

https://dspace.org/

A new model of population dynamics of HIV-1 in vivo

Cawse, Kerry-Anne

2009-09-08T09:00:44Z
http://hdl.handle.net/10539/7206

Downloaded from a DSpace repository. https://dspace.org/

MSc Dissertation:
A New Model of Population Dynamics of HIV-1 In Vivo

Kerry-Anne Cawse
Student number: 0408397A

Supervised by Prof D Sherwell

School of Computational and Applied Mathematics

March 12, 2009

Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Science in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.



Kerry-Anne Cawse
Johannesburg, 2009

Abstract

Using a simple model of HIV-1 viral load and infected CD4+ count, with easily measurable coefficients, valid on time scales of approximately a month, we find that HIV-1 is understood by sequential outburst of viral populations with delay between populations. AIDS is characterised by shorter delays between viral populations and HAART is characterised by longer delays between viral populations, as evidenced in observed viral blips. Infected CD4+ dynamics is understood by sequential outburst with overlap of CD4+ populations. We argue that viral load and CD4+ counts must be in phase and present a model in differential equations that supports this conclusion and is consistent with decay rates discovered by the outburst model.

Acknowledgments

I gratefully acknowledge a scholarship from the South African Centre for Epidemiological Modelling and Analysis (SACEMA). Assistance with the biology was given by Prof Tracy McLellan, Dr Alex Welte and Dr Eftyhia Vardas, and is much appreciated.

I would also like to express gratitude to my supervisor, Prof David Sherwell, for his support, effort and ideas.

Contents

1 Introduction **1**
1.1 Background . . . 1
1.2 The Existence of a Set-Point . . . 1
1.3 Review of Mathematical HIV-1 models . . . 3
1.4 The MACS Data [27] . . . 8
1.4.1 Inconsistencies . . . 9
1.4.2 Errors on Laboratory Measurements . . . 9
1.4.3 Default Viral Load Values . . . 10

2 The Models **11**
2.1 Objectives . . . 11
2.2 Primary Assumptions . . . 12
2.3 Methods . . . 13
2.4 The Overlap Model . . . 15
2.5 The Intermittent Model . . . 17

3 The ODE model **20**
3.1 The Equations . . . 21
3.2 The Constants . . . 22
3.3 Fixed Point . . . 31
3.4 Constants with Dimension . . . 32
3.5 Mean Viral Load and CD4+ . . . 35

4 Discussion **36**
4.1 Deciding the Model for Viral Load and CD4+ Data . . . 36
4.1.1 Viral Load Data . . . 36
4.1.2 CD4+ Data . . . 38
4.2 Time Scales and Phasing of HIV-1 and Infected CD4+ Cells . . . 41
4.3 Our Model vs Henrard's Results . . . 43
4.4 Window Period and End Phase . . . 44
4.5 The Effect of HAART . . . 46
4.6 The Effect of the Parabolic Shape . . . 47
4.6.1 On the Overlap model . . . 47
4.6.2 On the Intermittent Model . . . 48
4.6.3 On Ignoring Tails . . . 49

5 Conclusions **51**

6 References **53**

List of Figures

1.1 (Taken from [11]) A schematic graph of the progression of various biological measurements for an untreated HIV-1 infected patient. We focus on the line of "Plasma RNA copies" in the bottom graph representing viral load progression and the top graph showing the behaviour of the CD4+ count. 2
1.2 The viral load data points of one MACS patient. This data shows enormous amounts of variability, with the range significantly larger than the mean. This contrasts with the idea of a set point in the sense of equilibrium viral load. . 3
1.3 (Taken from Virus Dynamics [29] p.138) The solution curves to Nowak's differential equations (12.4). Each strain is born with probability and reaches an equilibrium. 5
1.4 (Taken from Virus Dynamics [29] p.158) The solution curves to Nowak's differential equations (14.1). 6
1.5 (Taken from [9]) The behavior of viral load that occurs in a patient on HAART during a blip. The time scale of this profile is similar to that of the window period, and will be used as an approximation for the lifetime of a viral population. 7
1.6 (Taken from [1]) Figure D shows the long time scales of Bajaria *et al*'s model and displays the set point. We find large variation around this set point and we believe this variation should be exploited further. 8
2.1 The mean height of a square is $\alpha = h$. 14
2.2 The mean height of a parabola is $\alpha = \frac{2}{3}h$. 14
2.3 The mean height of a gaussian $e^{\frac{-t^2}{\delta}}$ is $\alpha = 0.5652h$ for $\delta = 0.5991$, and the tails terminated at 10% of the maximum height. 14
2.4 The mean height of an exponential-type shape $\frac{t}{\epsilon}e^{\frac{-t}{\epsilon}+1}$ is $\alpha = 0.5337h$ for $\epsilon = 0.5116$, and the tail terminated at 10% of the maximum height. 14
2.5 The mean height of a triangle is $\alpha = \frac{1}{2}h$. 14
2.6 Points y_i on a parabolic wire are averaged to determine the mean height of a parabola. This mean is calculated to be $\frac{2}{3}h$. 15
2.7 Overlapping parabolas of constant height, baseline and distance between parabolas, sum to a mean and range around that mean. 16
2.8 The correspondence between the size of the standard deviation of the heights of the parabola and n, the number of overlapping parabolas. The sampling distribution has no effect on n. 18
2.9 The correspondence between the size of the standard deviation of the distance between starting times of the parabolas and n, the number of overlapping parabolas. The sampling distribution has a maximum error of 2.2% on n. . . 18
2.10 Intermittent parabolas allow for a large range $R \sim H_{max}$ while the mean of the sum over parabolas is decreased by the delay. 19
3.1 The solution to Nowak and May's equations (32)-(33), with arbitrary constants. 20
3.2 The solution to Perelson and Nelson's equations (34)-(36), with arbitrary constants. 21
3.3 The solution to our equations with the initial constant values. 23
3.4 The solution to our equations with $V(0) = 1$ (solid line) and $V(0) = 20$ (dashed line). 23
3.5 The solution to our equations with $X(0) = 0$ (solid line) and $X(0) = 10$ (dashed line). 24

3.6 The solution to our equations with $T^*(0) = 1$ (solid line) and $T^*(0) = 20$ (dashed line). 24
3.7 The solution to our equations with $T(0) = 0$ (solid line) and $T(0) = 20$ (dashed line). 25
3.8 The solution to our equations with $r = 5$ (solid line) and $r = 15$ (dashed line). 25
3.9 The solution to our equations with $d = 0.5$ (solid line) and $d = 1.5$ (dashed line). 26
3.10 The solution to our equations with $s_X = 0.1$ (solid line) and $s_X = 3$ (dashed line). 26
3.11 The solution to our equations with $c = 0.4$ (solid line) and $c = 1.2$ (dashed line). 27
3.12 The solution to our equations with $b = 0.01$ (solid line) and $b = 0.5$ (dashed line). 27
3.13 The solution to our equations with $k = 0.5$ (solid line) and $k = 1.5$ (dashed line). 28
3.14 The solution to our equations with $\delta = 1$ (solid line) and $\delta = 2$ (dashed line). 28
3.15 The solution to our equations with $p = 10$ (solid line) and $p = 500$ (dashed line). 29
3.16 The solution to our equations with $T_{max} = 1$ (solid line) and $T_{max} = 10$ (dashed line). 29
3.17 The solution to our equations with $d_T = 10$ (solid line) and $d_T = 50$ (dashed line). 30
3.18 The solution to our equations with $s_T = 0.1$ (solid line) and $s_T = 10$ (dashed line). 30
3.19 The solution curves to equations (37)-(40) with dimensions (using constants (77)-(88) and (93)-(96)). 34
4.1 Histograms of the n^{VL} and $\overline{\tau_L}^{VL}$ values for viral load data of MACS patients not on treatment. $n^{VL} < 1$ for 93% of patients and $\overline{\tau_L}^{VL} > 0$ for 98% of patients. 37
4.2 For untreated HIV-1 MACS data, observed mean and range gives $\overline{\tau_L}^{VL} = 1$. . 37
4.3 Histograms of the n^{CD} and $\overline{\tau_L}^{CD}$ values for CD4+ count data of MACS patients not on treatment. $n^{CD} > 1$ in 85% of patients and $\overline{\tau_L}^{CD} < 0$ in 61% of patients. 40
4.4 For untreated CD4+ MACS data, observed mean and range show $n^{CD} = 1.4$ overlapping parabolas. 40
4.5 The phasing between viral load and infected CD4+ cell solutions to ODE (37)-(40). (Using non-dimensionalised variables.) 42
4.6 Some possible phasings between viral load and CD4+ data. Time is non-dimensional and the heights of each parabola are arbitrary. 42
4.7 Toy histograms for n values in various possible scenarios. The left hand graph would represent a dataset such as MACS viral load data. The right hand graph would represent a dataset such as Henrard *et al*'s [18] trial. The middle graph would be some combination between the two datasets. 43
4.8 A series of parabolas depicting the window period. A series of parabolas with declining heights would represent a steady decay to the plateau, or one parabola of large height followed by regular parabolas of smaller height would represent a sudden change in "set point". 44

4.9 A series of parabolas depicting possible viral load end phase. Increasing heights of parabolas, in (a) growth to represent an ever-increasing viral load, or (b) step increase to represent a new "set point" occurring due to some biological change. 45
4.10 A series of parabolas depicting possible viral load end phase. Increasing baselines of parabolas, in (a) growth to represent an ever-increasing viral load, or (b) step increase to represent a new "set point" occurring due to some biological change. 45
4.11 A series of parabolas depicting possible viral load end phase. Decreasing distances between starting times of parabolas, in (a) growth to represent an ever-increasing viral load, or (b) step increase to represent a new "set point" occurring due to some biological change. 45
4.12 The graphs of the parabola, gaussian and exponential-type curve for CD4+. Each have the same area under the curve when terminated at height 10% of the maximum. The agreement between these curves is good. 48
4.13 Gaussians and exponential-type curves sum to reach observed mean and range for MACS CD4+ data for patients not on treatment. The agreement between these profiles and the parabola overlap model is good. 49

List of Tables

1.1 An extract from the MACS treatment data shows large inconsistencies, with more than one treatment value per visit. The treatment value 1 means the patient is not on treatment, 2 means the patient is on some antiretroviral and "." means treatment value is unknown. 9
1.2 The treatment data is reduced so that there are no duplicates of treatment data for one patient visit. 10
1.3 Default viral load values of 300 or 40 (depending on assay used) are displayed to show that viral load was undetectable. Patient 1018 shows a drop in viral load to these default values, presumably owing to antiretrovirals. 10
4.1 The statistics of viral load for MACS data, where P is the number of patients per group, # is the mean number of data points per patient in this group, M is the mean viral load, R is the range of viral load, and σ is the standard deviation of the viral load. All viral load data is measured in 10^4 copies/ml. A patient was only taken to belong to a particular category if they had more than 5 points in that category. Group A removed entries where viral load was 300 and group B removed entries where viral loads were 40 or 300. The intermittent model is clearly favoured for viral load data. 36
4.2 There are 7 patients in the MACS viral load data of patients not on treatment where both the overlap and intermittent model apply. For these patients, either $\overline{\tau_L}^{VL}$ is close to 0 or n^{VL} is close to 1. 38
4.3 The statistics of CD4+ count for MACS data, where P is the number of patients per group, # is the mean number of data points per patient in this group, M is the mean CD4+ count, R is the range of CD4+ count, and σ is the standard deviation of the CD4+ count. All CD4+ data is measured in 10^6 copies/l. A patient was only taken to belong to a particular category if they had more than 5 points in that category. Group A removed entries where viral load was 300 and group B removed entries where viral loads were 40 or 300. The overlap model is clearly favoured for CD4+ data. 39

1 Introduction

1.1 Background

It is well-known that HIV-1 has the strong characteristic of rapid mutation rates, with a point mutation occurring every replication cycle [34]. However, some mutations are "fitter" than others in their reproduction rate, and it has been found that only a tiny fraction of cells infected by HIV-1 are able to replicate [6]. It is only of interest to look at the mutations of HIV-1 which are able to do damage to the body, thus it is of some importance to determine, at least, the mutation rate of viable virus. For example, if the number is small, the genetics of these "viable strains" may inform vaccine design.

