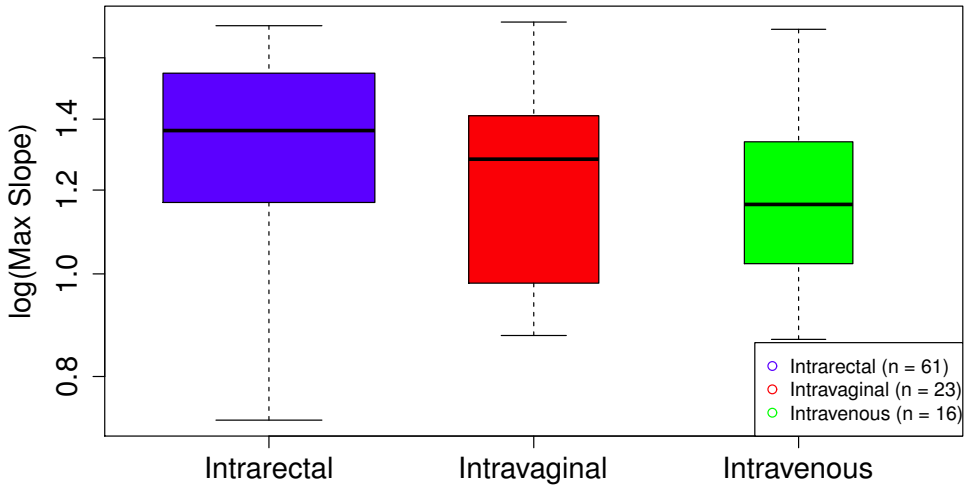


Figure 5.14: Boxplot of the log slope of the interval of maximum slope for each primate with a sampling rate near weekly grouped by challenge method.

5.5 Conclusions

Non-linear mixed-effects models offer a powerful tool allowing more complex models of within-host dynamics to be fit while efficiently utilizing data and allowing the application of biological knowledge of the within-host system. A major obstacle to estimating best-fit parameters for models of viral dynamics is the difficulty in parameter identifiability. Also incomplete understanding of the interplay between model parameters could lead to misidentifying the underlying cause for differences in best-fit parameters between groups.

Using our large compiled data-set we were able to identify differences in replicative fitness and burst size between SHIV subtypes, with lower replicative fitness but higher viral burst size in subtype C when compared to subtype B. A re-examination of the *in vitro* results considering this particular interplay between replicative fitness and burst size could help further elucidate the relationship between the *in vivo* and *in vitro* results.

We were also able to speculate that differences in immune response between primates lead to the outlying higher viral loads seen in some rhesus macaques infected with SHIV SF162. This difference between primates could be an important confounding variable and should be a controlled covariate, that was unavailable in our compiled dataset.

Chapter 6

Summary and Future Work

Using a combination of simulation, modelling, and statistical methods we identified evolutionary pressures on the life history traits of HIV-1 and the drivers of the observed difference in evolutionary rates within- and between-hosts. The extremely severe and competitive transmission bottleneck is an important factor in slower between-host evolution of HIV-1, relative to the rapid evolution of HIV-1 within-hosts, and subtype B and C have different relative fitness advantages. We speculate that the lower replicative fitness of HIV-1 subtype C drives the results of *in vitro* fitness experiments that find HIV-1 subtypes A, B, and D have a fitness advantage over subtype C. In addition, we suggest that *in vivo*, the importance of the extreme transmission bottleneck may increase the impact of viral burst size and that this, along with viral traits related to transmission fitness, could be the fitness advantages that drive the observed global expansion of HIV-1 subtype C.

To capture the effect of different relative within-host fitness advantages, details need to be added to our multi-scale model, yet the diversity and/or number of founding strains of HIV are still controversial. In future modelling work, we would like to include the possibility of transmitting more than one virion from the bottleneck inoculum. The increased number of virions transmitted would immediately begin a within-host fitness competition. We hypothesize that with the inclusion of a larger number of transmitted virions, those that are more fit, within-

host, would quickly establish the new infection and win the within-host competition. This might result in much faster between-host evolution or increase the importance of archiving virus with a higher transmission fitness but lower within-host fitness. The inclusion of a larger transmission bottleneck might elevate the importance of the latent reservoir.

We would also like to expand our multi-scale model to consider a non-homogeneous host population. Individuals have different human leukocyte antigen alleles and a fitness advantage in one host could be a disadvantage in another. We can consider this possibility by including two or more different types of hosts in our simulation, each with a distinct within-host fitness landscape.

Additionally we plan to incorporate incomplete drug treatment into the model to examine the effect of multiple activations from the latent reservoir on the course of evolution over the epidemic. Is there an increased risk of drug resistance seeding the reservoir and being transmitted, or does a stable latent reservoir and the transmission fitness costs of within-host evolution slow the transmission of drug resistance?

The result that subtype C has a lower replicative fitness but higher burst size when compared to subtype B raises questions that could possibly be addressed with additional *in vitro* research; for example, experimental environments could be devised to detect the increased fitness advantage conferred by a higher burst size. Further results from the literature regarding the relative fitness advantages of different life history traits in viral pathogens can also be leveraged to help understand HIV-1.

The work in this thesis highlights the advantage of using multiple tools, illustrating how modelling, simulation, and statistics can be combined to address complex problems. Careful analysis of the within-host models and the information available in the Uganda and Zimbabwe data-set uncovered the infeasibility of detecting replicative fitness differences from the data. The multi-scale model of HIV-1 evolution, combining deterministic within-host modelling with a stochastic bottleneck, highlighted the relative importance of the extreme transmission bottleneck. A model of within-host dynamics combined with a large, rich data-set, analysed with

a non-linear mixed effects model, uncovered the potential fitness differences between HIV-1 subtypes B and C. It is our hope that the use of multi-faceted, interdisciplinary approaches will continue to propel research forward in this critical field of pathogen evolution.

Bibliography

[Abraha et al., 2009] Abraha, A., Nankya, I. L., Gibson, R., Demers, K., Tebit, D. M., Johnston, E., Katzenstein, D., Siddiqui, A., Herrera, C., Fischetti, L., Shattock, R. J., and Arts, E. J. (2009). CCR5- and CXCR4-tropic subtype C human immunodeficiency virus type 1 isolates have a lower level of pathogenic fitness than other dominant group M subtypes: implications for the epidemic. Journal of virology, 83(11):5592–605.

[Abrahams et al., 2019] Abrahams, M.-R., Joseph, S. B., Garrett, N., Tyers, L., Moeser, M., Archin, N., Council, O. D., Matten, D., Zhou, S., Doolabh, D., Anthony, C., Goonetilleke, N., Karim, S. A., Margolis, D. M., Pond, S. K., Williamson, C., and Swanstrom, R. (2019). The Replication-Competent HIV-1 Latent Reservoir is Primarily Established Near the Time of Therapy Initiation. bioRxiv, page 512475.

[Alexander, 2016] Alexander, T. S. (2016). Human Immunodeficiency Virus Diagnostic Testing: 30 Years of Evolution. Clinical and Vaccine Immunology, 23(4):249–253.

[Ariën et al., 2005] Ariën, K. K., Abraha, A., Quiñones-mateu, M. E., Vanham, G., Arts, E. J., and Kestens, L. (2005). The Replicative Fitness of Primary Human Immunodeficiency Virus Type 1 (HIV-1) Group The Replicative Fitness of Primary Human Immunodeficiency Virus Type 1 (HIV-1) Group M , HIV-1 Group O , and HIV-2 Isolates †. Journal of virology, 79(14):8979–8990.

[Ball et al., 2003] Ball, S. C., Abraha, A., Collins, K. R., Andre, J., Baird, H., Quiñones-mateu, M. E., Murray, M., Richard, N., Zimmerman, P. a., Kawamura, T., Blauvelt, A.,

Arts, E. J., Marozsan, A. J., Quin, M. E., Penn-nicholson, A., and Lobritz, M. (2003). Comparing the Ex Vivo Fitness of CCR5-Tropic Human Immunodeficiency Virus Type 1 Isolates of Subtypes B and C Comparing the Ex Vivo Fitness of CCR5-Tropic Human Immunodeficiency Virus Type 1 Isolates of Subtypes B and C. Journal of virology, 77(2):1021–1038.

[Barre-Sinoussi et al., 2013] Barre-Sinoussi, F., Ross, A. L., and Delfraissy, J. F. (2013). Past, present and future: 30 years of HIV research. Nat Rev Microbiol, 11(12):877–883.

[Bbosa et al., 2019] Bbosa, N., Kaleebu, P., and Ssemwanga, D. (2019). HIV subtype diversity worldwide. Current Opinion in HIV and AIDS, 14(3):153–160.

[Bertels et al., 2018] Bertels, F., Marzel, A., Leventhal, G., Mitov, V., Fellay, J., Günthard, H. F., Böni, J., Yerly, S., Klimkait, T., Aubert, V., Battegay, M., Rauch, A., Cavassini, M., Calmy, A., Bernasconi, E., Schmid, P., Scherrer, A. U., Müller, V., Bonhoeffer, S., Kouyos, R., and Regoes, R. R. (2018). Dissecting HIV Virulence: Heritability of Setpoint Viral Load, CD4+ T-Cell Decline, and Per-Parasite Pathogenicity. Molecular Biology and Evolution, 35(1):27–37.

[Bertels et al., 2017] Bertels, F., Marzel, A., Leventhal, G., Mitov, V., Huldrych, F. G., Yerly, S., Klimkait, T., Battegay, M., Rauch, A., Cavassini, M., and Calmy, A. (2017). Dissecting HIV Virulence: Heritability of Setpoint Viral Load, CD4+ T Cell Decline and Per-Parasite Pathogenicity. pages 1–21.

[Blanquart et al., 2017] Blanquart, F., Wymant, C., Cornelissen, M., Gall, A., Bakker, M., Bezemer, D., Hall, M., Hillebregt, M., Ong, S. H., Albert, J., Bannert, N., Fellay, J., Fransen, K., Gourlay, A. J., Grabowski, M. K., Gunsenheimer-Bartmeyer, B., Günthard, H. F., Kivelä, P., Kouyos, R., Laeyendecker, O., Liitsola, K., Meyer, L., Porter, K., Ristola, M., van Sighem, A., Vanham, G., Berkhout, B., Kellam, P., Reiss, P., Fraser, C., Collaboration, B., Gourlay, J., Grabowski, M. K., Gunsenheimer-Bartmeyer, B., Meyer, L., Porter, K., Ristola, M., Sighem, A. V., Vanham, G., Berkhout, B., Kellam, P., Reiss, P., and Fraser,

C. (2017). Viral genetic variation accounts for a third of variability in HIV-1 set-point viral load in Europe. PLOS Biology, 15(6):1–26.

[Boritz et al., 2016] Boritz, E. A., Darko, S., Swaszek, L., Wolf, G., Wells, D., Wu, X., Henry, A. R., Laboune, F., Hu, J., Ambrozak, D., Hughes, M. S., Hoh, R., Casazza, J. P., Vostal, A., Bunis, D., Nganou-Makamdop, K., Lee, J. S., Migueles, S. A., Koup, R. A., Connors, M., Moir, S., Schacker, T., Maldarelli, F., Hughes, S. H., Deeks, S. G., and Douek, D. C. (2016). Multiple Origins of Virus Persistence during Natural Control of HIV Infection. Cell, 166(4).

[Carlson et al., 2016] Carlson, J. M., Du, V. Y., Pfeifer, N., Bansal, A., Tan, V. Y. F., Power, K., Brumme, C. J., Kreimer, A., DeZiel, C. E., Fusi, N., Schaefer, M., Brockman, M. A., Gilmour, J., Price, M. A., Kilembe, W., Haubrich, R., John, M., Mallal, S., Shapiro, R., Frater, J., Harrigan, P. R., Ndung'u, T., Allen, S., Heckerman, D., Sidney, J., Allen, T. M., Goulder, P. J. R., Brumme, Z. L., Hunter, E., and Goepfert, P. A. (2016). Impact of pre-adapted HIV transmission. Nature medicine, 22(6):606–613.