HIV-1 is an RNA retrovirus, because its genome encodes an enzyme called reverse transcriptase, which allows transcription from RNA to DNA, which is the opposite order of normal transcription [33]. The reverse transciptase also allows for a high level of errors to occur in replication, thus resulting in the high mutation rate of the virus [34]. There are two major types of HIV, HIV-1 and HIV-2. We are interested particularly in HIV-1, since this type is responsible for the majority of HIV infections [41]. HIV-1 primarily infects activated CD4+ cells, and since parts of the immune system are dependent on these cells [22], HIV-1 handicaps the body's defenses against it. It is for this reason that this dissertation will take into consideration CD4+ cells as well as numbers of free viral RNA.

The immune system has several cells with which it counters attacks on the body. The spearheads of the immune system are the T-cells and B-cells [22][38] (the names refer to where the cells are created, as T-cells are created in the thymus and B-cells are created in bone marrow [38]). T-cells are split into two main groups, the CD8+ cells (also known as killer T-cells or CTLs) and the CD4+ cells (also known as helper T-cells), both have receptors that recognise virus particles as being a threat and are able to neutralise infected cells or trigger apoptosis (cell suicide) [16]. An infected cell displays peptides on the cell surface via a molecule called the Major Histocompatability Complex (MHC) [23][38] and groups of these peptides form epitopes which are then recognized by T-cells. An epitope is usually 9 peptides long, but can range from 8-11 peptides in length [17]. There are 216 known HIV-1 epitopes that elicit a response from the CTLs [23].

It is now known that the apparent equilibrium of the asymptomatic phase actually represents a rapid replication and clearance of the virus [19][43], and it has been found that the level of HIV-1 RNA in plasma is a good representative of the mutation rate of the virus occurring in the body [11]. The diversity of the virus in the body is central to the treatment of this disease [10].

1.2 The Existence of a Set-Point

The notion of a "set point" in the asymptomatic period of untreated HIV-1 progression is accepted in the literature. Feinburg (1996) [11] and Langford (2007) [22] define this as a long term stable mean of viral load. The viral load of a patient is used as an important marker to determine a patient's progression to AIDS [18] and often, a single measurement is thought to be representative of the set point.

Henrard *et al* [18] is widely cited for his group's study into HIV-1 RNA levels for as long as 11 years. The result from this study was that viral load remained stable over the 3-11 year follow up time, with occasional 10-fold and sometimes 100-fold increases in viral load [18]. Lindbäck *et al* [24] discovered that this steady-state level was reached a median of 2 months after infection with very small rate of change thereafter.

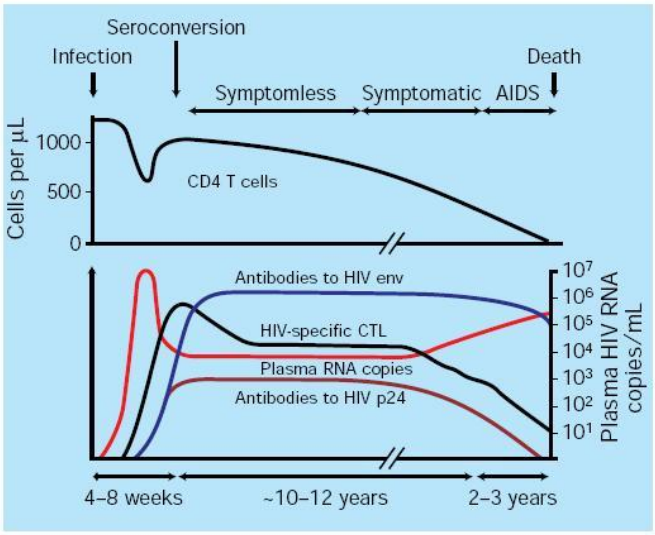


Figure 1.1: (Taken from [11]) A schematic graph of the progression of various biological measurements for an untreated HIV-1 infected patient. We focus on the line of "Plasma RNA copies" in the bottom graph representing viral load progression and the top graph showing the behaviour of the CD4+ count.

Results such as these encourage mathematical models that develop equilibrium viral load on time scales of overall disease progression in untreated patients [28]. The importance is that the equilibrium models might approximate reality. This is epitomised by the long plateau of plasma RNA copies in the schematic diagram of Figure 1.1. Yet it is equally well known that viral RNA concentrations in blood can fluctuate between 10^3 and 10^6 copies/ml [3] [7] [8] [21] [25], which casts doubt on the meaningfulness of equilibrium viral load. Henrard *et al* [18] noted these fluctuations in some patients as well. In the MACS dataset studied here [27], we find that 98% of all 754 patients with more than ten recorded datapoints (treated and untreated) display fluctuations of free virus with range more than twice their mean. The data shows, for viral load

$$H_{min} \sim 10^3, H_{max} \sim 10^5 \tag{1}$$

where the above are measured in copies/ml and are averaged over all patients.

Studies done by Fraser *et al* [13], Vidal *et al* [42] and Schacker *et al* [40] find large variation between and within patients and Vidal *et al* find no evidence of a set point at all during chronic infection. However, a significant increase in viral load in Vidal's study [42] was considered to be an increase $> 0.5log_{10}$ of baseline viral load, which occurred in 19% of this trial. Contrastingly, Henrard's trial considered the viral loads to be stable because only 14% of his trial had increases of $1log_{10}$ (i.e. $|v_{mean} - v_i| > 1$ where v is measured on a log scale) [18]. As such, the data in both cases could be similar, with only the definition of a stable set point differing in each case.

Figure 1.2 clearly shows the high level of variability in viral load in one patient as part

of the MACS dataset [27]. It should be noted that both Henrard *et al* [18] and Vidal *et al* [42] would find no stable set point in this particular patient.

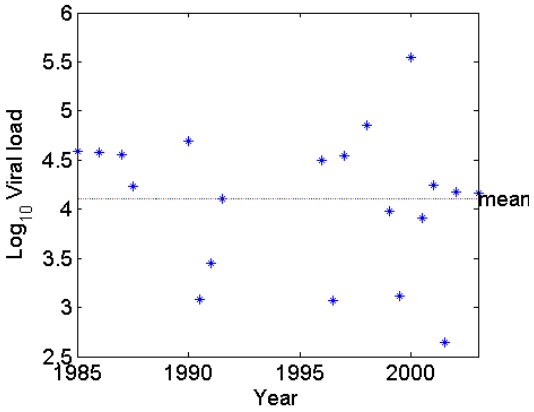


Figure 1.2: The viral load data points of one MACS patient. This data shows enormous amounts of variability, with the range significantly larger than the mean. This contrasts with the idea of a set point in the sense of equilibrium viral load.

The fluctuations seen in the MACS data are dramatic and ask that the fluctuations themselves be understood. In this dissertation I will discuss models with a stable mean, but without equilibrium, that account for variability in MACS data.

1.3 Review of Mathematical HIV-1 models

Since HIV was discovered in the early 1980's, scientists have been searching for a cure. Much was not known about the virus, and the first big breakthroughs came in 1995, where the development of antiretrovirals allowed Wei [43] and Ho [20], to discover that the so called "plateau" in viral load actually concealed a very high replication rate of HIV. This high replication rate was also explored by Perelson [36] (1996) and many others from then onwards.

The early antiretrovirals proved unsuccessful in long-term control of the virus, and control was achieved in 1996 in the combination of 3 separate antiretrovirals in the initiation of HAART (Highly Active Anti-Retroviral Therapy) [4]. This approach was only able to control HIV, not eliminate it, and more than 20 years later, we are still unable to eliminate the virus. In the 20 years, many mathematical models have been developed to explain the dynamics of HIV. They have covered many different areas and applications of the virus and together they have increased our understanding of HIV-1 [29] [44].

Without any immune system response, the dynamics between different viral populations would be simple - the fittest population will survive [29] [31]. Overbaugh *et al* [34] found two forces that determine which virus mutations survive - number of host cells available, and pressure from the immune system. Thus it is the presence of the immune system that results in competition between the viral populations, as the fastest reproducing viral population may be more of a target for the CTL responses [29]. As such, the efficient mutations are those with a neutral fitness value [39]. The immune system response is continuously evolving [15] [45] and so the dominant viral population that is able to replicate quickly and also evade

these responses must change as well. The evolution of the dominant viral population is examined in greater detail by Ball *et al* in a very recent paper [2], but the time scale of each dominant viral population is in doubt.

These ideas led to the concept of a quasi-species - "populations of closely related but distinct viral genomes" [31]. From this idea, and using their notation, Nowak, May and Anderson (1990) [31] developed a model

$$\dot{v_i} = v_i(r - sz - px_i), i = 1, ..., n \tag{2}$$

where v_i represents viral population i, x_i represents an immune system response aimed at viral population i, and z represents a general immune system response aimed at all viral populations. This basic idea is built upon in the paper [31] and in further papers [30] [32] and many others, summarised in a book [29]. The interesting part of the model is that mutants v_i are created with a certain probability over a given time interval. Thus we have birth of new strains throughout infection.

These models display oscillations and, by choice of constants, the viral load may reach an equilibrium, as in the model from [29] (p.128)

$$\begin{aligned} \dot{v_i} &= v_i(r - px_i - qz), i = 1, ..., n & (3)\\ \dot{x_i} &= cv_i - bx_i, i = 1, ..., n & (4)\\ \dot{z} &= kv - bz & (5) \end{aligned}$$

where

$$v = \sum_{i=1}^{n} v_i.$$

Here v_i, x_i and z are as discussed above, where the general immune system response is aimed at the sum over all viral populations. With choice of constants, this leads to an equilibrium in total viral load, as seen in Figure 1.3. We do not find evidence of equilibrium in MACS data.

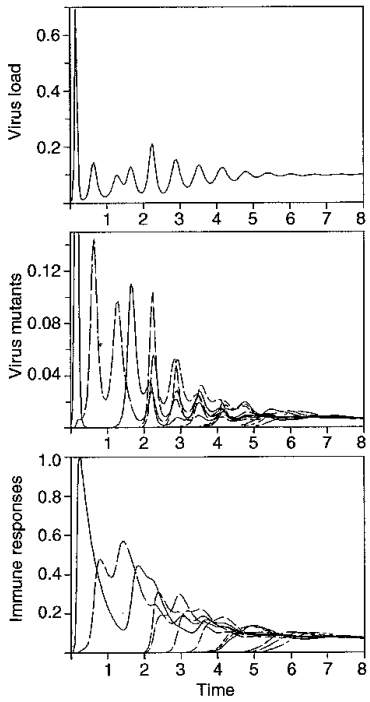
Figure 1.4 shows the solution curves for Nowak and May's [29] equation (14.1) as follows (where all indices are free indices and there is no summation over repeated indices):

$$\begin{aligned} \dot{v}_{ij} &= v_{ij}(r - px_i - qy_j) & (6)\\ \dot{x}_i &= x_i(cv_{i*} - b) & (7)\\ \dot{y}_j &= y_j(kv_{*j} - b) & (8) \end{aligned}$$

$$\text{with } i = 1, ..., n_1,\ j = 1, ..., n_2,$$

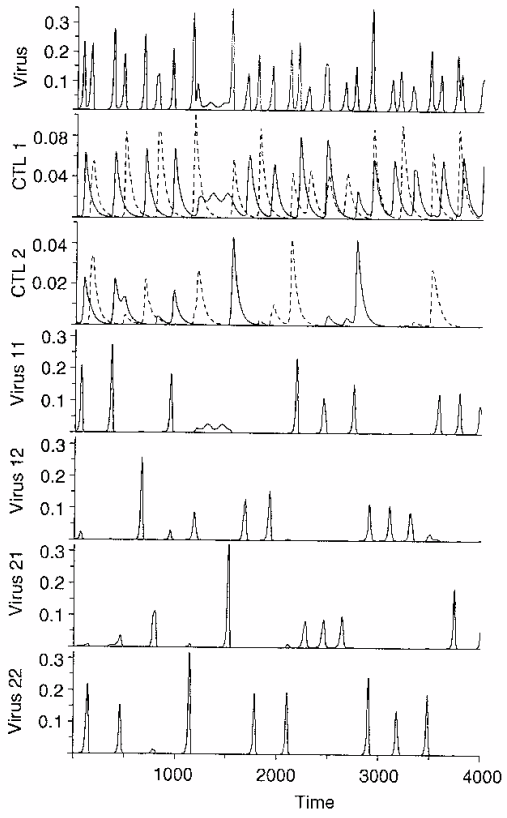
$$v_{i*} = \sum_{j=1}^{n_2} v_{ij}$$

$$v_{*j} = \sum_{i=1}^{n_1} v_{ij}$$



Antigenic variation in the presence of strain-specific and cross-reactive immune responses. The simulation begins with a single viral strain, which induces both strain-specific and cross-reactive immune responses. Subsequent strains escape from the strain-specific response, but not from the cross-reactive response. The equilibrium virus load is an increasing function of antigenic diversity, but saturates at high levels of antigenic diversity. The simulation is based on eqn (12.4) with the parameter values $r = 2.5$, $p = 2$, $q = 2.4$, $c = k = 1$, $b = 0.1$. The probability that a new mutant arises in the time interval $[t, t + dt]$ is given by Pdt with $P = 0.1$.

Figure 1.3: (Taken from Virus Dynamics [29] p.138) The solution curves to Nowak's differential equations (12.4). Each strain is born with probability and reaches an equilibrium.



Computer simulation of the basic model given by eqn (14.1). There are two different epitopes with two sequence variants in each epitope (thus altogether two different virus species). All virus mutants replicate at the same rate and are present at the beginning ($t = 0$) in different abundances. There is no subsequent production of new antigenic material. Nevertheless, we observe sequential peaks in viral abundance that correspond to antigenically different variants. Thus antigenic oscillations can occur without antigenic drift. As discussed in the text, we have a clear understanding of the long-term behaviour of the system: the y_i will converge to zero, and there are undamped neutral oscillations with the x_i and v_{i*}. The parameters are: $n_1 = n_2 = 2$, $r = 0.1$, $p = 5$, $c = 1.1$, $k = 1$, $b = 0.02$. The time axes is in arbitrary units (but the biological observations suggest these oscillations to occur on a time-scale of weeks or months).

Figure 1.4: (Taken from Virus Dynamics [29] p.158) The solution curves to Nowak's differential equations (14.1).

This is a fairly complex system, where each virus particle v_{ij} displays two epitopes, A and B, with sequence i in epitope A and sequence j in epitope B. There are n_1 possible sequences for epitope A and n_2 possible sequences for eptiope B. There are also CTL responses x_i directed at sequence i of epitope A and y_j directed at sequence j of epitope B. In this model, all virus particles reproduce at rate r and are killed by CTL responses at rates p and q. The CTLs have a fixed death rate b and are stimulated to grow by the presence of all virus particles. As in Figure 1.3, the total virus levels are obtained by adding over the different viral mutations. Each v_{ij} is defined to be a separate viral strain. [29]

This model describes the dynamics of many interacting populations of the virus, with a time scale of weeks or months per population. No new populations are born, there are fixed possible populations with one dominating for a short time period, before another population recovers. The interest in the model is the apparently sudden outburst of a strain, but most importantly, it is clear that if this total viral load is randomly sampled, the range of data can exceed the mean. This picture is consistent with MACS data (see Figure 1.2).

Others have developed influential models as well, such as Perelson and Nelson [35] who looked at very simple models to describe HIV-1 behaviour, and Callaway [5] who modelled HIV-1 behaviour under HAART. This last paper looks directly at steady state viral load during treatment with HAART, but a steady state is perturbed by viral blips [9]. We again propose that these perturbations be examined more closely, rather than focus on the steady state. Blips are modelled by Di Mascio *et al* [9] and Figure 1.5 shows the presumed shape and size of a viral blip during HAART.

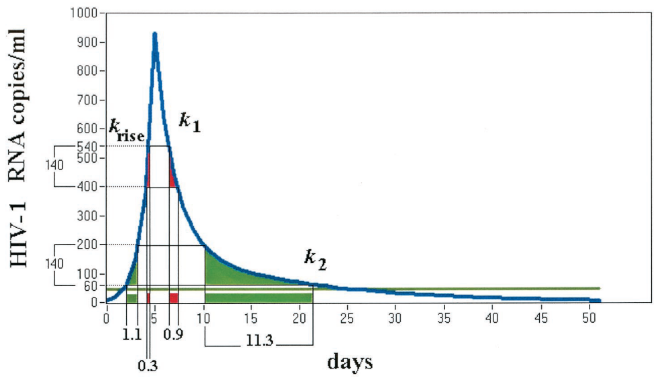
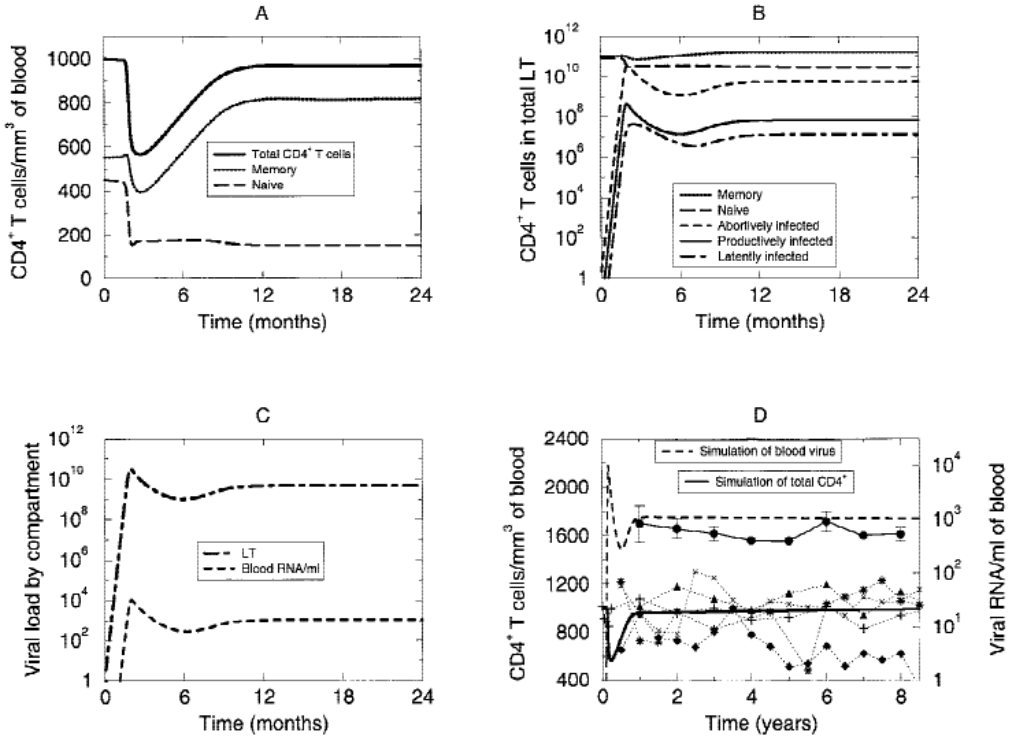


Figure 1.5: (Taken from [9]) The behavior of viral load that occurs in a patient on HAART during a blip. The time scale of this profile is similar to that of the window period, and will be used as an approximation for the lifetime of a viral population.

Wu and Zhang [44] use a semi-parametric model and incorporate individual levels of variation in viral load. However this paper is interested in modelling HIV-1 dynamics under HAART over patient lifetime and cannot capture the fluctuations of Figure 1.2.

Bajaria *et al* [1] develop a sophisticated, modern model to describe HIV-1 behaviour, with 9 coupled differential equations and 16 constants. These equations model latently infected cells, failed infections of a cell, infected cells that are able to replicate, naive vs memory T-cells, movement of cells between blood and lymph, effects of treatment, among others. This model comprehensively covers many aspects of HIV-1 infection, but again this model

focuses on time scales of order years and on set point, as seen in Figure 1.6.



Long-term nonprogressor (LTNP) model results. **(A–C)** Acute and asymptomatic stages for cell classes in blood, cells in lymphatic tissues (LT), and virus in both LT and blood, respectively. **(D)** Comparison of our model solution for longer than 8 years of infection with data from Pantaleo et al. (45), Greenough et al. (46), and Fauci et al. (47) for CD4+ counts (◆, +, ×, ▲, ✱) and from Pantaleo et al. (45) for viral load (●) in the typical LTNP. The parameters and initial values are given in Table 2 in the Appendix (see Table 3 for further comparison with data).

Figure 1.6: (Taken from [1]) Figure D shows the long time scales of Bajaria *et al*'s model and displays the set point. We find large variation around this set point and we believe this variation should be exploited further.

It will be of interest to examine the short-term dynamics of different HIV-1 populations, on time scales of weeks to a month or so. These dynamics are hidden in long-term models on the overall scale of a patient lifetime.

1.4 The MACS Data [27]

A total of 4,954 homosexual and bisexual men voluntarily participated since the beginning of the MACS study in 1984. They were followed up semi-annually. In each visit the data sets of physical examination, questionnaires and laboratory results were generated. From April 1987 to September 1991, patients such as partners of the original cohort were targeted. This release of the data terminates in 2003.

In order to preserve patient confidentiality, only the year of visit and a visit number were given. As such, each measurement was ordered by the visit number and our results are

independent of measurement within a year. Only patients with 10 or more recorded viral load and CD4+ count measurements were considered in this dissertation.

The first treatment entries were given in 1992, and the anti-retroviral questionnaire was only introduced at visit 13 (1990 if visit one was at the beginning of 1984 when the trial started and visits took place semi-annually every year). Unfortunately, there were not enough patients confirmed on treatment to make any conclusions for that group. The data may be grouped according to: 1984-1992 (before treatment was recorded); and 1992-2003 (treatment may be indicated).