[Carlson et al., 2014a] Carlson, J. M., Schaefer, M., Monaco, D. C., Batorsky, R., Claiborne, D. T., Prince, J., Deymier, M. J., Ende, Z. S., Klatt, N. R., DeZiel, C. E., Lin, T.-H., Peng, J., Seese, A. M., Shapiro, R., Frater, J., Ndung'u, T., Tang, J., Goepfert, P., Gilmour, J., Price, M. A., Kilembe, W., Heckerman, D., Goulder, P. J. R., Allen, T. M., Allen, S., and Hunter, E. (2014a). HIV transmission. Selection bias at the heterosexual HIV-1 transmission bottleneck. Science (New York, N.Y.), 345(6193):1254031.

[Carlson et al., 2014b] Carlson, J. M., Schaefer, M., Monaco, D. C., Batorsky, R., Claiborne, D. T., Prince, J., Deymier, M. J., Ende, Z. S., Klatt, N. R., DeZiel, C. E., Lin, T. H., Peng, J., Seese, A. M., Shapiro, R., Frater, J., Ndung'u, T., Tang, J., Goepfert, P., Gilmour, J., Price, M. A., Kilembe, W., Heckerman, D., Goulder, P. J., Allen, T. M., Allen, S., and Hunter, E. (2014b). Selection bias at the heterosexual HIV-1 transmission bottleneck. Science, 345(6193).

[Champredon et al., 2013] Champredon, D., Bellan, S., and Dushoff, J. (2013). HIV Sexual Transmission Is Predominantly Driven by Single Individuals Rather than Discordant Couples: A Model-Based Approach. PLoS ONE, 8(12):e82906.

[Chen et al., 2018] Chen, J., Ren, Y., Daharsh, L., Liu, L., Kang, G., Li, Q., Wei, Q., Wan, Y., and Xu, J. (2018). Identification of unequally represented founder viruses among tissues in very early SIV rectal transmission. Frontiers in Microbiology, 9(MAR).

[Conway et al., 2019] Conway, J. M., Perelson, A. S., and Li, J. Z. (2019). Predictions of time to HIV viral rebound following ART suspension that incorporate personal biomarkers. PLOS Computational Biology, 15(7):e1007229.

[Cuevas and Sanjuán, 2015] Cuevas, Geller, G. L.-A. J. and Sanjuán, R. (2015). Extremely High Mutation Rate of HIV-1 In Vivo. PLoS Biology, 13(9):1–19.

[Davenport et al., 2019] Davenport, M. P., Khoury, D. S., Cromer, D., Lewin, S. R., Kelleher, A. D., and Kent, S. J. (2019). Functional cure of HIV: the scale of the challenge. Nature Reviews Immunology, 19(1):45–54.

[Descours et al., 2017] Descours, B., Petitjean, G., López-Zaragoza, J.-L., Bruel, T., Raffel, R., Psomas, C., Reynes, J., Lacabaratz, C., Levy, Y., Schwartz, O., Lelievre, J. D., and Benkirane, M. (2017). CD32a is a marker of a CD4 T-cell HIV reservoir harbouring replication-competent proviruses. Nature.

[Deymier et al., 2015] Deymier, M. J., Ende, Z., Fenton-May, A. E., Dilernia, D. A., Kilembe, W., Allen, S. A., Borrow, P., and Hunter, E. (2015). Heterosexual Transmission of Subtype C HIV-1 Selects Consensus-Like Variants without Increased Replicative Capacity or Interferon-α Resistance. PLoS Pathogens, 11(9):1–22.

[Doekes et al., 2017] Doekes, H. M., Fraser, C., and Lythgoe, K. A. (2017). Effect of the Latent Reservoir on the Evolution of HIV at the Within- and Between-Host Levels. PLOS Computational Biology, 13(1):e1005228.

[Eigen, 1971] Eigen, M. (1971). Selforganization of Matter and the Evolution of Biological Macromolecules.

[Eugene et al., 2013] Eugene, H. S., Pierce-Paul, B. R., Cragio, J. K., and Ross, T. M. (2013). Rhesus macaques vaccinated with consensus envelopes elicit partially protective immune responses against SHIV SF162p4 challenge. Virology journal, 10(1):102.

[Eyre-Walker and Keightley, 2007] Eyre-Walker, A. and Keightley, P. D. (2007). The distribution of fitness effects of new mutations. Nature Reviews Genetics, 8(8):610–618.

[Eyre-Walker et al., 2006] Eyre-Walker, A., Woolfit, M., and Phelps, T. (2006). The distribution of fitness effects of new deleterious amino acid mutations in humans. Genetics, 173(2):891–900.

[Feder et al., 2017] Feder, A. A. F., Kline, C., Polacino, P., Cottrell, M., Angela, D. M., Keele, B. F., Hu, S.-L., Petrov, D. A., Pennings, P. S., and Ambrose, Z. (2017). High resolution spatio-temporal assessment of simian / human immunodeficiency virus (SHIV) evolution reveals a highly dynamic process within the host. bioRxiv, 13(5):1–31.

[Fraser et al., 2007] Fraser, C., Hollingsworth, T. D., Chapman, R., de Wolf, F., and Hanage, W. P. (2007). Variation in HIV-1 set-point viral load: epidemiological analysis and an evolutionary hypothesis. Proceedings of the National Academy of Sciences of the United States of America, 104(44):17441–6.

[Fraser et al., 2014] Fraser, C., Lythgoe, K., Leventhal, G. E., Shirreff, G., Hollingsworth, T. D., Alizon, S., and Bonhoeffer, S. (2014). Virulence and pathogenesis of HIV-1 infection: an evolutionary perspective. Science (New York, N.Y.), 343(6177):1243727.

[García et al., 2010] García, A., Siddappa, N. B., Li, Q., Haase, A. T., Paul, K., Stroud, F., Zhang, X., Fountain, J. A., Villinger, F., Novembre, F. J., Else, J. G., Evan Secor, W., and Ruprecht, R. M. (2010). ORIGINAL ARTICLE: AIDS and optic neuritis in a rhesus

monkey infected with the R5 clade C SHIV-1157ipd3N4. Journal of Medical Primatology, 39(5):356–360.

[Gianella et al., 2016] Gianella, S., Kosakovsky Pond, S. L., Oliveira, M. F., Scheffler, K., Strain, M. C., De la Torre, A., Letendre, S., Smith, D. M., and Ellis, R. J. (2016). Compartmentalized HIV rebound in the central nervous system after interruption of antiretroviral therapy. Virus Evolution, 2(2):vew020.

[Haaland et al., 2009] Haaland, R. E., Hawkins, P. A., Salazar-Gonzalez, J., Johnson, A., Tichacek, A., Karita, E., Manigart, O., Mulenga, J., Keele, B. F., Shaw, G. M., Hahn, B. H., Allen, S. A., Derdeyn, C. A., and Hunter, E. (2009). Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. PLoS Pathogens, 5(1).

[Harbison et al., 2014] Harbison, C., Zhuang, K., Gettie, A., Blanchard, J., Knight, H., Didier, P., Cheng-Mayer, C., and Westmoreland, S. (2014). Giant cell encephalitis and microglial infection with mucosally transmitted simian-human immunodeficiency virus SHIVSF162P3N in rhesus macaques. Journal of NeuroVirology, 20(1):62–72.

[Immonen et al., 2012] Immonen, T., Gibson, R., Leitner, T., Miller, M. A., Arts, E. J., Somersalo, E., and Calvetti, D. (2012). A hybrid stochastic-deterministic computational model accurately describes spatial dynamics and virus diffusion in HIV-1 growth competition assay. Journal of Theoretical Biology, 312:120–132.

[Immonen et al., 2015] Immonen, T. T., Conway, J. M., Romero-Severson, E. O., Perelson, A. S., and Leitner, T. (2015). Recombination Enhances HIV-1 Envelope Diversity by Facilitating the Survival of Latent Genomic Fragments in the Plasma Virus Population. PLoS Computational Biology, 11(12):1–26.

[Immonen and Leitner, 2014] Immonen, T. T. and Leitner, T. (2014). Reduced evolutionary rates in HIV-1 reveal extensive latency periods among replicating lineages. Retrovirology, 81(11):1—-11.

[Joseph et al., 2015] Joseph, S. B., Swanstrom, R., Kashuba, A. D. M., and Cohen, M. S. (2015). Bottlenecks in HIV-1 transmission: insights from the study of founder viruses. Nature reviews. Microbiology, 13(7):414–425.

[Kariuki et al., 2017] Kariuki, S. M., Selhorst, P., Ariën, K. K., and Dorfman, J. R. (2017). The HIV-1 transmission bottleneck.

[Karlsson and Mittler, 2000] Karlsson, A. C. and Mittler, J. (2000). Stefan Lindba Anders So Primary HIV Infection Study Group. (May).

[Keele et al., 2008] Keele, B. F., Giorgi, E. E., Salazar-Gonzalez, J. F., Decker, J. M., Pham, K. T., Salazar, M. G., Sun, C., Grayson, T., Wang, S., Li, H., Wei, X., Jiang, C., Kirchherr, J. L., Gao, F., Anderson, J. A., Ping, L.-H., Swanstrom, R., Tomaras, G. D., Blattner, W. A., Goepfert, P. A., Kilby, J. M., Saag, M. S., Delwart, E. L., Busch, M. P., Cohen, M. S., Montefiori, D. C., Haynes, B. F., Gaschen, B., Athreya, G. S., Lee, H. Y., Wood, N., Seoighe, C., Perelson, A. S., Bhattacharya, T., Korber, B. T., Hahn, B. H., and Shaw, G. M. (2008). Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proceedings of the National Academy of Sciences of the United States of America, 105(21):7552–7.

[Konrad et al., 2016] Konrad, B. P., Taylor, D., Conway, J. M., Ogilvie, G. S., and Coombs, D. (2016). On the duration of the period between exposure to HIV and detectable infection. Epidemics.

[Kuo and Lichterfeld, 2018] Kuo, H.-h. and Lichterfeld, M. (2018). Recent progress in understanding hiv reservoirs. Curr Opin HIV AIDS., pages 137–142.

[Kyeyune et al., 2013] Kyeyune, F., Nankya, I., Metha, S., Akao, J., Ndashimye, E., Tebit, D. M., Rodriguez, B., Kityo, C., Salata, R. a., Mugyenyi, P., and Arts, E. (2013). Treatment failure and drug resistance is more frequent in HIV-1 subtype D versus subtype A-infected Ugandans over a 10-year study period. AIDS (London, England), 27(12):1899–909.

[Lakhashe et al., 2014] Lakhashe, S. K., Byrareddy, S. N., Zhou, M., Bachler, B. C., Hemashettar, G., Hu, S. L., Villinger, F., Else, J. G., Stock, S., Lee, S. J., Vargas-Inchaustegui, D. A., Cofano, E. B., Robert-Guroff, M., Johnson, W. E., Polonis, V. R., Forthal, D. N., Loret, E. P., Rasmussen, R. A., and Ruprecht, R. M. (2014). Multimodality vaccination against clade C SHIV: Partial protection against mucosal challenges with a heterologous tier 2 virus. Vaccine, 32(48):6527–6536.

[Lederman et al., 2015] Lederman, M. M., Veazey, R. S., Offord, R., Mosier, D. E., Mefford, M., Jr, M. P., Lifson, J. D., Salkowitz, J. R., Blauvelt, A., and Hartley, O. (2015). No Title. 306(5695):485–487.

[Li et al., 2010] Li, H., Bar, K. J., Wang, S., Decker, J. M., Chen, Y., Sun, C., Salazar-Gonzalez, J. F., Salazar, M. G., Learn, G. H., Morgan, C. J., Schumacher, J. E., Hraber, P., Giorgi, E. E., Bhattacharya, T., Korber, B. T., Perelson, A. S., Eron, J. J., Cohen, M. S., Hicks, C. B., Haynes, B. F., Markowitz, M., Keele, B. F., Hahn, B. H., and Shaw, G. M. (2010). High multiplicity infection by HIV-1 in men who have sex with men. PLoS Pathogens, 6(5):1–17.

[Lythgoe and Fraser, 2012] Lythgoe, K. A. and Fraser, C. (2012). New insights into the evolutionary rate of HIV-1 at the within-host and epidemiological levels. Proceedings. Biological sciences / The Royal Society, 279(1741):3367–3375.

[Lythgoe et al., 2017] Lythgoe, K. A., Gardner, A., Pybus, O. G., and Grove, J. (2017). Short-Sighted Virus Evolution and a Germline Hypothesis for Chronic Viral Infections. Trends in Microbiology, 25(5):336–348.