1.4.1 Inconsistencies

In the treatment data file there are many inconsistencies with more than one anti-viral value per visit, as seen in a particular patient below:

Patient ID	Visit number	Year	Treatment value
1857	200	1993	1
1857	200	1993	1
1857	200	1993	1
1857	210	1994	1
1857	210	1994	1
1857	210	1994	.
1857	220	1994	.
1857	220	1994	1
1857	220	1994	.
1857	230	1995	2
1857	230	1995	1
1857	230	1995	2
1857	240	1995	1
1857	240	1995	1
1857	240	1995	.

Table 1.1: An extract from the MACS treatment data shows large inconsistencies, with more than one treatment value per visit. The treatment value 1 means the patient is not on treatment, 2 means the patient is on some antiretroviral and "." means treatment value is unknown.

Note that there were many repeated entries for the same visit and patient. If all these repeated entries showed the same treatment value, the duplicates were simply eliminated. If these repeated entries showed contradicting treatment values, the treatment value at that visit was taken to be unknown. If repeated entries showed a contradiction between known and unknown treatment plan, the known treatment plan was taken as the treatment value only if that value was confirmed in another repeated entry for that visit. The data from Table 1.1 was processed into Table 1.2 for our data analysis.

1.4.2 Errors on Laboratory Measurements

Data from the MACS trial ranged from 1984-2003, over which time the quality of viral RNA assays have improved. Mellors *et al* used the MACS data in 1995 (where patients

Patient ID	Visit number	Year	Treatment value
1857	200	1993	1
1857	210	1994	1
1857	220	1994	.
1857	230	1995	.
1857	240	1995	1

Table 1.2: The treatment data is reduced so that there are no duplicates of treatment data for one patient visit.

had maximum follow-up time of 8.3 years) and determined the coefficient of variation of HIV-1 RNA to be 11.4% with the Quantiplex RNA assay [26] (this assay is based on bDNA signal amplification). Also in 1995, Henrard *et al* cites errors of 10^2 copies/ml for PCR amplification (in a different data set) [18]. Giorgio *et al* [14] examine MACS patients whose blood specimens were sent to various different laboratories between 1985-1988 to check for consistency in T-cell results. The maximum coefficient of variation between these laboratories for CD4+ cell results was 11.9%.

1.4.3 Default Viral Load Values

Patient ID	Year	CD4+ count	Viral load
1018	1993	560	1199
1018	1995	537	300
1018	1996	496	300
1018	1997	546	40
1018	1997	476	40
1018	1998	443	40
1018	1998	532	40
1018	1999	475	40
1018	1999	515	40
1018	2000	537	40
1018	2000	655	40
1018	2001	484	90

Table 1.3: Default viral load values of 300 or 40 (depending on assay used) are displayed to show that viral load was undetectable. Patient 1018 shows a drop in viral load to these default values, presumably owing to antiretrovirals.

Two groups are defined: Group A removed values of 300; and Group B removed values of 300 and 40. These viral load values are indicative of treatment and will be a useful control in our analysis below. Previously discussed inconsistencies in the treatment file cast doubt on these values, thus a patient who is declared untreated may skew our results in the untreated category if they display constant undetectable levels of virus (presumably owing to treatment).

2 The Models

2.1 Objectives

The aim of this dissertation is to explore the dynamics of viable HIV-1 populations in vivo, for untreated patients. This will be related to patients on antiretrovirals. It is of particular interest to find the average number of viable HIV-1 in vivo at a particular time, especially during the asymptomatic phase of infection. "Viable" is defined to be the viral populations able to replicate and able to achieve a certain population size. Now there are a few ways to catagorise these viral populations. The first and perhaps most obvious is to group together all viral RNA with the same gene sequence, supposing that the fitness of the virus is due to the viral structure. Some parts of the viral RNA sequence encode specific tasks, such as the protein *env* (a section of the HIV RNA which codes for the envelope protein), which contains a region called the V3 loop, about 30 amino acids long, which seems to be able to counter some immune system responses. The *env* gene is essential to the binding of a free virion to a host cell and seems to have a high mutation rate, which could lead to the virus being insusceptible to immune system responses [28][29]. To my knowledge there is no published data showing exactly which regions of the HIV-1 RNA are essential to the functionality of the virus, although it appears that only about one third of the HIV-1 gene is invariant [34]. Then viable virus could be characterised by this invariant part.

The second catagorisation of "viable" has to do with how the body recognises an infected cell as being a threat. As mentioned above, the T-cells recognise infected cells by the epitope they display. The body is capable of recognising more than one epitope as a threat simultaneously, but it may not recognise the epitopes of all infected cells [29]. The immune system is able to adapt over time, with more epitopes being recognised [45]. There are also subtypes of killer and helper T-cells called memory T-cells. These may live for years or even decades [38] and are responsible for fast immune responses to secondary infection (as seen in diseases such as measles). However, these memory CD4+ cells are a target for HIV-1 infection and the killer T-cell memory response seems to depend on these memory CD4+ cells [29], which decline during the course of infection (see Figure 1.1). Yet, viable HIV-1 could be characterised by the related sets of one or more epitopes.

We will assume here that MACS viral load and CD4+ measurement are simply the sum of such variants, respectively. We will seek to model the totals and will refer simply to HIV-1 and CD4+ populations. There is preliminary evidence of viral load by gene sequence [10], but we remain far from a detailed picture of quantitative viral load for sequence in a phylogenetic tree. The importance of such a picture is that the invariant parts of the HIV-1 gene could be identified; also, that the genetic characteristics of successful strains could be identified and the characteristics of particular gene alterations that lead to a successful strain could be identified. This last could be regarded as a "trigger" for onset of the new viable strain.

Yet we are far from this information. In particular, phylogenetic trees do not resolve accurately the time of onset of such strain. Thus we have the information that every base pair might mutate in one replication cycle [34], which sets a lower time interval between blood samples for sequencing. Contrasting with this, there is evidence that the genetic information of viral RNA in latent CD4+ cells is stable over periods of years [12], which determines the ideal duration of daily sampling. We are interested that the observed, extreme fluctuations in viral load are owing to the emergence of distinct populations, each triggered by some (unknown) mechanism. Because strong fluctuations are observed, each of these must decay.

We will suppose in particular that each population has a time scale of a few weeks to a month or so.

That viral populations can exist on this time scale is evidenced by the period of initial acute viremia before the "set point" emerges [24] and periods of transient viremia during HAART [9]. In the latter, blips are expressed within approximately a month, thus our upper limit of a month or so is justified. The landmark determination by Ho *et al* [20] and Wei *et al* [43] of free virus lifetime of order hours, with the help of mathematical models on time scales of days, remains valid if our model is taken to develop over timescales of many days. Indeed, the repeatability of these experiments suggests that natural change in viral load is not significant over days and thereby sets a lower bound for our timescales.

In the mathematical models used below, we will not presume the exact nature of the viable populations. These populations are a group of virus mutations that grow and die together, regardless of the structures involved. The only restriction we place on these populations is that they should have a lifetime of weeks to a month or so, as discussed above. It is possible that other, distinct, populations may be long lived, as perhaps in latently infected CD4+ cells [12].

An original contribution of this dissertation is the nature of the mathematical model proposed. In contrast to the standard population dynamics explored in for example Bajaria *et al* [1], whose models hold over a timespan of years, this dissertation will use the biological timespan of weeks to months as seen in blips and the window period [9] [11]. The dramatic mutation rates of the HIV-1 virus will be exploited, and total viral load is proposed to be a sum over viral populations of quite short life. The latter makes this model quite distinct, even compared to the model of Figure 1.4, where any short-lived viral outburst is seen to recur at later times, that is, a strain persists for the lifetime of the patient.

The primary objective of this dissertation is to use such models to determine the number of viral populations at any moment and their amplitudes in MACS patients.

2.2 Primary Assumptions

The primary assumption is that the time scale of the window period as seen in Figure 1.1 is representative of the lifetime of the initial HIV-1 populations, and therefore of the time scale for the immune system to significantly clear those viral populations. This model assumes that not only is the window period representative of the first viral populations, but of every viral population. It is interesting to note that the blips that have been reported in many patients currently on HAART have a time scale of roughly 20-30 days [9] (similar to the time scale of the window period). This supports the primary assumption, as the blip would represent a single viral population that has temporarily escaped the effects of the treatment to rise to a high viral load before it is once again controlled.

The second assumption is that the immune system can in principle target multiple viral populations at any one time, which has been confirmed as possible [29].

Restrictions will be placed on viable populations so that only those populations with significant amplitude are considered, in this dissertation viable populations must clearly exceed H_{min} (1). This model will then ignore certain low-level populations.

Our model will suppose that the rise and fall of the CD4+ and HIV-1 viral populations are adequately modeled by simple profiles with the above-mentioned life time. Finally, this model will hold throughout the asymptomatic period (Figure 1.1).

Of course these primary assumptions will be tested in the modelling. It is known that patients on ART and who might have ignorable viral loads, always suffer a recovery to full

viral loads if they default; thus insignificant viral populations with reasonable life time, may mutate to eventually give "viable strains" (in the sense of this model). This too may be modelled by our methods, but an essential aim of this investigation is to determine whether the primary assumptions are supported by the results. In all cases, the mathematical models must be interpreted in physiological terms.

2.3 Methods

Population dynamics at its most detailed, comprises sets of differential equations for the various inter-related populations. Recalling the models reviewed in Section 1.3, the simplest of them involves equations (3)-(5) with 7 coupling constants, which are not easily determined. The most sophisticated of them (for example Bajaria *et al* [1]) involves 9 coupled equations in 16 constants. It is a strength of the breakthrough mathematical models [19] and [43], where viral lifetime $\sim$ 6 hours was discovered, that they comprised a few, easily measurable coefficients. Here we aim to build the simplest possible model with easily measured parameters. To be specific, we will model growth and decay of HIV and CD4+ dynamics by approximating each population by a simple profile. The parameters of the model will be the mean, maximum and range of patient data.

Several profiles could be considered. In Figures 2.1-2.5 we examine various possible profiles and we are satisfied that in all of these cases, the mean height$\geq 0.5h$, where h is the maximum height of the profile.

To illustrate the calculation of the mean height of a parabola, we look at a parabolic wire, shown in Figure 2.6 and given by $y(x) = -c(x-a)(x-b)$. Taking N, the number of points to be $N = \frac{b-a}{\Delta x}$ and x_i to be $x_i = a + i\Delta x$, we find the average height over points on the wire as follows:

$$\begin{aligned}
\overline{y} &= \frac{\sum_{i=1}^{N} y_i}{N} && (9)\\
&= \overset{lim}{\Delta x \to 0} \frac{\sum_{i=1}^{(b-a)/\Delta x} c(x_i - a)(x_i - b)}{(b-a)/\Delta x} && (10)\\
&= \frac{c}{b-a} \overset{lim}{\Delta x \to 0} \frac{(b-a)(b-a+\Delta x)(b-a-\Delta x)\Delta x}{6\Delta x} && (11)\\
&= \frac{c}{6}(b-a)^2 && (12)\\
y_{max} &= \frac{c}{4}(b-a)^2 && (13)\\
\Rightarrow \overline{y} &= \frac{2}{3} y_{max}. && (14)
\end{aligned}$$

Of all these profiles, we choose to use parabolas in our model since they are very simple profiles with algebraic representations for sums over many parabolas. This will allow exact solutions for variables in our model. They are also comparable with other curved biological models, as shown below. Once the parabola analysis is complete, we will determine the effect of choosing this profile over others (see Section 4.6).

Note that in Figures 2.1-2.5, profiles with tails are terminated at height 10% of the maximum height. This termination height is due to the relationship between the mean and minimum values in MACS viral load data (see Section 4.1). For the profiles with tails, the

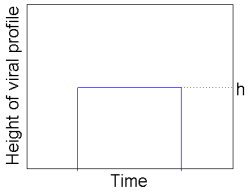


Figure 2.1: The mean height of a square is $\alpha = h$.

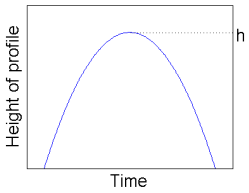


Figure 2.2: The mean height of a parabola is $\alpha = \frac{2}{3}h$.

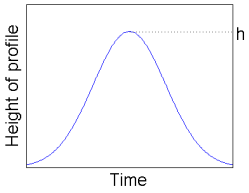


Figure 2.3: The mean height of a gaussian $e^{\frac{-t^2}{\delta}}$ is $\alpha = 0.5652h$ for $\delta = 0.5991$, and the tails terminated at 10% of the maximum height.

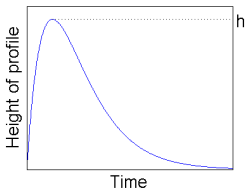


Figure 2.4: The mean height of an exponential-type shape $\frac{t}{\epsilon}e^{\frac{-t}{\epsilon}+1}$ is $\alpha = 0.5337h$ for $\epsilon = 0.5116$, and the tail terminated at 10% of the maximum height.

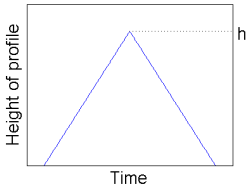


Figure 2.5: The mean height of a triangle is $\alpha = \frac{1}{2}h$.

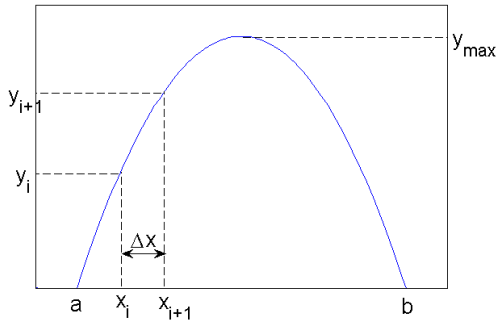


Figure 2.6: Points y_i on a parabolic wire are averaged to determine the mean height of a parabola. This mean is calculated to be $\frac{2}{3}h$.

values for ϵ and δ are determined so that the area under the graph is comparable to the area under the parabola shown (see Section 4.6).

Taking into account the observed amplitude and fluctuation in the viral load, mathematical and statistical techniques will be used to estimate the amplitude and the number of viral populations at any moment in plasma. Various scenarios are considered. For example, the parabolas may overlap, or, the switch-on of a parabola may be delayed well beyond the decay of the previous population. The case may be considered where each parabola has the same height and the parabolas are evenly spaced. Simple analytic results are possible in this case. Extending the theory, a computer program can be designed to simulate the random occurrence of viable populations in the body by generating parabolas with randomly distributed heights and starting times.

Further, the same technique may be applied to CD4+ data. It will be determined if this method gives us comparable results for viral load and CD4+ data.

We will find it useful to create a simple model for a single viable population in the form of a set of ordinary differential equations. The behaviour of these models will provide insight into the dynamics of our parabola models. This will be discussed in Section 3.

2.4 The Overlap Model

Suppose each parabola is a viable HIV-1 population, and that more than one viral population exists simultaneously. The parabola is defined by

$$H_i(t) = -\frac{h_i}{\tau*^2}(t - t_i)(t - (t_i + \tau)), \tag{15}$$

where $H_i(t)$ depicts the i^{th} viral population, t_i is the starting time of the i^{th} population, τ is the baseline length of the parabola, and $\tau*$ is a convenient non-dimensionalising timescale.

Choosing $\tau* = \tau/2$ we find that

$$H_i(\overline{t}) = -h_i(\overline{t} - \overline{t_i})(\overline{t} - (\overline{t_i} + 2)), \tag{16}$$

where the bars denote normalisation, and

$$h_i = (H_{max})_i \tag{17}$$

is the maximum viral load of the viral population. Then the observed viral load at any normalised time $\bar{t}$ may in principle be the sum over all viral populations, or in our model, a sum over parabolas. We will conveniently drop all bars so that t denotes non-dimensional time. H_i and h_i will be measured in RNA copies/ml.

Suppose that all parabolas are equal and that the switch-on time is constant. Then $t_i = i\Delta t$, $\tau = n\Delta t$, (where Δt is the constant gap between starting times of parabolas) and $h_i = h$. This sum (Figure 2.7) is given by

$$H(t) = \sum_{i=1}^{n} H_i(t) \tag{18}$$

$$= -\frac{1}{6}hn(\Delta t^2 + 6\Delta t^2 n + 5\Delta t^2 n^2 - 6t\Delta t - 12t\Delta t n + 6t^2). \tag{19}$$

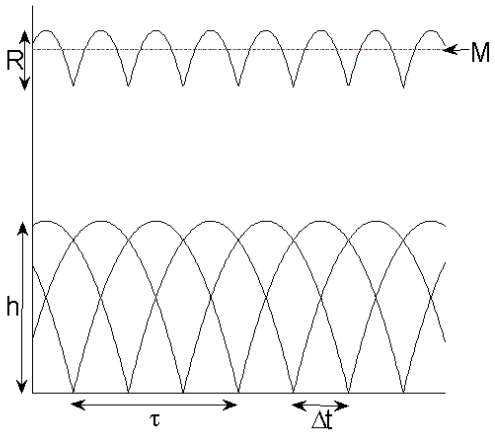


Figure 2.7: Overlapping parabolas of constant height, baseline and distance between parabolas, sum to a mean and range around that mean.

It is apparent that if parabolas overlap, the sum over parabolas yields a small fluctuation about a mean (Figure 2.7).

Figure 2.7 shows that the sum over parabolas is a set of sequential parabolas. Clearly, the mean over the first parabola is the mean over the sequence of parabolas. As seen in Figure 2.7, the minimum of this sum occurs at $t = n\Delta t$ and the maximum occurs at $t = (n + \frac{1}{2})\Delta t$. From our formula, we find that

$$H_{min} = H(n\Delta t) = \frac{1}{6}\Delta t^2 hn(n^2 - 1), n \geq 1 \tag{20}$$

$$H_{max} = H((n + \tfrac{1}{2})\Delta t) = \frac{1}{12}\Delta t^2 hn(2n^2 + 1), n \geq 1 \tag{21}$$

$$\Rightarrow H_{max} - H_{min} = \frac{1}{4}\Delta t^2 hn \tag{22}$$

Using the height of the parabola, we find the mean, M to be

$$M = \frac{2}{3}\frac{1}{4}\Delta t^2 hn + \frac{1}{6}\Delta t^2 hn(n^2-1) \quad (23)$$

$$= \frac{1}{6}\Delta t^2 hn^3 \quad (24)$$

and clearly the range R is given above by

$$R = H_{max} - Hmin = \frac{1}{4}\Delta t^2 hn. \quad (25)$$

Using $\Delta t = \frac{2}{n}$ (time is normalised by $\frac{\tau}{2}$), we find

$$M = \frac{2}{3}hn \quad (26)$$

$$R = \frac{h}{n} \quad (27)$$

These expressions are independent of time t and so they model a stable mean (Figure 2.7). By observing M and R in real data, we may calculate n and h by

$$n = \sqrt{\frac{3M}{2R}} \quad (28)$$

$$h = Rn. \quad (29)$$

Note that n is a number, h and R will be measured in RNA copies/ml, or CD4+ cells/l for viral load and CD4+ count data respectively.

To simulate random sequences of overlapping parabolas, we have varied h_i and $t_i = i\Delta t$ (by varying Δt) by sampling from normal distributions, $N(h, \sigma_h \leq \frac{h}{3})$ and $N(\Delta t, \sigma_{\Delta t} \leq \frac{\Delta t}{3})$ respectively. We find (30) and (31) to be good estimates (with 2.2% error in the case of Δt and no error with respect to h), as seen in Figures 2.8 and 2.9. This model clearly only holds if an overlap of HIV-1 viral populations exists, in other words, if $n \geq 1$ from data (Equation (28)).

It is important to note that these formulae do not depend on the real-time baseline of a parabola, which is not available from MACS data. Similarly, the formulae do not depend on the date of measurement, only the mean and range of the MACS data.

2.5 The Intermittent Model

If viral load parabolas do not overlap, the fluctuation in total viral load is of order H_{max}. Since $H_{max} \sim 10^5$ (1), we can now model VL (viral load) fluctuation. We model viral load as sequentially composed of a parabola of period τ', followed by a low-level viral reservoir of period τ_L (see Figure 2.10); without inconsistencies, we approximate viral load in the reservoir by zero. Since $H_{max} \sim 10^5$ copies/ml, $H_{min} \sim 10^3$ copies/ml, the effect of the low-level VL can be ignored. Assuming that measurements are made uniformly in time on this structure, we calculate the mean height to be

$$M' = \frac{2H_{max}\tau'}{3(\tau' + \tau_L)}, \quad (30)$$

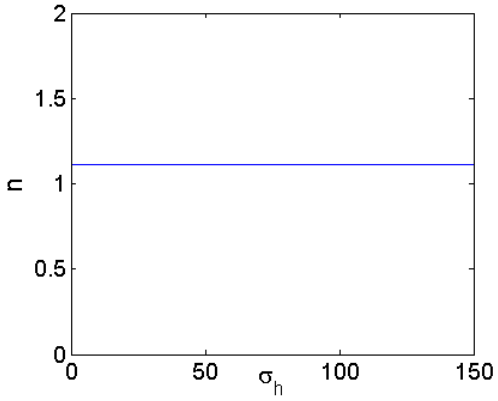


Figure 2.8: The correspondence between the size of the standard deviation of the heights of the parabola and n, the number of overlapping parabolas. The sampling distribution has no effect on n.

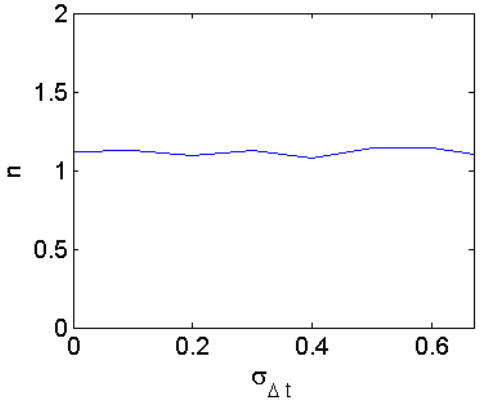


Figure 2.9: The correspondence between the size of the standard deviation of the distance between starting times of the parabolas and n, the number of overlapping parabolas. The sampling distribution has a maximum error of 2.2% on n.