[Lythgoe et al., 2013] Lythgoe, K. A., Pellis, L., and Fraser, C. (2013). Is Hiv Short-Sighted? Insights From A Multistrain Nested Model. Evolution, 67(10):2769–2782.

[Marozsan et al., 2005] Marozsan, A. J., Moore, D. M., Lobritz, M. A., Fraundorf, E., Abraha, A., Reeves, J. D., and Arts, E. J. (2005). Differences in the Fitness of Two Diverse Wild-Type Human Immunodeficiency Virus Type 1 Isolates Are Related to the Efficiency of Cell Binding and Entry. Journal of Virology, 79(11):7121–7134.

[Morrison et al., 2010] Morrison, C. S., Demers, K., Kwok, C., Bulime, S., Rinaldi, A., Munjoma, M., Dunbar, M., Chipato, T., Byamugisha, J., Van Der Pol, B., Arts, E., and Salata, R. A. (2010). Plasma and cervical viral loads among Ugandan and Zimbabwean women during acute and early HIV-1 infection. AIDS (London, England), 24(4):573–82.

[Morrison et al., 2007] Morrison, C. S., Richardson, B. A., Mmiro, F., Chipato, T., Celentano, D. D., Luoto, J., Mugerwa, R., Padian, N., Rugpao, S., Brown, J. M., Cornelisse, P., and Salata, R. A. (2007). Hormonal contraception and the risk of HIV acquisition. Aids, 21(1):85–95.

[Mumbauer et al., 2013] Mumbauer, A., Gettie, A., Blanchard, J., and Cheng-Mayer, C. (2013). Efficient Mucosal Transmissibility but Limited Pathogenicity of R5 SHIVSF162P3N in Chinese-Origin Rhesus Macaques. JAIDS Journal of Acquired Immune Deficiency Syndromes, 62(5):496–504.

[Nowak and May, 2000] Nowak, M. and May, R. (2000). Virus Dynamics: Mathematical Principles of Immunology and Virology. OUP Oxford.

[Olabode et al., 2019] Olabode, A. S., Avino, M., Ng, G. T., Abu-Sardanah, F., Dick, D. W., and Poon, A. F. Y. (2019). Evidence for a recombinant origin of HIV-1 Group M from genomic variation. Virus Evolution, 5(1).

[Park and Bolker, 2017] Park, S. W. and Bolker, B. M. (2017). Effects of contact structure on the transient evolution of HIV virulence. PLOS Computational Biology, 13(3):e1005453.

[Payne et al., 2014] Payne, R., Muenchhoff, M., Mann, J., Roberts, H. E., Matthews, P., Adland, E., Hempenstall, A., Huang, K.-H., Brockman, M., Brumme, Z., Sinclair, M., Miura, T., Frater, J., Essex, M., Shapiro, R., Walker, B. D., Ndung'u, T., McLean, A. R., Carlson, J. M., and Goulder, P. J. R. (2014). Impact of HLA-driven HIV adaptation on virulence in populations of high HIV seroprevalence. Proceedings of the National Academy of Sciences of the United States of America, 111(50):E5393–400.

[Peterson et al., 2014] Peterson, C. W., Younan, P., Polacino, P. S., Maurice, N. J., Miller, H. W., Prlic, M., Jerome, K. R., Woolfrey, A. E., and Hu, S.-l. (2014). infection. 42(5):237–246.

[Ping et al., 2013] Ping, L.-H., Joseph, S. B., Anderson, J. A., Abrahams, M.-R., Salazar-Gonzalez, J. F., Kincer, L. P., Treurnicht, F. K., Arney, L., Ojeda, S., Zhang, M., Keys, J., Potter, E. L., Chu, H., Moore, P., Salazar, M. G., Iyer, S., Jabara, C., Kirchherr, J., Mapanje, C., Ngandu, N., Seoighe, C., Hoffman, I., Gao, F., Tang, Y., Labranche, C., Lee, B., Saville, A., Vermeulen, M., Fiscus, S., Morris, L., Karim, S. A., Haynes, B. F., Shaw, G. M., Korber, B. T., Hahn, B. H., Cohen, M. S., Montefiori, D., Williamson, C., and Swanstrom, R. (2013). Comparison of Viral Env Proteins from Acute and Chronic Infections with Subtype C Human Immunodeficiency Virus Type 1 Identifies Differences in Glycosylation and CCR5 Utilization and Suggests a New Strategy for Immunogen Design. Journal of Virology, 87(13):7218–7233.

[Polacino et al., 2008] Polacino, P., Larsen, K., Galmin, L., Suschak, J., Kraft, Z., Stamatatos, L., Anderson, D., Barnett, S. W., Pal, R., Bost, K., Bandivdekar, A., Miller, C. J., and Hu, S.-L. (2008). Differential pathogenicity of SHIV SF162 P4 infection in pig-tailed and rhesus macaques. Journal of Medical Primatology, 37:13–23.

[Poon et al., 2018] Poon, A. F. Y., Prodger, J. L., Lynch, B. A., Lai, J., Reynolds, S. J., Kasule, J., Capoferri, A. A., Lamers, S. L., Rodriguez, C. W., Bruno, D., Porcella, S. F., Martens,

C., Quinn, T. C., and Redd, A. D. (2018). Quantitation of the latent HIV-1 reservoir from the sequence diversity in viral outgrowth assays. Retrovirology, 15(1):47.

[Redd et al., 2012] Redd, A. D., Collinson-Streng, A. N., Chatziandreou, N., Mullis, C. E., Laeyendecker, O., Martens, C., Ricklefs, S., Kiwanuka, N., Nyein, P. H., Lutalo, T., Grabowski, M. K., Kong, X., Manucci, J., Sewankambo, N., Wawer, M. J., Gray, R. H., Porcella, S. F., Fauci, A. S., Sagar, M., Serwadda, D., and Quinn, T. C. (2012). Previously transmitted HIV-1 strains are preferentially selected during subsequent sexual transmissions. Journal of Infectious Diseases, 206(9):1433–1442.

[Ribeiro et al., 2010] Ribeiro, R. M., Qin, L., Chavez, L. L., Li, D., Self, S. G., and Perelson, A. S. (2010). Estimation of the initial viral growth rate and basic reproductive number during acute HIV-1 infection. Journal of virology, 84(12):6096–6102.

[Robinson et al., 2010] Robinson, J. E., Franco, K., Elliott, D. H., Maher, M. J., Reyna, A., Montefiori, D. C., Zolla-Pazner, S., Gorny, M. K., Kraft, Z., and Stamatatos, L. (2010). Quaternary epitope specificities of anti-HIV-1 neutralizing antibodies generated in rhesus macaques infected by the simian/human immunodeficiency virus SHIVSF162P4. Journal of virology, 84(7):3443–53.

[Sengupta and Siliciano, 2018] Sengupta, S. and Siliciano, R. F. (2018). Targeting the Latent Reservoir for HIV-1. Immunity, 48(5):872–895.

[Shaw and Hunter, 2012] Shaw, G. M. and Hunter, E. (2012). HIV transmission. Cold Spring Harbor Perspectives in Medicine, 2(11):1–23.

[Shirreff et al., 2011] Shirreff, G., Pellis, L., Laeyendecker, O., and Fraser, C. (2011). Transmission selects for HIV-1 strains of intermediate virulence: A modelling approach. PLoS Computational Biology, 7(10).

[Sholukh et al., 2015] Sholukh, A. M., Watkins, J. D., Vyas, H. K., Gupta, S., Lakhashe, S. K., Thorat, S., Zhou, M., Hemashettar, G., Bachler, B. C., Forthal, D. N., Villinger, F., Satten-

tau, Q. J., Weiss, R. A., Agatic, G., Corti, D., Lanzavecchia, A., Heeney, J. L., and Ruprecht, R. M. (2015). Defense-in-depth by mucosally administered anti-HIV dimeric IgA2 and systemic IgG1 mAbs: Complete protection of rhesus monkeys from mucosal SHIV challenge. Vaccine, 33(17):2086–2095.

[Stafford et al., 2000] Stafford, M. A., Coreya, L., Caob, Y., Daar " ", E. S., Hob, D. D., and Perelson, A. S. (2000). Modeling Plasma Virus Concentration during Primary HIV Infection. J. theor. Biol, 203:285–301.

[Sui et al., 2014] Sui, Y., Gordon, S., Franchini, G., and Berzofsky, J. A. (2014). Non-human primate models for HIV/AIDS vaccine development, volume 102.

[Tebit and Arts, 2011] Tebit, D. M. and Arts, E. J. (2011). Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. The Lancet Infectious Diseases, 11(1):45–56.

[Theys et al., 2018] Theys, K., Libin, P., Pineda-Peña, A. C., Nowé, A., Vandamme, A. M., and Abecasis, A. B. (2018). The impact of HIV-1 within-host evolution on transmission dynamics. Current Opinion in Virology, 28:92–101.

[Tully et al., 2016] Tully, D. C., Ogilvie, C. B., Batorsky, R. E., Bean, D. J., Power, A., Ghebremichael, M., Bedard, H. E., Gladden, A. D., Seese, M., Amero, M. A., Lane, K., Mcgrath, G., Bazner, S. B., Tinsley, J., Lennon, N. J., Henn, M. R., Brumme, Z. L., Norris, P. J., Rosenberg, E. S., Mayer, K. H., Jessen, H., Pond, S. L. K., Walker, B. D., Altfeld, M., Carlson, J. M., Allen, T. M., Power, K. A., Ghebremichael, M., Bedard, H. E., Gladden, A. D., Seese, A. M., Amero, M. A., Lane, K., Mcgrath, G., Bazner, S. B., Tinsley, J., Lennon, N. J., Henn, M. R., Brumme, Z. L., Norris, P. J., Rosenberg, E. S., Mayer, K. H., Jessen, H., Kosakovsky Pond, S. L., Walker, B. D., Altfeld, M., Carlson, J. M., Allen, T. M., Power, A., Ghebremichael, M., Bedard, H. E., Gladden, A. D., Seese, M., Amero, M. A., Lane, K., Mcgrath, G., Bazner, S. B., Tinsley, J., Lennon, N. J., Henn, M. R., Brumme, Z. L., Norris,

P. J., Rosenberg, E. S., Mayer, K. H., Jessen, H., Pond, S. L. K., Walker, B. D., Altfeld, M., Carlson, J. M., and Allen, T. M. (2016). Differences in the Selection Bottleneck between Modes of Sexual Transmission Influence the Genetic Composition of the HIV-1 Founder Virus. PLoS Pathogens, 12(5):1–29.

[U et al., 2001] U, T. N., Lu, Y., Renjifo, B., Touzjian, N., Kushner, N., Pena-cruz, V., Novitsky, V. A., Lee, T.-h., and Essex, M. A. X. (2001). Infectious simian / human immunodeficiency virus with human immunodeficiency virus type 1 subtype c from an african isolate : Rhesus macaque model. 75(23):11417–11425.

[van Dorp et al., 2014] van Dorp, C. H., van Boven, M., and de Boer, R. J. (2014). Immunoepidemiological Modeling of HIV-1 Predicts High Heritability of the Set-Point Virus Load, while Selection for CTL Escape Dominates Virulence Evolution. PLoS Computational Biology, 10(12).

[Venner et al., 2016] Venner, C. M., Nankya, I., Kyeyune, F., Demers, K., Kwok, C., Chen, P. L., Rwambuya, S., Munjoma, M., Chipato, T., Byamugisha, J., Van Der Pol, B., Mugyenyi, P., Salata, R. A., Morrison, C. S., and Arts, E. J. (2016). Infecting HIV-1 Subtype Predicts Disease Progression in Women of Sub-Saharan Africa. EBioMedicine, 13:305–314.

[Volz et al., 2017] Volz, E. M., Romero-Severson, E., and Leitner, T. (2017). Phylodynamic Inference across Epidemic Scales. Molecular Biology and Evolution, 34(5):1276–1288.

[Vrancken et al., 2014] Vrancken, B., Rambaut, A., Suchard, M. A., Drummond, A., Baele, G., Derdelinckx, I., Van Wijngaerden, E., Vandamme, A.-M., Van Laethem, K., and Lemey, P. (2014). The Genealogical Population Dynamics of HIV-1 in a Large Transmission Chain: Bridging within and among Host Evolutionary Rates. PLoS Computational Biology, 10(4):e1003505.