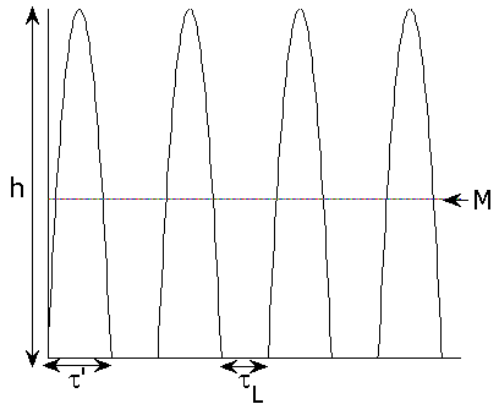


Figure 2.10: Intermittent parabolas allow for a large range $R \sim H_{max}$ while the mean of the sum over parabolas is decreased by the delay.

where τ' is the baseline of the parabola (where τ' not necessarily equal to τ above) and τ_L is the latent period. The delay ratio, $\overline{\tau_L} = \tau_L/\tau'$, if it exists, is determined from M' and H_{max} from data.

$$\overline{\tau_L} = \frac{\tau_L}{\tau'} = \frac{2H_{max}}{3M'} - 1. \tag{31}$$

As in the overlap model, sampling the heights of the parabolas, h, from a normal distribution N(h,$\sigma_h \leq \frac{h}{3}$) does not significantly change n. Since $\overline{\tau_L}$ is to be determined, this may not be sampled. Thus this model is stable under variations in the heights of the parabolas.

Clearly this model only holds if such a gap exists, in other words $\overline{\tau_L} \geq 0$. Else the overlap model must hold.

We have given the above models in terms of viral loads. It is clear that they may be similarly applied to CD4+ count data.

3 The ODE model

The main finding of this dissertation is the result of the algebraic model. However, we would like to show that similar results follow using an ODE model.

First, we would like to look at two standard HIV models. The first, ([29] p.125) models the relationship between free virus particles and immune system response for a single population as follows:

$$\dot{v} = rv - pxv \tag{32}$$
$$\dot{x} = cv - bx \tag{33}$$

Where v represents free virus and x represents the immune system response aimed at that particular virus population. The virus grows at exponential rate r and is eliminated at a rate pxv. The immune system response is stimulated by the presence of free virions at a rate c and the immune system cells are cleared at a rate b.

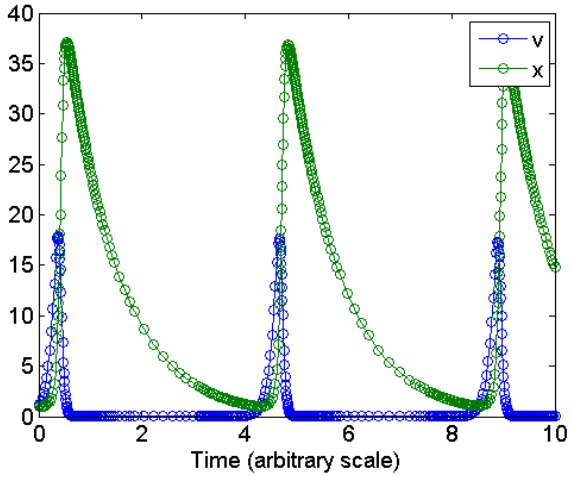


Figure 3.1: The solution to Nowak and May's equations (32)-(33), with arbitrary constants.

This model (Figure 3.1) has a very interesting time scale difference between the two variables which we would like to exploit. It is also interesting to note the cyclical outburst and rest of v. However, in out model, the oscillations are not appropriate. We bear in mind the idea of memory CTLs, where the example of measles is used - once a person has had measles, the vast majority of patients cannot catch it again [38]. Measles does not even appear with diminished effect after a second exposure, making damped oscillations also inappropriate. We would also like to use a model that encompasses the behaviour of CD4+ cells as hosts for the virus.

The next model (by Perelson and Nelson [36]) models the interaction between free virus and infected and uninfected CD4+ cells as follows:

$$\dot{T} = s + pT(1 - T/T_{max}) - d_T T - kVT \tag{34}$$
$$\dot{T^*} = kVT - \delta T^* \tag{35}$$
$$\dot{V} = N\delta T^* - cV \tag{36}$$

Here V represents free virions, T^* infected CD4+ cells and T uninfected CD4+ cells. The free virions are produced by the infected cells at a rate N and have a natural death rate c. The uninfected cells are converted to infected cells at a rate kVT, proportional to both the number of virions and the number of available hosts. The infected cells have a natural death rate δ. The uninfected CD4+ cells are kept at a constant level by the terms $s + pT(1 - T/T_{max}) - d_T T$ where s is a constant source term, the middle term is the logistic model of saturating growth, and d_T is a natural death rate.

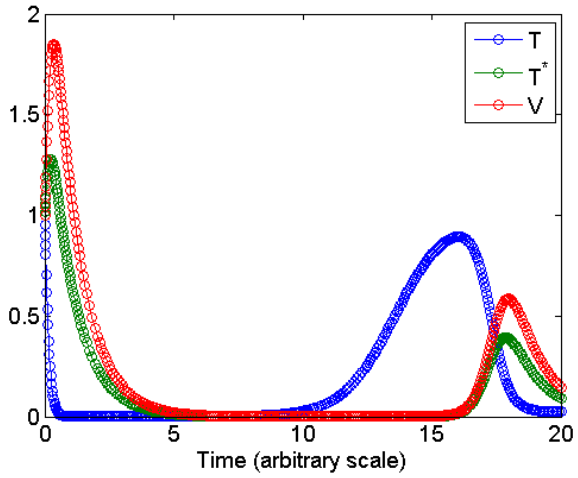


Figure 3.2: The solution to Perelson and Nelson's equations (34)-(36), with arbitrary constants.

This model has very interesting interactions between free virus and CD4+ cells, however it does not take into account the immune system, and we would prefer different time scales between free virus and infected CD4+ cells. The damped oscillations are not appropriate, as discussed above, and the damped oscillations for viral load and infected CD4+ cells reach an equilibrium greater than zero. It is likely that introducing immune system effects to this model will effective kill each virus population. It should be noted that we are only modelling a single virus population in the body and we rely on virus mutation for the next outburst to occur.

3.1 The Equations

We combine the above models as follows (where all values are dimensionless):

$$\dot{V} = rT^* - dXV \tag{37}$$
$$\dot{X} = s_X + cT^* - bX \tag{38}$$
$$\dot{T^*} = kVT - \delta T^* \tag{39}$$
$$\dot{T} = s_T + pT(1 - T/T_{max}) - d_T T - kVT \tag{40}$$

As in above models, V represents free virus, X the CTL effect, T^* the infected CD4+ cells and T the uninfected CD4+ cells. The equations for T and T^* are identical to those in Perelson and Nelson's model. The free virions are produced by infected CD4+ cells at a rate r and are eliminated by the immune system at a rate d. The natural death rate is ignorable over the timescales of weeks to months, which is the lifetime we assume for each viral population. The immune system is stimulated by the presence of infected CD4+ cells rather than by free virions as in Nowak and May's model above. This is due to the fact that the immune system recognises epitopes displayed on the surface of infected cells, as expained in Section 1. The death rate remains the same. The source term in the equation for the immune system is representative of the memory CTLs. These cells have such long lifetimes compared to the time scales of the above equations that they can be seen to be "produced" at a constant rate.

3.2 The Constants

We set the following values to the dimensionless constants

$$V(0) = 10 \tag{41}$$
$$X(0) = 0 \tag{42}$$
$$T^*(0) = 10 \tag{43}$$
$$T(0) = 10 \tag{44}$$
$$r = 10 \tag{45}$$
$$d = 1 \tag{46}$$
$$s_X = 1 \tag{47}$$
$$c = 0.8 \tag{48}$$
$$b = 0.05 \tag{49}$$
$$k = 1 \tag{50}$$
$$\delta = 1.5 \tag{51}$$
$$s_T = 1 \tag{52}$$
$$p = 100 \tag{53}$$
$$T_{max} = 4 \tag{54}$$
$$d_T = 30 \tag{55}$$

and find the solution curve given by Figure 3.3.

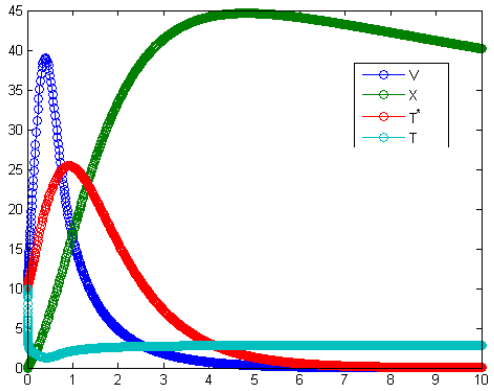


Figure 3.3: The solution to our equations with the initial constant values.

Figure 3.3 is the solution to the ODE (37)-(40) using constants (41)-(55). This figure will be used as a reference when the constant values are changed in order to discuss their impact on the solution. We note that V and T^* fall to zero, that is, there is no recurrence of viremia. The rate of fall of T^* is slower than that of V. The immune system clearly shows memory and uninfected cells recover to an equilibrium. Therefore there are available host cells for the next outburst.

The effects of choosing these constants will be examined in the graphs below.

Increasing the initial amount of virus in the blood (Figure 3.4) does raise the graph of viral load and slightly raises the graphs of the infected CD4+ cells and the immune system response. It is interesting to note that a very low initial level of virus rises very quickly and the overall shapes of the solution curves are unchanged.

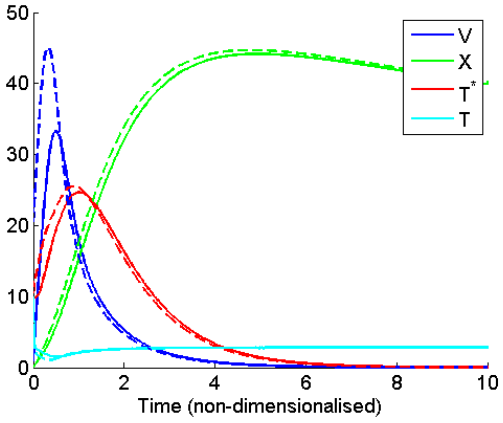


Figure 3.4: The solution to our equations with $V(0) = 1$ (solid line) and $V(0) = 20$ (dashed line).

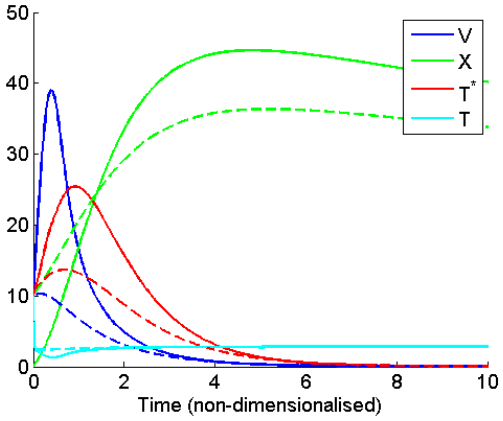


Figure 3.5: The solution to our equations with $X(0) = 0$ (solid line) and $X(0) = 10$ (dashed line).

As seen in Figure 3.5, a high initial immune system response prevents the viral load from growing initially and lowers the amplitude of the infected CD4+ cells. The numbers of uninfected cells remain stable throughout without any significant drop. However the general shape of all the graphs are the same.

A high initial number of infected host cells results in a rise in all graphs in Figure 3.6. Again, the shapes of all the solutions remain the same and viral load and infected CD4+ cells drop very close to zero.