[Whitney et al., 2014] Whitney, J. B., Hill, A. L., Sanisetty, S., Penaloza-Macmaster, P., Liu, J., Shetty, M., Parenteau, L., Cabral, C., Shields, J., Blackmore, S., Smith, J. Y., Brinkman, A. L., Peter, L. E., Mathew, S. I., Smith, K. M., Borducchi, E. N., Rosenbloom, D. I., Lewis, M. G., Hattersley, J., Li, B., Hesselgesser, J., Geleziunas, R., Robb, M. L., Kim, J. H., Michael, N. L., and Barouch, D. H. (2014). Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. Nature, 512(1).

[Wolinsky et al., 1992] Wolinsky, S. M., Wike, C. M., Korber, B. T., Hutto, C., Parks, W. P., Rosenblum, L. L., Kunstman, K. J., Furtado, M. R., and Muñoz, J. L. (1992). Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. Science (New York, N.Y.), 255(September):1134–1137.

[Zanini et al., 2017] Zanini, F., Puller, V., Brodin, J., Albert, J., and Neher, R. A. (2017). In vivo mutation rates and the landscape of fitness costs of HIV-1. Virus Evolution, 3(1).

[Zimmer, 2002] Zimmer, C. (2002). Envelope-Constrained Neutralization-Sensitive HIV-1 after Heterosexual Transmission. Science, 296(5568):633–635.

Appendix A

Multi-scale model of the evolution of HIV-1

A.1 Parameter and Trial Definitions

Trial: Blueberry

Number	Name	Control	μ(BN)	Latent	α	γ	ρ	μ	λ_μ	λ_δ	ϕ
271	Blueberry271	Full	1	–	0.045	0	0.045	0.08	3.55	0.003	0.26
272	Blueberry272	Full	1	LR	0.045	0.01	0.035	0.08	3.55	0.003	0.26
273	Blueberry273	Full	1	LR	0.045	0.02	0.025	0.08	3.55	0.003	0.26
274	Blueberry274	Full	1	LR	0.045	0.05	0	0.08	3.55	0.003	0.26
275	Blueberry275	Full	1	LR	0.045	0.1	0	0.08	3.55	0.003	0.26
276	Blueberry276	Fixed	1	–	0.045	0	0.045	0.08	3.55	0.003	0.26
277	Blueberry277	Fixed	1	LR	0.045	0.01	0.035	0.08	3.55	0.003	0.26
278	Blueberry278	Fixed	1	LR	0.045	0.02	0.025	0.08	3.55	0.003	0.26
279	Blueberry279	Fixed	1	LR	0.045	0.05	0	0.08	3.55	0.003	0.26
280	Blueberry280	Fixed	1	LR	0.045	0.1	0	0.08	3.55	0.003	0.26

Trial: Primary

#	Name	Control	μ(BN)	Latent	α	γ	ρ	μ	λ_μ	λ_δ	ϕ
1	Primary1	Full	1	–	0.055	0	0.055	0.08	3.55	0.015	0.248
2	Primary2	Full	1	LR	0.055	0.02	0.035	0.08	3.55	0.015	0.248
3	Primary3	Fixed	1	–	0.055	0	0.055	0.08	3.55	0.015	0.248
4	Primary4	Fixed	1	LR	0.055	0.02	0.035	0.08	3.55	0.015	0.248
5	Primary5	Full	2	–	0.055	0	0.055	0.08	3.55	0.015	0.248
6	Primary6	Full	2	LR	0.055	0.02	0.035	0.08	3.55	0.015	0.248
7	Primary7	Fixed	2	–	0.055	0	0.055	0.08	3.55	0.015	0.248
8	Primary8	Fixed	2	LR	0.055	0.02	0.035	0.08	3.55	0.015	0.248
9	Primary9	Full	1	–	0.055	0	0.055	0.08	3.55	0.02	0.248
10	Primary10	Full	1	LR	0.055	0.02	0.035	0.08	3.55	0.02	0.248
11	Primary11	Fixed	1	–	0.055	0	0.055	0.08	3.55	0.02	0.248
12	Primary12	Fixed	1	LR	0.055	0.02	0.035	0.08	3.55	0.02	0.248
13	Primary13	Full	2	–	0.055	0	0.055	0.08	3.55	0.02	0.248
14	Primary14	Full	2	LR	0.055	0.02	0.035	0.08	3.55	0.02	0.248
15	Primary15	Fixed	2	–	0.055	0	0.055	0.08	3.55	0.02	0.248
16	Primary16	Fixed	2	LR	0.055	0.02	0.035	0.08	3.55	0.02	0.248
17	Primary17	Full	1	–	0.055	0	0.055	0.08	3.55	0.025	0.248
18	Primary18	Full	1	LR	0.055	0.02	0.035	0.08	3.55	0.025	0.248
19	Primary19	Fixed	1	–	0.055	0	0.055	0.08	3.55	0.025	0.248
20	Primary20	Fixed	1	LR	0.055	0.02	0.035	0.08	3.55	0.025	0.248
21	Primary21	Full	2	–	0.055	0	0.055	0.08	3.55	0.025	0.248
22	Primary22	Full	2	LR	0.055	0.02	0.035	0.08	3.55	0.025	0.248
23	Primary23	Fixed	2	–	0.055	0	0.055	0.08	3.55	0.025	0.248
24	Primary24	Fixed	2	LR	0.055	0.02	0.035	0.08	3.55	0.025	0.248

Trial: Sensitivity

#	Name	Control	μ(BN)	Latent	α	γ	ρ	μ	$\lambda\mu$	$\lambda\delta$	ϕ
1	Sensitivity1	Full	1	LR	0.055	0.02	0.035	0.08	3.55	0.003	0.248
2	Sensitivity2	Full	1	LR	0.055	0.018	0.037	0.08	3.55	0.003	0.248
3	Sensitivity3	Full	1	LR	0.055	0.022	0.033	0.08	3.55	0.003	0.248
4	Sensitivity4	Full	1	LR	0.055	0.0145	0.035	0.08	3.55	0.003	0.248
5	Sensitivity5	Full	1	LR	0.055	0.0255	0.035	0.08	3.55	0.003	0.248
6	Sensitivity6	Full	1	LR	0.0495	0.02	0.035	0.08	3.55	0.003	0.248
7	Sensitivity7	Full	1	LR	0.0605	0.02	0.035	0.08	3.55	0.003	0.248
8	Sensitivity8	Full	1	LR	0.055	0.02	0.035	0.08	3.195	0.003	0.248
9	Sensitivity9	Full	1	LR	0.055	0.02	0.035	0.08	3.905	0.003	0.248
10	Sensitivity10	Full	1	LR	0.055	0.02	0.035	0.08	3.55	0.0027	0.248
11	Sensitivity11	Full	1	LR	0.055	0.02	0.035	0.08	3.55	0.0033	0.248
12	Sensitivity12	Full	1	LR	0.055	0.02	0.035	0.08	3.55	0.003	0.2232
13	Sensitivity13	Full	1	LR	0.055	0.02	0.035	0.08	3.55	0.003	0.2728

Within-host Parameters				
parameter	description	units	default value	source
λ	growth rate of healthy immune cells	cells·(ml·day)$^{-1}$	$x_0^2(5.7 \times 10^{-5})$	[Stafford et al., 2000]
δ	death rate of infected cells	day^{-1}	0.0433	[Stafford et al., 2000]
d	death rate of healthy immune cells	day^{-1}	0.0057	[Stafford et al., 2000]
β_0	initial β	ml·(cells·day)$^{-1}$	.0005	
β_M	maximum β	ml·(cells·day)$^{-1}$	.000525	
β_λ	shape of β fitness kernel	mutation class^{-1}	.07	
δ_0	initial δ	day^{-1}	.8	
δ_m	minimum δ	day^{-1}	0.0057	
δ_λ	shape of δ fitness kernel	mutation class^{-1}	.07	
μ	mean of mutation kernel	day^{-1}	5×10^{-5}	[Lythgoe et al., 2013]
λ_μ	shape of mutation kernel	mutation class^{-1}	4	
γ	latency entrance rate	day^{-1}	0.01	[Doekes et al., 2017, Lythgoe et al., 2013]
α	activation rate from latent reservoir	day^{-1}	0 to 0.01	[Doekes et al., 2017, Lythgoe et al., 2013]
ρ	latent reservoir cellular proliferation rate	day^{-1}	$\alpha - \gamma/r_L$	[Doekes et al., 2017, Lythgoe et al., 2013]
r_L	relative size of the latent reservoir	cells·ml^{-1}	0.01 to 2	[Doekes et al., 2017, Lythgoe et al., 2013]
T_M	maximum infection time	years	5	
$y_1(0)$	mean # of contacts		1 to 5	

Table A.1: Default within-host parameters

A.2 Simulation Sketch

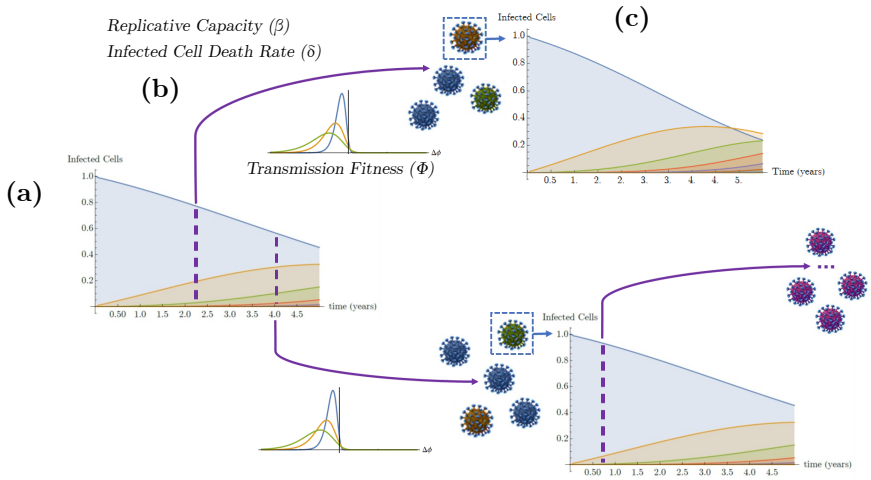


Figure A.1: (a) The results of a within-host model of viral competition with a latent reservoir and distinct mutation classes. Each mutation class may have, for example, an increased replicative capacity, β. (b) A transmission event where transmission fitness (ϕ) replicative capacity (β) and infected cell death rate (δ) are selected and transmitted to the next host. The minimum of 1 and a Poisson distributed number of infected cells are randomly selected from the transmitter across, and proportional to, the within-host mutation classes. Transmission fitness for each selected virion is selected from a flipped inverse gamma distribution where the variance of the distribution is determined by the mutation class of the selected virus. β and δ are inherited from the mutation class from which the potential founder was selected. (c) The next infection in the transmission chain is founded by the virion with the maximum transmission fitness and this ϕ then becomes the mean of the flipped inverse gamma distribution from which ϕ is selected for any following transmissions. This new host inherits mutation class, β, and δ from the selected virion which then are used as initial conditions for the next within-host model.

A.3 Transmission Timing

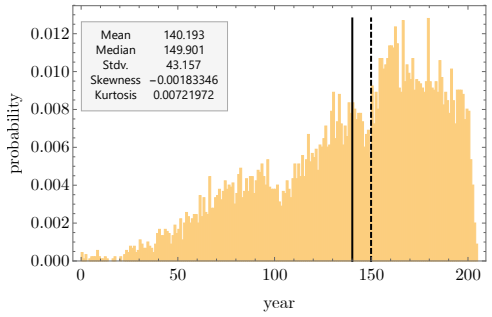


Figure A.2: Transmission time histogram of a surviving tree. The histogram shows the transmission times from the experiment *Primary* replicate number 65. The solid line is the mean and the dashed line in the median.

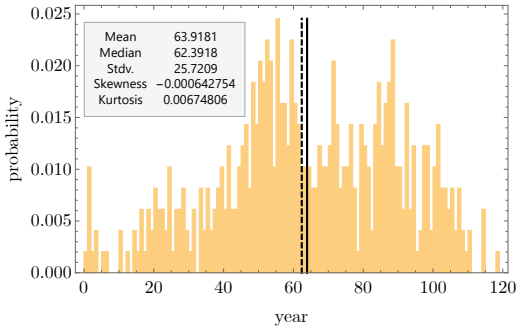


Figure A.3: Transmission time histogram of a extinct tree. The histogram shows the transmission times from the experiment *Primary* replicate number 660. The solid line is the mean and the dashed line in the median.

A.4 Trial: *Primary* comparisons

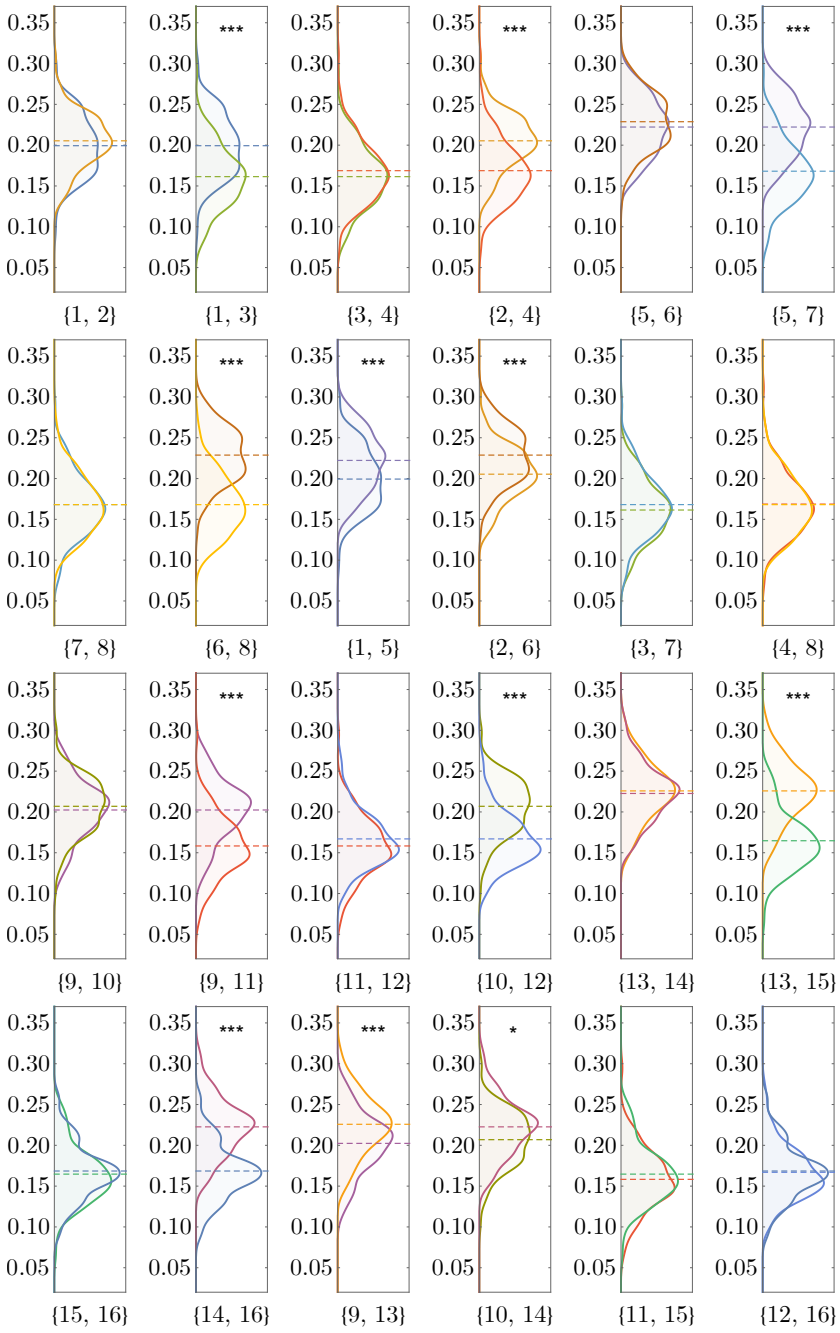


Figure A.4: Comparison of experiments from *Primary* at 200 years of evolution. Significance shown is for a two-tailed Mann-Whitney test at the *** 1%, ** 5%, and * 10% levels including the Bonferroni correction $\left(\frac{\alpha}{24}\right)$ for 24 comparisons.

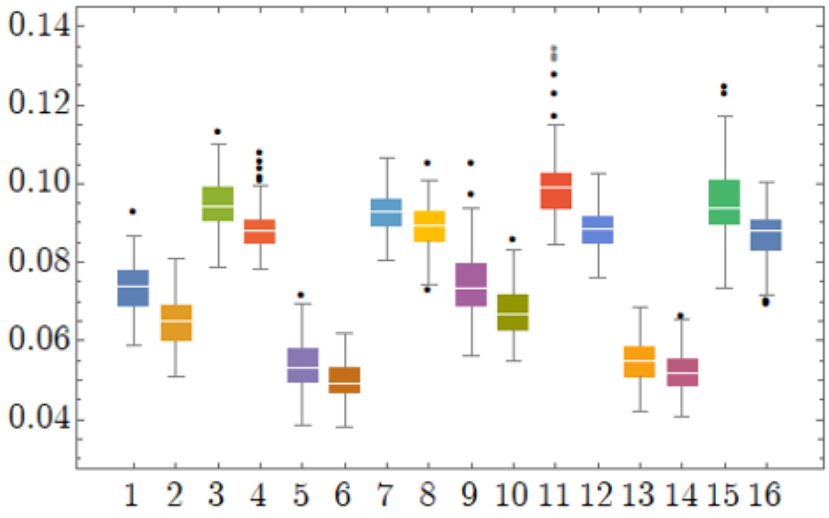


Figure A.5: Mean transmitted mutation class averaged over the epidemic for experiments from *Primary*.

Appendix B

HIV subtype fitness differences *in vivo*

B.1 Inter-appointment waiting times

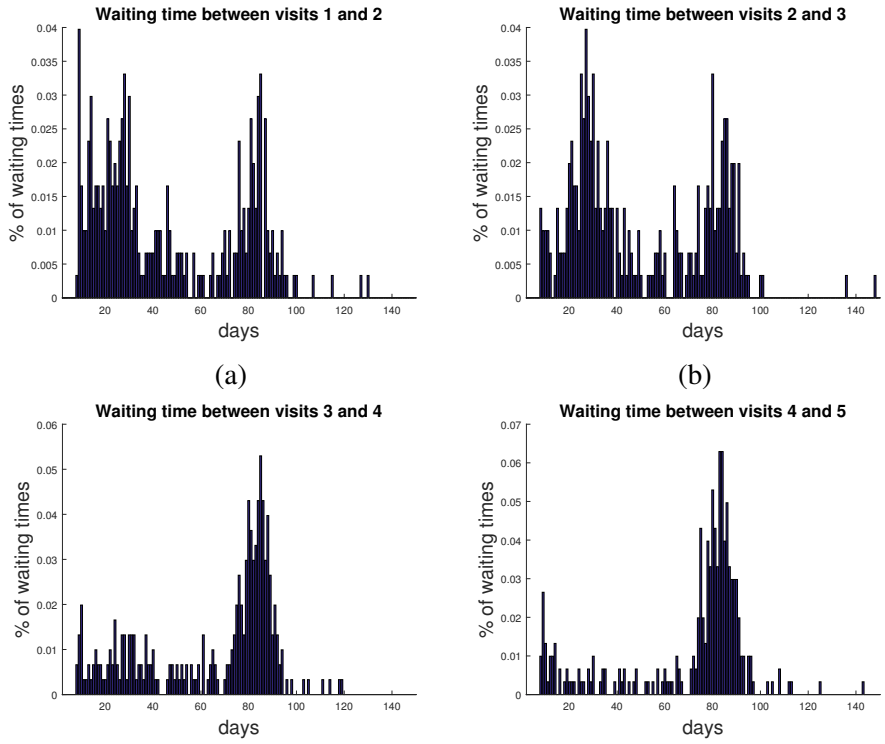
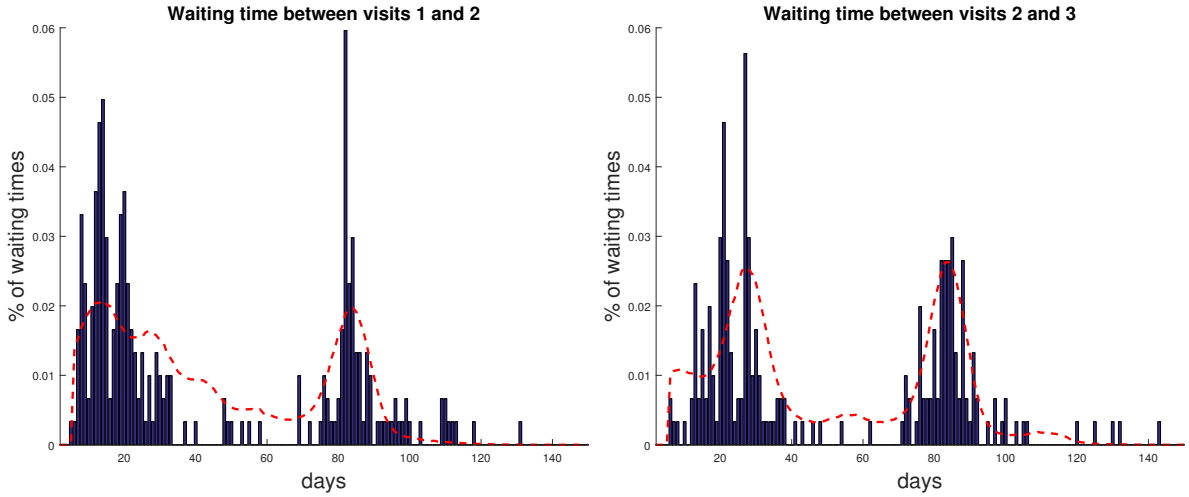


Figure B.1: Histogram of the waiting times between the first and second, (a), the second and third, (b), third and fourth, (c), and fourth and fifth appointments, (d).

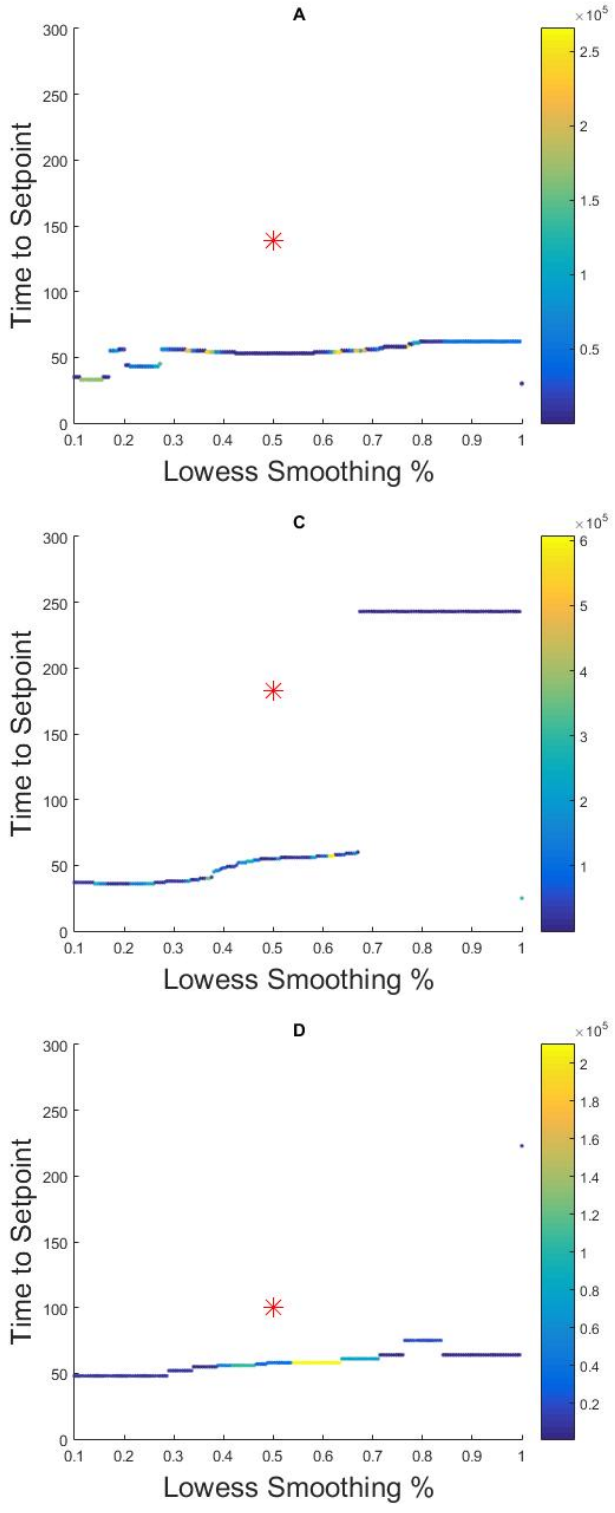
B.1.1 Simulated inter-appointment waiting times



(a) (b)

Figure B.2: Histogram of randomly generated realistic appointment timing fit the time between the first and second visit, panel (a), the second and thrid visit, panel (b).