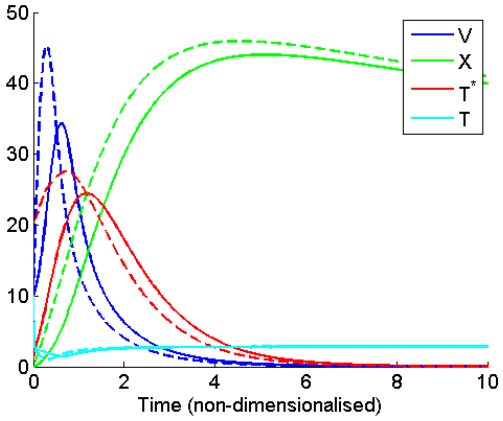


Figure 3.6: The solution to our equations with $T^*(0) = 1$ (solid line) and $T^*(0) = 20$ (dashed line).

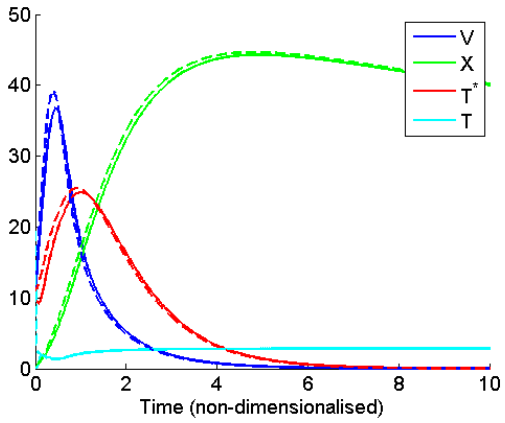


Figure 3.7: The solution to our equations with $T(0) = 0$ (solid line) and $T(0) = 20$ (dashed line).

The number of host cells initially available has little impact on the solution in Figure 3.7. This is due to the fact that the host cells reach equilibrium very quickly, so the initial value is soon replaced with the equilibrium value. More host cells results in slightly more infected cells (and consequentially more free virus and a slightly larger immune system response), but the changes in the graphs are negligible.

A higher rate of free virion production (Figure 3.8) will obviously result in a faster growth of virus. This results in more infected cells, an immune system response that is more highly stimulated, and a more noticeable drop in uninfected host cells. The only change in shape is the slight irregularity in the curve of the infected CD4+ cells. This is mirrored in the sharper drop in uninfected cells and the point of inflection is seen to correspond with the peak in the viral load.

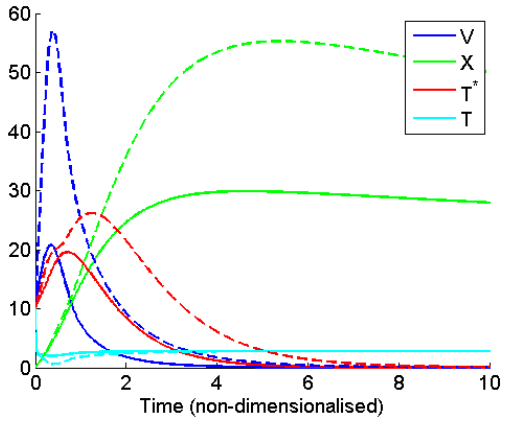


Figure 3.8: The solution to our equations with $r = 5$ (solid line) and $r = 15$ (dashed line).

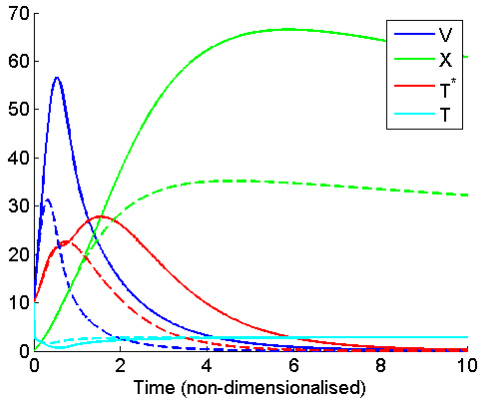


Figure 3.9: The solution to our equations with $d = 0.5$ (solid line) and $d = 1.5$ (dashed line).

Lowering the death rate of the free virions has a similar effect on the graphs in Figure 3.9. The inflection is again observed in the infected CD4+ cells. An interesting result of this model is that, after possible inflection, the infected CD4+ cell population continues to grow, even when the free virions are declining. This is due to the longer lifespan associated with the infected cells, and the fact that the free virions are targeted by the immune system. It is interesting to note the change in lifetimes of viral load and infected CD4+ populations.

Changing the source term associated with the immune system response produces interesting results, as seen in Figure 3.10. It is seen that a lower source term results in oscillations in viral load, infected CD4+ cells and the immune system response. A high source term results in a rapid drop to zero for viral load and infected cells. The uninfected cells are not visibly affected by the source term. Thus by choice of constant, we can achieve oscillations as in models discussed previously, but this is not appropriate to our model definition. Thus we favour larger values for s_X.

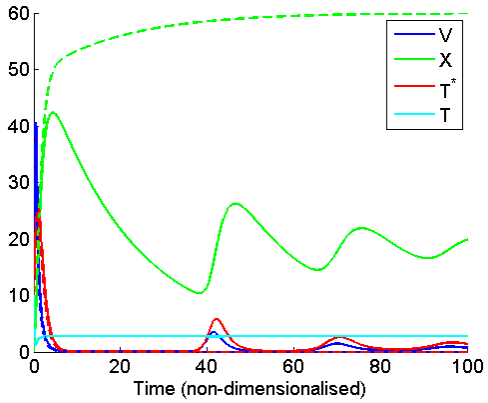


Figure 3.10: The solution to our equations with $s_X = 0.1$ (solid line) and $s_X = 3$ (dashed line).

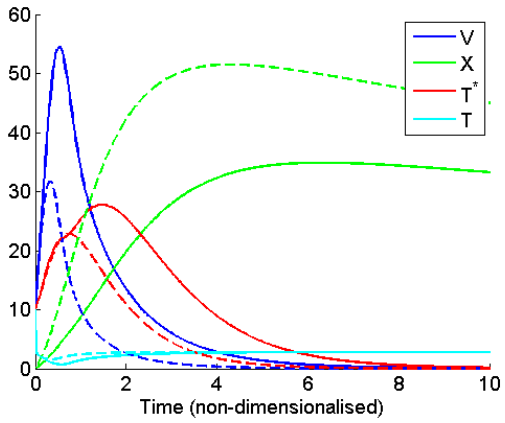


Figure 3.11: The solution to our equations with $c = 0.4$ (solid line) and $c = 1.2$ (dashed line).

As seen in Figure 3.11, increasing the rate at which the immune system is stimulated by infected cells results in lower viral load and infected CD4+ count both with shorter lifetime and a higher immune system value. The uninfected CD4+ cells also recover quicker. The shapes of the graphs are mostly unchanged.

The clearance rate of immune system cells, however, has a large impact on the graph shapes in Figure 3.12. We see oscillations for high values of b and in that case, the viral load and infected CD4+ count reach a non-zero equilibrium. Thus we favour weaker clearance rate b.

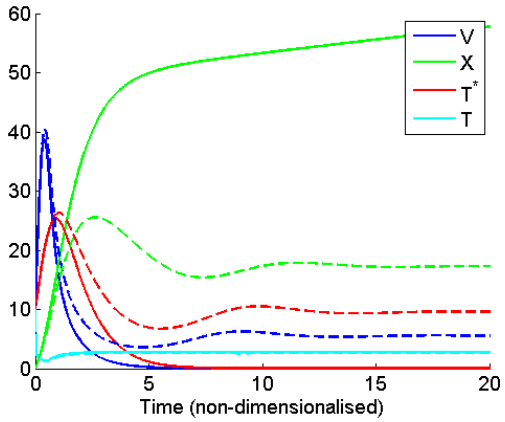


Figure 3.12: The solution to our equations with $b = 0.01$ (solid line) and $b = 0.5$ (dashed line).

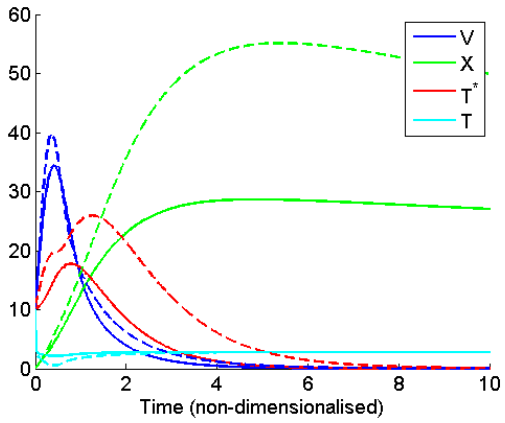


Figure 3.13: The solution to our equations with $k = 0.5$ (solid line) and $k = 1.5$ (dashed line).

The rate of conversion between uninfected and infected CD4+ cells has only a small impact on viral load and uninfected CD4+ cells in Figure 3.13. However, for high k, infected CD4+ cells are raised and have longer lifetime, and in response, the immune system is stimulated. It is the forced constant rate of the immune system that allows the longer time span for the infected CD4+ cells as they effectively have an unlimited supply of host cells before they are removed by the lack of virions to infect new cells.

The death rate of infected cells (Figure 3.14) also has little impact on viral load and uninfected cells. A higher death rate predictably lowers the infected CD4+ cell count and the lifetime of infected CD4+ and in response, the immune system is less stimulated. The shapes of all the graphs remain the same.

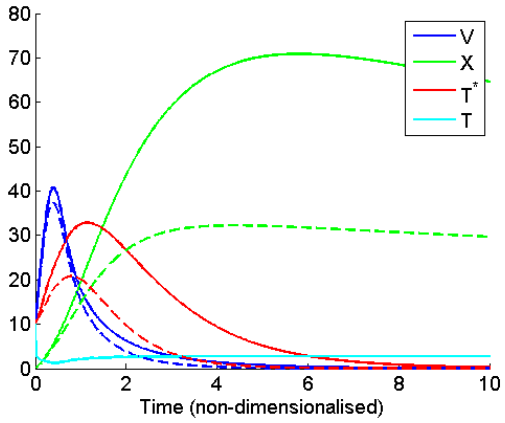


Figure 3.14: The solution to our equations with $\delta = 1$ (solid line) and $\delta = 2$ (dashed line).

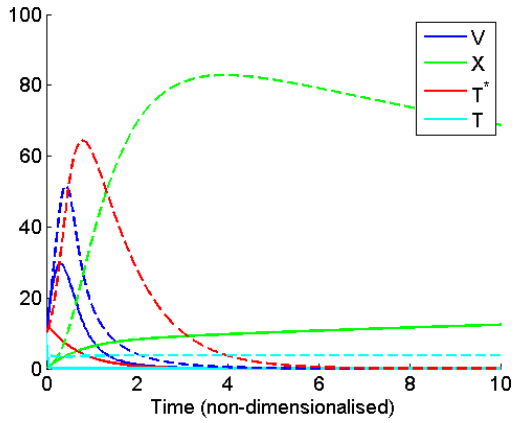


Figure 3.15: The solution to our equations with $p = 10$ (solid line) and $p = 500$ (dashed line).