B.2 Estimation of SPVL using LOWESS smoothing



The whole data set was used and then time to setpoint was estimated using lowess smoothing. Each dots location with respect to the vertical axis represents the estimated time to setpoint using the first minimum after the highest rate of descent method and each dots location w.r.t the horizontal axis represents the local percent of the first 300 days of data used in the local smoothing. The colour of the dots represents the plasma viral RNA/ml at setpoint. The red '*' represents the results from Morrison et. all 2010 were the percent of smoothing was arbitrarily chosen to be 50% since the authors did not restrict the data to the first 300 days.

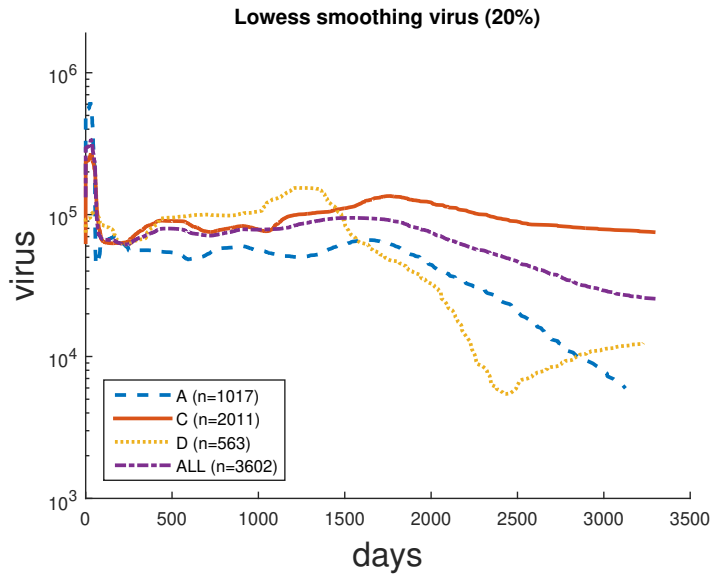


Figure B.6: Estimated SPVL over time. Lowess smoothing at 20% of available data. *n* is all datapoints used.

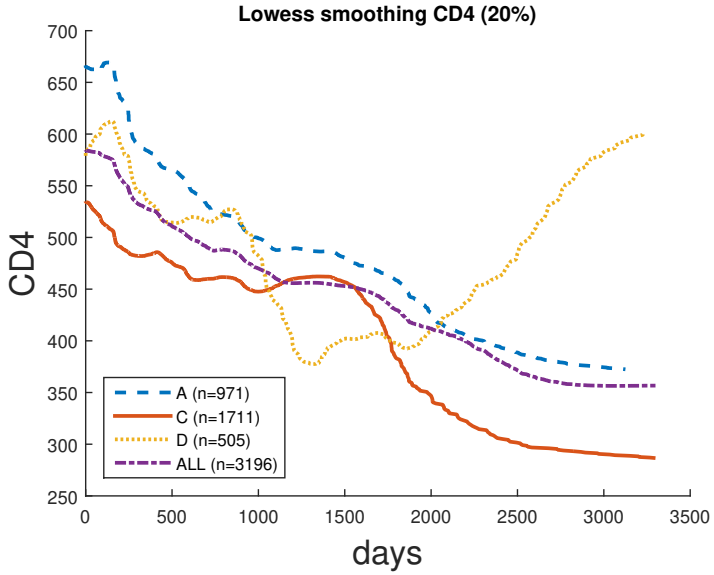


Figure B.7: Estimated CD4 cell counts over time. Lowess smoothing at 20% of available data. *n* is all datapoints used.

B.3 Realistic simulated appointment timing

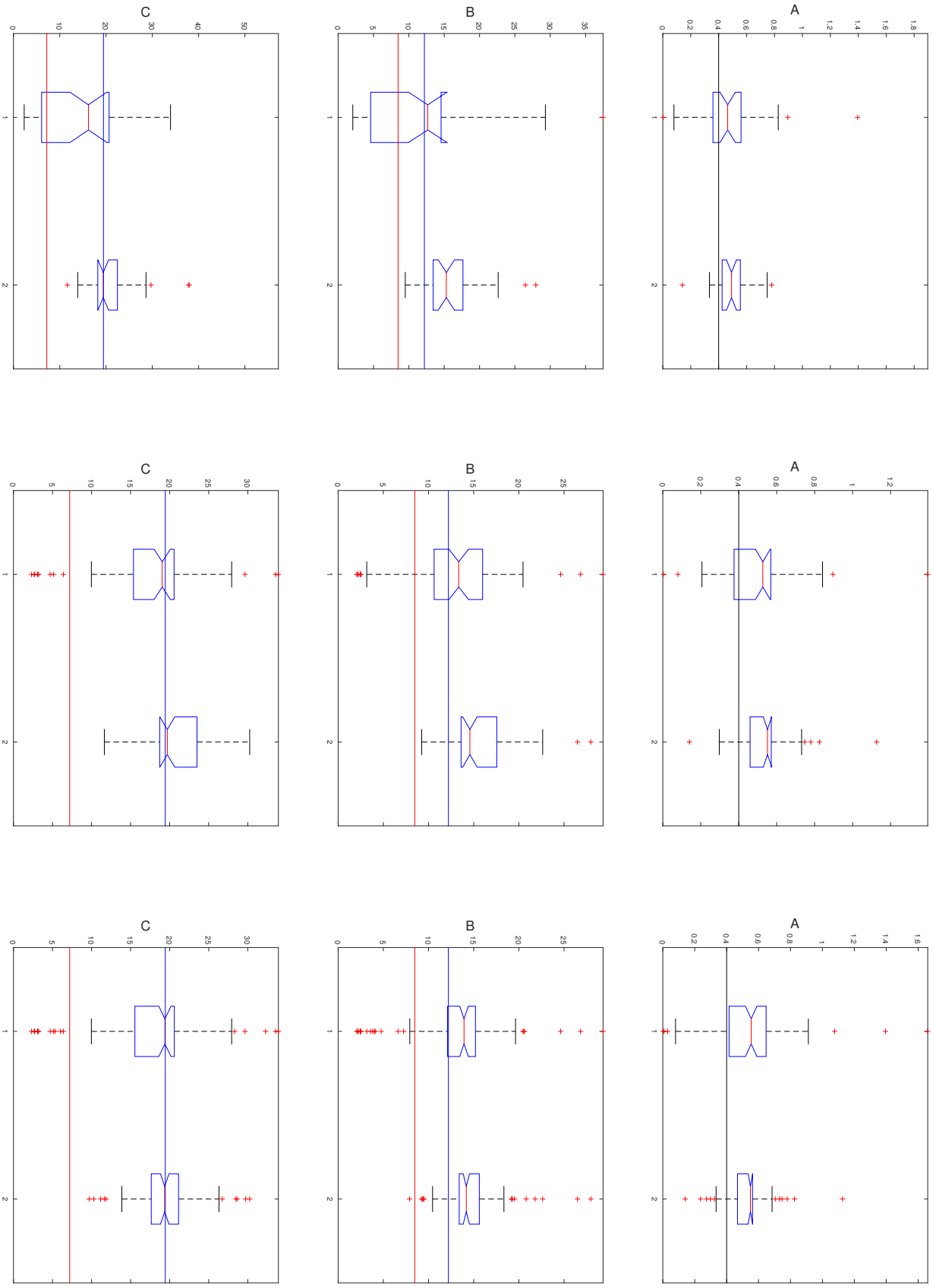


Figure B.8: Boxplot of the fit parameters for realistic appointment timing. The solid lines represent the actual parameter values.

Appendix C

SHIV Data Summary

C.1 Interval of maximum slope

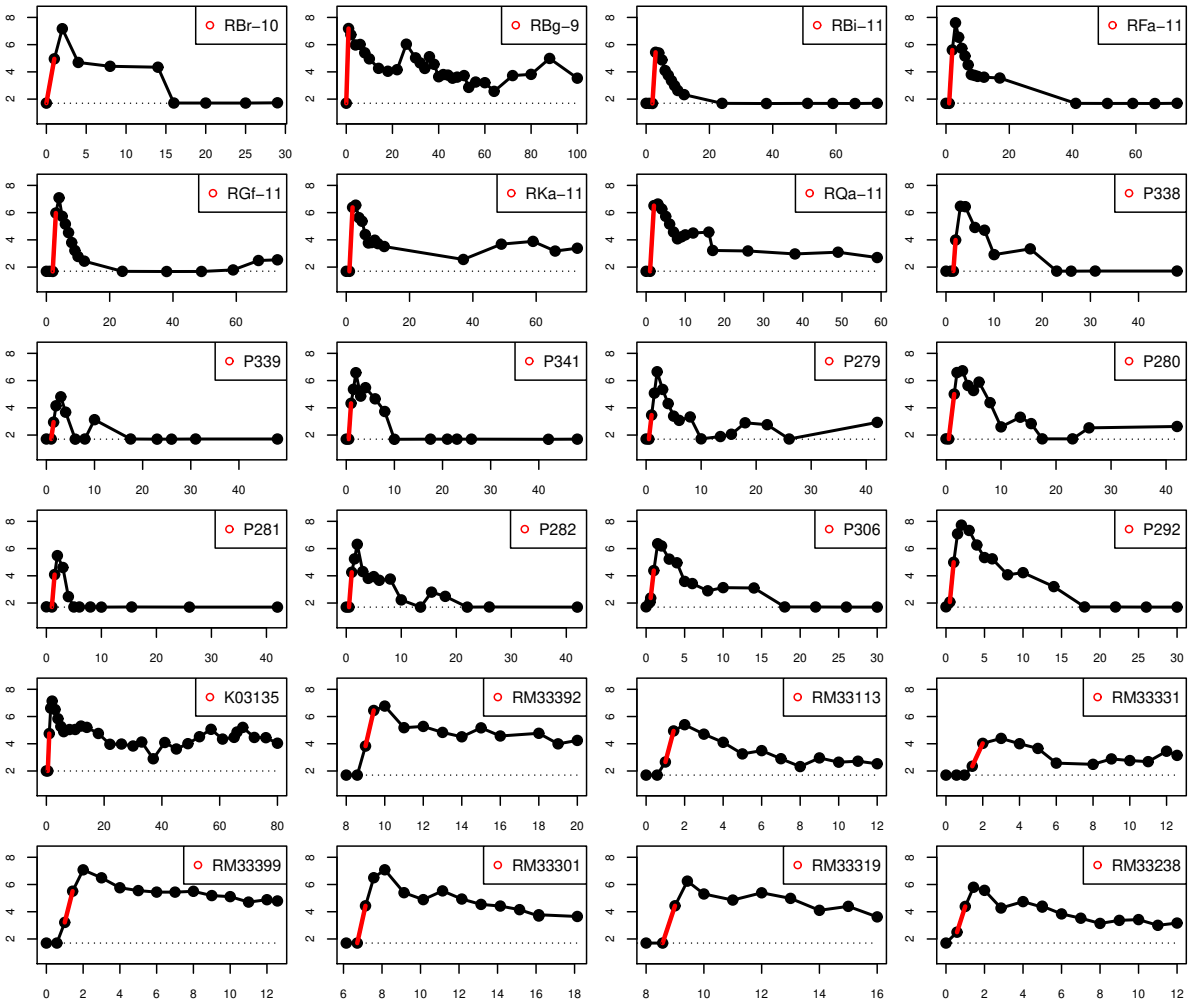


Figure C.1: Interval of maximum slope for all 24 subtype C primates

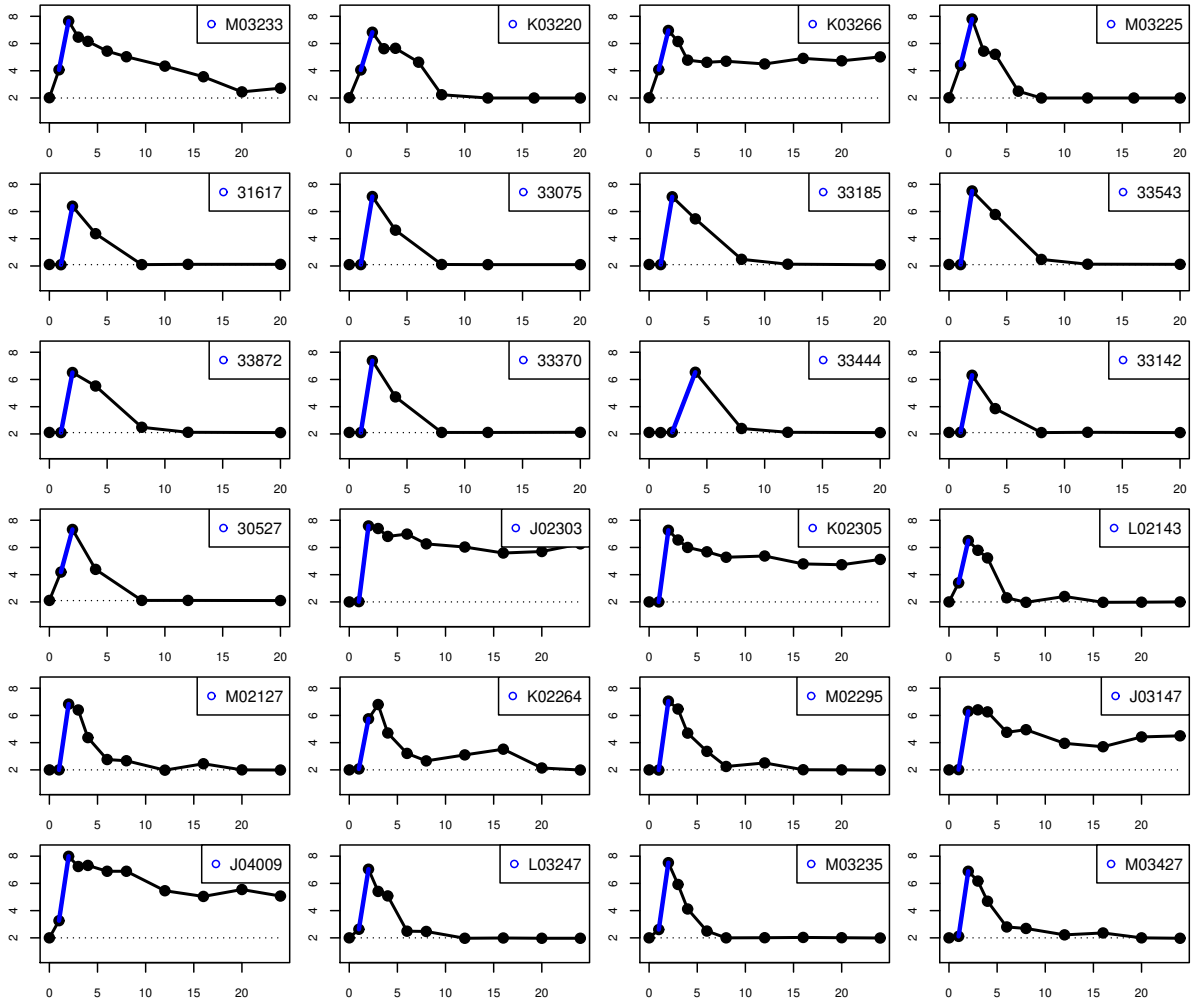


Figure C.2: Interval of maximum slope for 24 subtype B primates

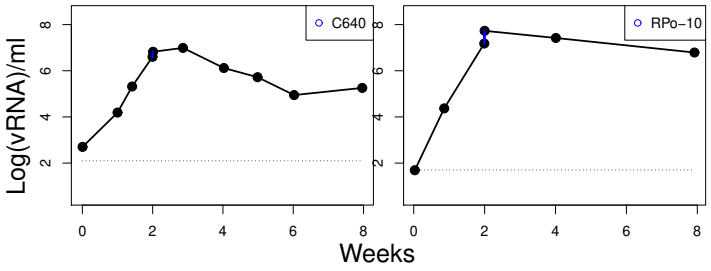


Figure C.3: Interval of maximum slope outliers

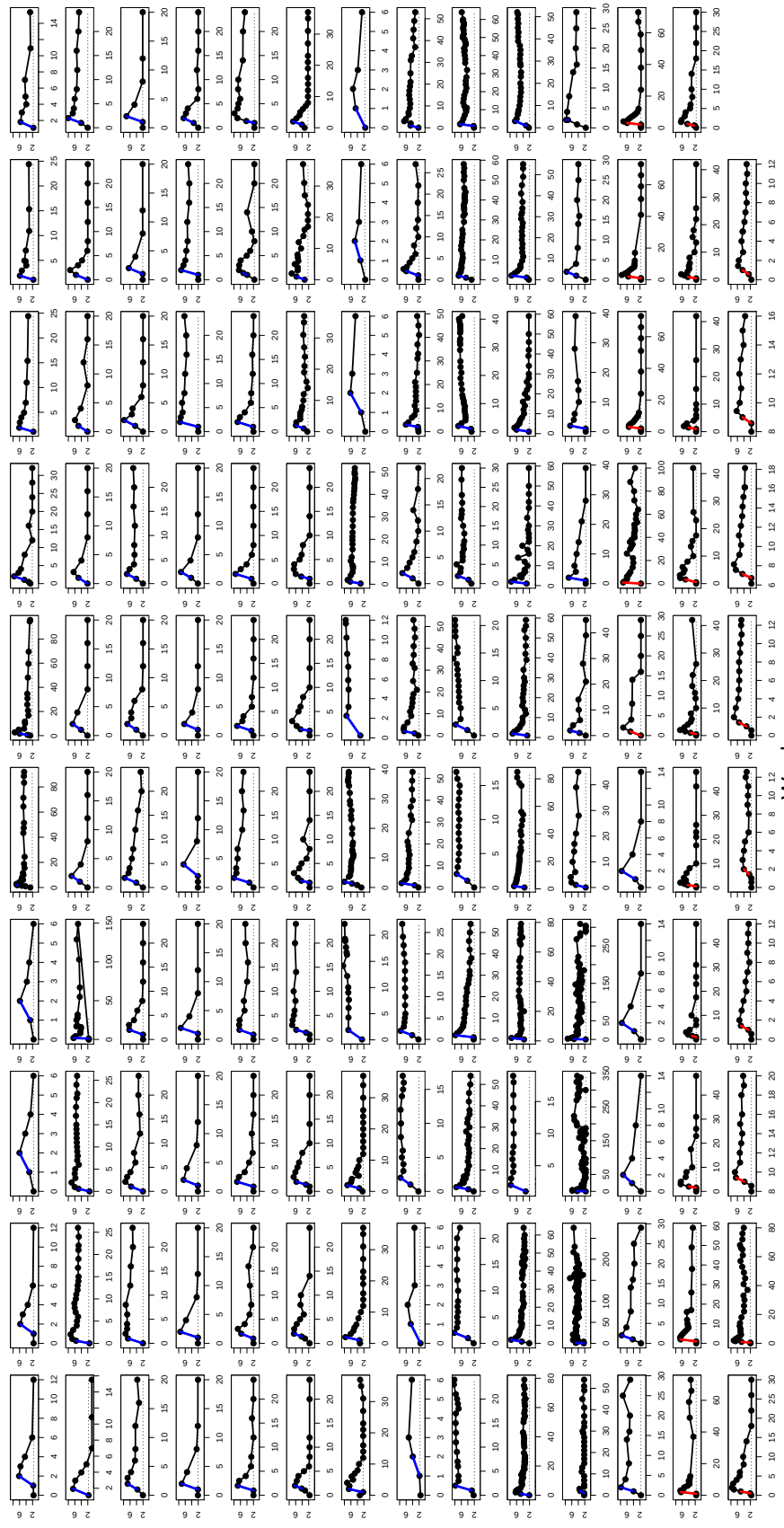


Figure C.4: Interval of maximum slope for all primates

C.2 With-in host replication rate β

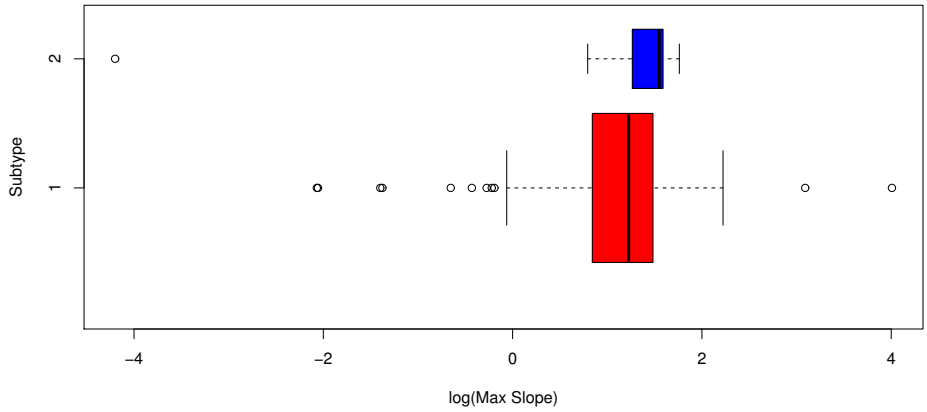


Figure C.5: A boxplot of the maximum *log*(slope) between sample points for subtypes 1 = *B* and 2 = *C*, with 115 primates infected with subtype *B* and 17 infected with subtype *C*. The Wilcoxon rank sum test for equal mean maximum slopes between subtypes yield a p-value of 0.03521.

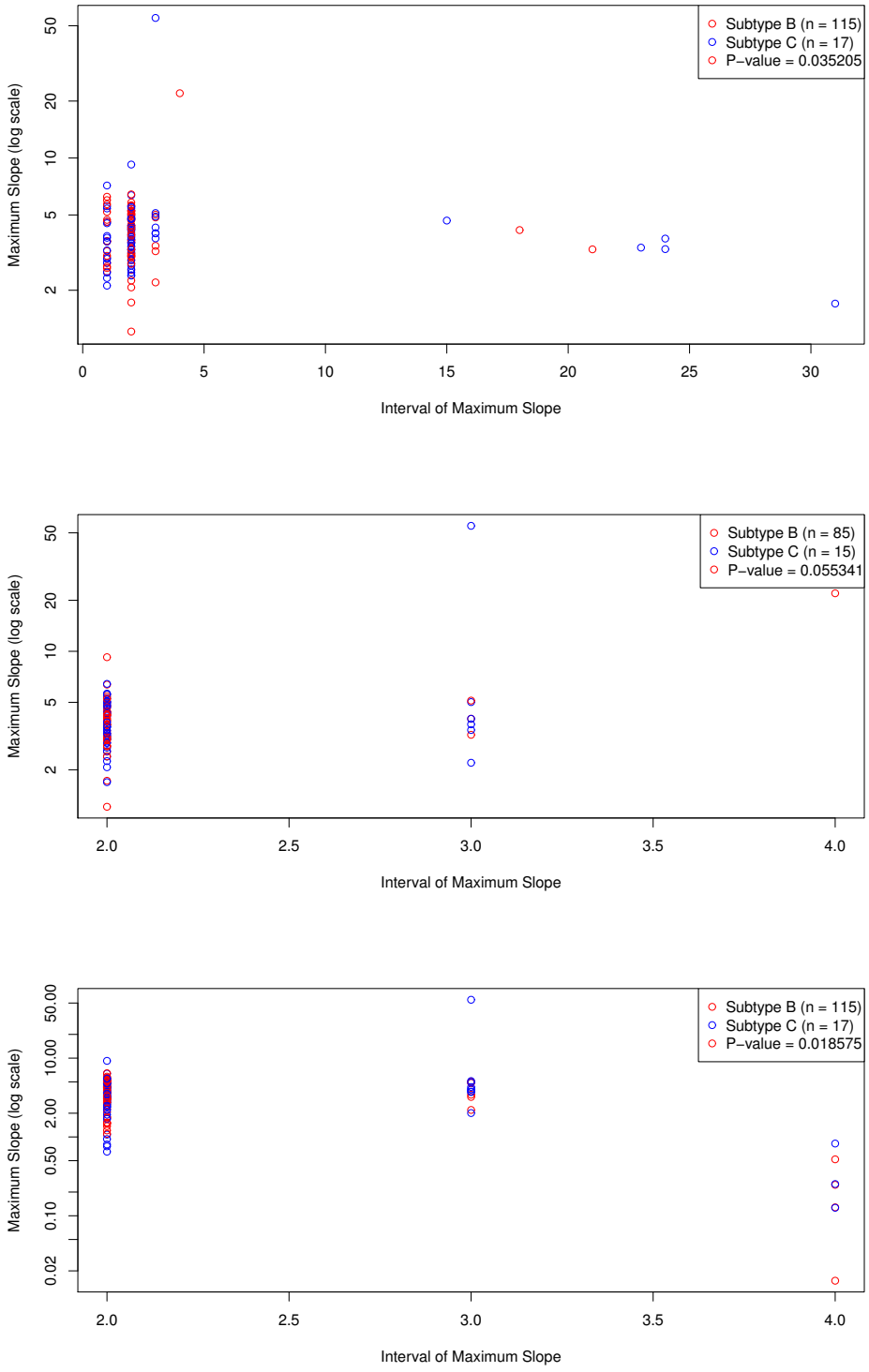


Figure C.6: Plot of the maximum *log*(slope) between sample points, over all data (top plot) and between in intervals 2 to 5 (bottom plot), for subtypes $1 = B$ and $2 = C$. The P-value is reported from the Wilcoxon rank sum test for equal mean maximum slopes between subtypes.

Appendix D

SHIV Figures and Tables

D.1 Initial parameter distributions

Parameter	Distribution	Covariate	Population IE
λ	Lognormal	Species	8000
β	Lognormal	Subtype	2.92e-6
d_x	Lognormal		0.04
d_y	Lognormal		0.8
p_0	Lognormal	Species,Subtype	205
c_v	Lognormal		2
T_0	Lognormal		10000
v_0	Lognormal		0.03

Figure D.1: IRIE model initial parameter distributions, covariates, and initial estimates.

Parameter	Distribution	Covariate	Population IE
λ	Lognormal	Species	8000
β	Lognormal	Subtype	2.92e-6
d_x	Lognormal		0.04
d_y	Lognormal		0.8
p_0	Lognormal	Species,Subtype	205
c_v	Lognormal		2
T_0	Lognormal		10000
v_0	Lognormal		0.03

Figure D.2: RestingIE model initial parameter distributions, covariates, and initial estimates.

Parameter	Distribution	Covariate	Population IE
λ	Lognormal	Species	8000
β	Lognormal	Subtype	2.92e-6
d_x	Lognormal		0.04
d_y	Lognormal		0.8
p_0	Lognormal	Species,Subtype	205
c_v	Lognormal		2
T_0	Lognormal		10000
v_0	Lognormal		0.03

Figure D.3: RIRIE model initial parameter distributions, covariates, and initial estimates.

D.2 RIR-Linear Stratified prediction errors

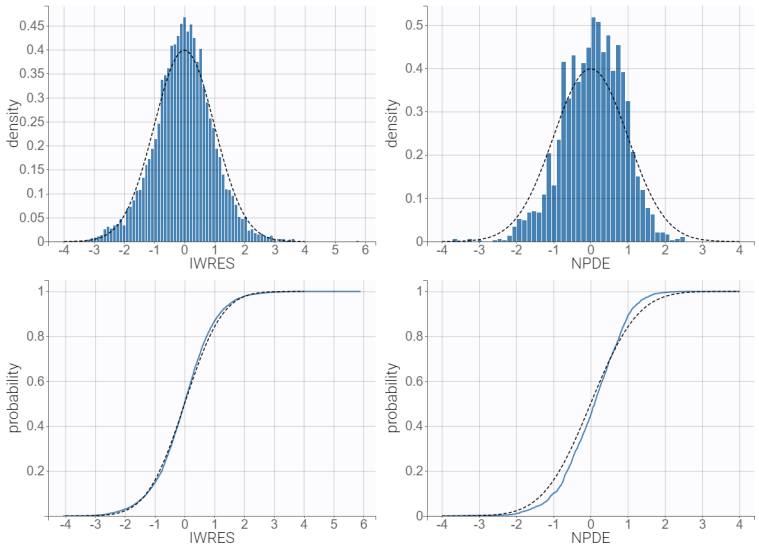


Figure D.4: Individual weighted residuals and Normalized prediction distribution errors for the RIR-Linear Stratified model.

Curriculum Vitae

DAVID W. DICK

EDUCATION

2015 - Present **PhD Candidate, Applied Mathematics**, Western University, London, Ontario, Canada
Supervisors: Lindi Wahl, Art Poon

2013 - 2015 **Masters of Science, Mathematics**, Wilfrid Laurier University, Waterloo, Ontario, Canada
Thesis Title: Introduction of stochastic compartmental disease model with arbitrary interaction and recovery distributions
Supervisor: Adam Metzler

2001 - 2013 **Bachelor of Arts, Mathematics**, Wilfrid Laurier University, Waterloo, Ontario, Canada

RESEARCH AND TEACHING INTERESTS

Modeling complex multiscale problems in virus evolution using mathematical tools with a focus on the evolution of HIV and emerging diseases

Stochastic modelling and simulation applied to viral dynamics

Phylogenetic clustering and visualization methods with interest in elucidating the long term evolution and recombination of HIV

RESEARCH EXPERIENCE

2015 - present **Doctoral Candidate,** Western University, London, Ontario, Canada
HIV-1 viral dynamics, stochastic modeling and simulation, ODE modeling and data fitting, phylogenetic distance metrics, and bacteriophage evolution

2013 - 2015 **Graduate Student,** Wilfrid Laurier University, Waterloo, Ontario, Canada
Stochastic modeling and simulation, deterministic disease modeling

CONFERENCES, TALKS, AND POSTERS

International Talks

2019	Virginia Tech MathBio Seminar, Blacksburg, Virginia ,USA Invited Speaker: Modeling the multiscale evolution of HIV-1
2019	Society for Industrial and Applied Mathematics Conference on Applications of Dynamical Systems (DS19), Snowbird, Utah, USA Invited Mini Symposium Talk: Evolution of HIV-1 Across the Within- and Between-host Scales: The Role of Transmission Bottlenecks
2019	26th HIV Dynamics & Evolution, Cascais, Portugal The role of transmission bottlenecks in the evolution of HIV-1
2017	Applied Mathematics, Modeling and Computational Science, Wilfrid Laurier University, Waterloo, Ontario, Canada Modelling fitness evolution of HIV-1 Including the Impact of Both Viral Replication and Transmission
2015	Fifth International Conference on Mathematical Modeling and Analysis of Populations in Biological Systems, Western University, London, Ontario, Canada Determining viral load set point and time to reach viral load set point among patients infected with HIV-1 subtypes A,C and D.
2015	Canadian Industrial and Applied Mathematics Society/Applied Mathematics, Modeling and Computational Science, Wilfrid Laurier University, Waterloo, Ontario, Canada In-host HIV model describes differences in disease progression among patients infected with HIV-1 subtypes A, C and D.

International Posters

2018	25th HIV Dynamics & Evolution, Leavenworth, Washington, USA Modeling two distinct selective challenges faced by HIV-1 at the within- and between-host levels
2017	Keystone Symposia: Modeling Viral Infections and Immunity, Estes Park, Colorado, USA Using within host mathematical models of acute SHIV infection with Bayesian tools to assess in vivo fitness differences in subtypes B and C.
2016	23rd HIV Dynamics & Evolution, Woods Hole, Massachusetts, USA Modeling Acute SHIV Infection to assess in vivo fitness differences in subtypes B and C: A meta-analysis

National Talks

2019	CMS Winter Meeting, Toronto, ON, Canada Invited Mini Symposium Talk: Modeling the evolution of HIV-1 across the within- and between-host scales

2018 Canadian Industrial and Applied Mathematics Society, Ryerson University, Toronto, Ontario Canada
Invited Mini Symposium Talk: Modeling two distinct selective challenges faced by HIV-1 at the within- and between-host levels to elucidate the evolution of set-point viral load as a proxy of Virulence

2017 Canadian Industrial and Applied Mathematics Society, Dalhousie University, Halifax, Nova Scotia, Canada
Modeling the fitness evolution of HIV at the within- and between-host levels

Institutional Talks

2017 Undergraduate Society of Applied Mathematics, Western University, London, Ontario, Canada
Invited Talk: Modeling the evolution of virulence of HIV-1

National Posters

2018 Canadian Association for HIV Research, Vancouver, British Columbia, Canada
Modeling Two Distinct Evolutionary Forces on HIV-1 at the Within- and Between-host Levels

Conference Attendee

2019 Probabilistic Structures in Evolution, Bielefeld University, Bielefeld, Germany

2015 2nd Workshop on Virus Dynamics, Fields Institute, Toronto, Ontario, Canada

Non-Presenting Author

2019 26th HIV Dynamics & Evolution, Cascais, Portugal
Characterizing the global recombinant history of HIV-1 group m with dynamic network community detection

2018 25th HIV Dynamics & Evolution, Leavenworth, Washington, USA
Evidence for a fragmented origin of HIV-1 group M from a sliding window analysis of near full-length genomes

2016 Physics Undergraduate Conference, London, Canada
Agent Based Modelling: A simulation of HIV1 dynamics in an environment of interacting CD4+ cells

PUBLICATIONS

2019 Abayomi S Olabode, Mariano Avino, Garway T Ng, Faisal Abu-Sardanah, David W Dick, Art F Y Poon, Evidence for a recombinant origin of HIV-1 Group

M from genomic variation, *Virus Evolution*, Volume 5, Issue 1, January 2019, vey039, https://doi.org/10.1093/ve/vey039

2018 Wahl, L. M., Betti, M. I., Dick, D. W., Pattenden, T. and Puccini, A. J. (2019), Evolutionary stability of the lysis-lysogeny decision: Why be virulent?. Evolution, 73: 92-98. doi:10.1111/evo.13648

In Preparation

David W. Dick, Jason Knapp, Eric Arts, Ruth M. Ruprecht, Art F. Y. Poon and Lindi M. Wahl
Meta-analysis of in vivo fitness differences between SHIV subtypes B and C in acute infection.

David W. Dick, Jacqueline Doan, Art F. Y. Poon and Lindi M. Wahl
Evolution of HIV-1 Across the Within- and Between-host Scales: The Role of Transmission Bottlenecks

SUPERVISING EXPERIENCE

Co-supervisor and Primary Contact, Western University, London, Ontario, Canada

2019 Jacqueline Doan, Undergraduate research student supported by a national undergraduate student research award

2019 Jacqueline Doan, Undergraduate research volunteer
Project: Multiscale simulation of HIV evolution implemented in Mathematica

2015-2016 Dustin Sokolowski, Undergraduate research volunteer
Project: Agent based simulation of cell-to-cell transmission of HIV-1 implemented in Python.

TEACHING EXPERIENCE

Guest Lecture, Western University, London, Ontario, Canada

2017 Calculus and Probability with Biological Applications, AM1201b
Created and presented three guest lectures on probability

2017 Advanced Applied Mathematics for Electrical Engineering, AM3415a

Coordinating Teaching Assistant, Western University, London, Ontario, Canada

2019 Calculus and Probability with Biological Applications, AM1201b

Teaching Assistant, Western University, London, Ontario, Canada

2019 Applied Mathematics for Engineers, AM 1413

2015, 2016, 2018	Calculus and Probability with Biological Applications, AM1201b
2017	Applied Mathematics for Electrical and Mechanical Engineering III, AM2276A
2017	Advanced Applied Mathematics for Electrical Engineering, AM3415a
2016	Ordinary Differential Equations, AM2402A
2015	Applied Mathematics for Engineering II, AM2270A

2013, 2014

Teaching Assistant, Wilfrid Laurier University, Waterloo, Ontario, Canada
Introductory Calculus for Business and Social Sciences, MA129

2012, 2013

Mathematics Assistance Center Tutor, Wilfrid Laurier University, Waterloo, Ontario, Canada
Homework session: Introduction to Mathematical Proofs, MA121
Drop in help center: All first and second year mathematics and statistics courses

ACADEMIC AND ADMINISTRATIVE EXPERIENCE

2013 - 2014 Elected Graduate Representative, Wilfrid Laurier University, Waterloo, Ontario, Canada

TECHNICAL SKILLS

Computer algebra: Mathematica, Maple
Programing languages: Mathematica, MATLAB, R, Python
Data fitting packages: Monolix, RStan