The term p is responsible for the uninfected CD4+ cells reaching an equilibrium. It is clearly seen in Figure 3.15 that a low p results in almost no uninfected CD4+ cells being available. This results in the infected cells being unable to rise due to a lack of hosts, and as a result the immune system is not stimulated very much. The viral load solution keeps its shape but has a lower amplitude and shorter lifetime.

The value T_{max} determines the value of the uninfected CD4+ cell equilibrium. A low value for T_{max} has the same result on the solutions as a low value for p discussed above, as seen in Figure 3.16.

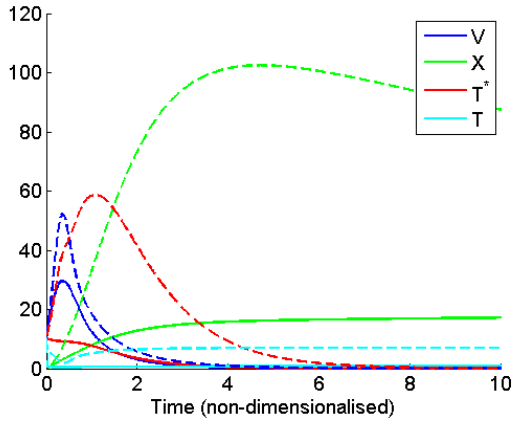


Figure 3.16: The solution to our equations with $T_{max} = 1$ (solid line) and $T_{max} = 10$ (dashed line).

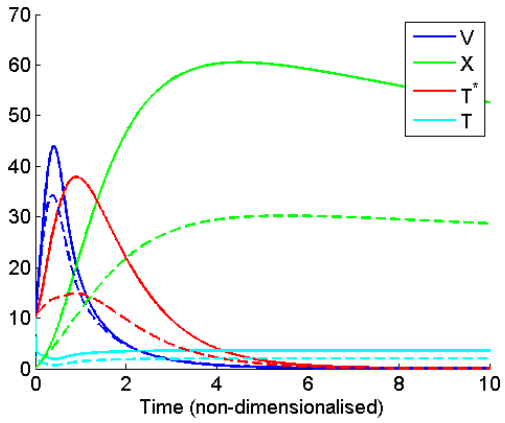


Figure 3.17: The solution to our equations with $d_T = 10$ (solid line) and $d_T = 50$ (dashed line).

The clearance rate for uninfected CD4+ cells (Figure 3.17) has an effect on the equilibrium value for uninfected cells and a high clearance rate results in much lower infected CD4+ cells (due to lack of hosts) and a lower value for the immune system reaction. The effect on viral load is minimal.

The source term for uninfected cells does not have a large impact on the solution curves in Figure 3.18. A larger source means more hosts for infection, so infected CD4+ cells are slightly increased and as a result the immune system effect is slightly increased. The effect on uninfected CD4+ cells and viral load is minimal.

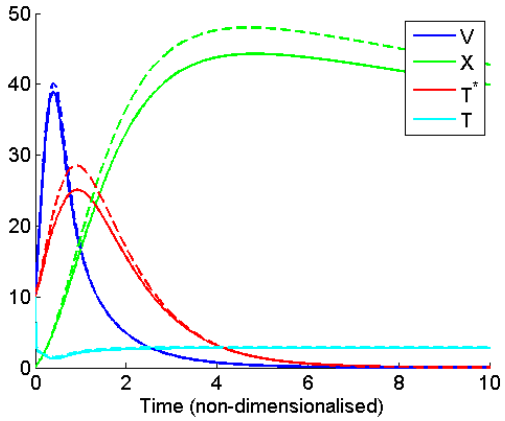


Figure 3.18: The solution to our equations with $s_T = 0.1$ (solid line) and $s_T = 10$ (dashed line).

3.3 Fixed Point

To find the fixed points, we set

$$\begin{aligned} 0 &= rT^* - dXV && (56)\\ 0 &= s_X + cT^* - bX && (57)\\ 0 &= kVT - \delta T^* && (58)\\ 0 &= s_T + pT(1 - T/T_{max}) - d_T T - kVT && (59) \end{aligned}$$

and find the fixed point at

$$\begin{aligned} V &= 0 && (60)\\ X &= \frac{s_X}{b} && (61)\\ T^* &= 0 && (62)\\ T &= \frac{-(d_T + p)T_{max} \pm \sqrt{4ps_T T_{max} + (d_T T_{max} - pT_{max})^2}}{2p} && (63) \end{aligned}$$

or

$$\text{let } \gamma = \sqrt{T_{max}(4cd^2p(\delta s_X + cs_T) + (cd(d_T - p) + bkr)^2 T_{max})}$$

$$\begin{aligned} V &= \frac{(cd\delta s_X T_{max})(p - d_T) + bkrT_{max}(\delta s_X + 2cs_T) \pm \delta s_X \gamma}{2cdk(\delta s_X + cs_T)T_{max}} && (64)\\ X &= -\frac{kr(cdd_T T_{max} - cdpT_m ax + bkrT_m ax \pm \gamma)}{2cd^2\delta p} && (65)\\ T^* &= -\frac{2cd^2\delta p s_X + bcdd_T krT_{max} - bcdkp_T rT_{max} + b^2k^2r^2T_{max} \pm bkr\gamma}{2c^2d^2\delta p} && (66)\\ T &= -\frac{cdd_T T_{max} - cdpT_{max} + bkrT_{max} \pm \gamma}{2cdp} && (67) \end{aligned}$$

When substituting in the defined constant values, we find only one solution with all variables positive. This is given (in non-dimensional variables) by

$$\begin{aligned} V &= 0.0 && (68)\\ X &= 20.0 && (69)\\ T^* &= 0.0 && (70)\\ T &= 2.8 && (71) \end{aligned}$$

It is seen that every variable reaches an equilibrium. It is important to note that viral load only escapes when

$$2cdk(\delta s_X + cs_T)T_{max} = 0 \quad (72)$$

In other words, viral load only escapes when the immune system reaches a zero equilibrium, or free virions do not die, for example.

3.4 Constants with Dimension

The above graphs are shown with dimensionless constants in order to make the comparisons easier. It must be shown that these constants can be returned to their correct dimensions.

We rescale so that $V = V'/\hat{V}$, $X = X'/\hat{X}$, $T^* = T^{*'}/\hat{T^*}$ and $T = T'/\hat{T}$, where the hatted variables contain the dimensions. Then, equations (37)-(40) become:

$$\hat{V}^{-1}\dot{V'} = r\frac{T^{*'}}{\hat{T^*}} - d\frac{X'}{\hat{X}}\frac{V'}{\hat{V}} \tag{73}$$

$$\hat{X}^{-1}\dot{X'} = s_X + c\frac{T^{*'}}{\hat{T^*}} - b\frac{X'}{\hat{X}} \tag{74}$$

$$\hat{T^*}^{-1}\dot{T^{*'}} = k\frac{V'}{\hat{V}}\frac{T'}{\hat{T}} - \delta\frac{T^{*'}}{\hat{T^*}} \tag{75}$$

$$\hat{T}^{-1}\dot{T'} = s_T + p\frac{T'}{\hat{T}}(1 - \frac{T'}{\hat{T}}/T_{max}) - d_T\frac{T'}{\hat{T}} - k\frac{V'}{\hat{V}}\frac{T'}{\hat{T}} \tag{76}$$

Our constants are re-defined (now with dimension) as:

$$\hat{r} = \frac{r\hat{V}}{\hat{T^*}} \tag{77}$$

$$\hat{d} = \frac{d}{\hat{X}} \tag{78}$$

$$\hat{s_X} = s_X\hat{X} \tag{79}$$

$$\hat{c} = \frac{c\hat{X}}{\hat{T^*}} \tag{80}$$

$$\hat{b} = b \tag{81}$$

$$\hat{k} = \frac{k\hat{T^*}}{\hat{V}\hat{T}} \tag{82}$$

$$\hat{\delta} = \delta \tag{83}$$

$$\hat{s_T} = s_T\hat{T} \tag{84}$$

$$\hat{p} = p \tag{85}$$

$$\hat{T_{max}} = T_{max}\hat{T} \tag{86}$$

$$\hat{d_T} = d_T \tag{87}$$

$$\hat{k} = \frac{k}{\hat{V}} \tag{88}$$

Note above that there are two possible dimensions for k ((82) and (88)). In order for this to be resolved, $\hat{T^*} = \hat{T}$. These are both CD4+ counts with the same dimension, so this is possible.

From Figure 3.3 above, we seen that definitive values for the variables are as follows:

$$V_{max} \approx 40 \tag{89}$$
$$X_{max} \approx 45 \tag{90}$$
$$T^*_{max} \approx 25 \tag{91}$$
$$T_{max} \approx 4 \tag{92}$$

We only have accurate data for total viral load and for total CD4+ count. Since the dimensions for $\hat{T^*}$ and $\hat{T}$ must be the same, we will simply split the total to get these values. During HIV-1 infection, the ratio between CD4+ and CD8+ (CTL) cells rapidly reaches 1:2 [1], so we will assume the value for the immune system response to be double that of the total CD4+ count. As such, we have the following definitive values for variables with dimension (using MACS data for patients not on treatment):

$$V'_{max} \approx 10^5 \tag{93}$$
$$X'_{max} \approx 10^3 \tag{94}$$
$$T^{*'}_{max} \approx 250 \tag{95}$$
$$T'_{max} \approx 250 \tag{96}$$

From this, we determine the constants holding dimension to be:

$$\hat{V} \approx 250 virions/ml \tag{97}$$
$$\hat{X} \approx 20\ cells/l \tag{98}$$
$$\hat{T^*} \approx 10\ cells/l \tag{99}$$
$$\hat{T} \approx 60\ cells/l \tag{100}$$

Therefore the new constants (with dimension) become:

$$\hat{r} = 250\ thousand\ virions\ per\ infected\ CD4+\ cell \tag{101}$$
$$\hat{d} = 0.05\ CD8+\ cells^{-1} \tag{102}$$
$$\hat{s_X} = 20\ CD8+\ cells \tag{103}$$
$$\hat{c} = 1.6\ CD8+\ cells\ per\ infected\ CD4+\ cell \tag{104}$$
$$\hat{b} = 0.05\ (non-dimensional\ clearance\ rate) \tag{105}$$
$$\hat{k} = 6 \times 10^{-4}\ virions^{-1} \tag{106}$$
$$\hat{\delta} = 1.5\ (non-dimensional\ clearance\ rate) \tag{107}$$
$$\hat{s_T} = 60\ uninfected\ CD4+\ cells \tag{108}$$
$$\hat{p} = 100\ (non-dimensional\ rate) \tag{109}$$
$$\hat{T_{max}} = 240\ uninfected\ CD4+\ cells \tag{110}$$
$$\hat{d_T} = 30\ (non-dimensional\ clearance\ rate) \tag{111}$$

Where, from the minimum values in MACS data (for untreated patients) and from previous assumptions:

$$V(0) \approx 10^3 \ virions/ml \tag{112}$$
$$X(0) \approx 0 \ cells/l \tag{113}$$
$$T^*(0) \approx 200 \ cells/l \tag{114}$$
$$T(0) \approx 200 \ cells/l \tag{115}$$

Note that time was not added into the dimensionalisation because we have no data with which to accurately set the time axis. The time scale of free virions in Figure 3.19 is seen to be of order months, which is in agreement with our arguments in Section 2.1.

All values with dimensions are estimates only. The above constants lead to the graphs shown in Figure 3.19.

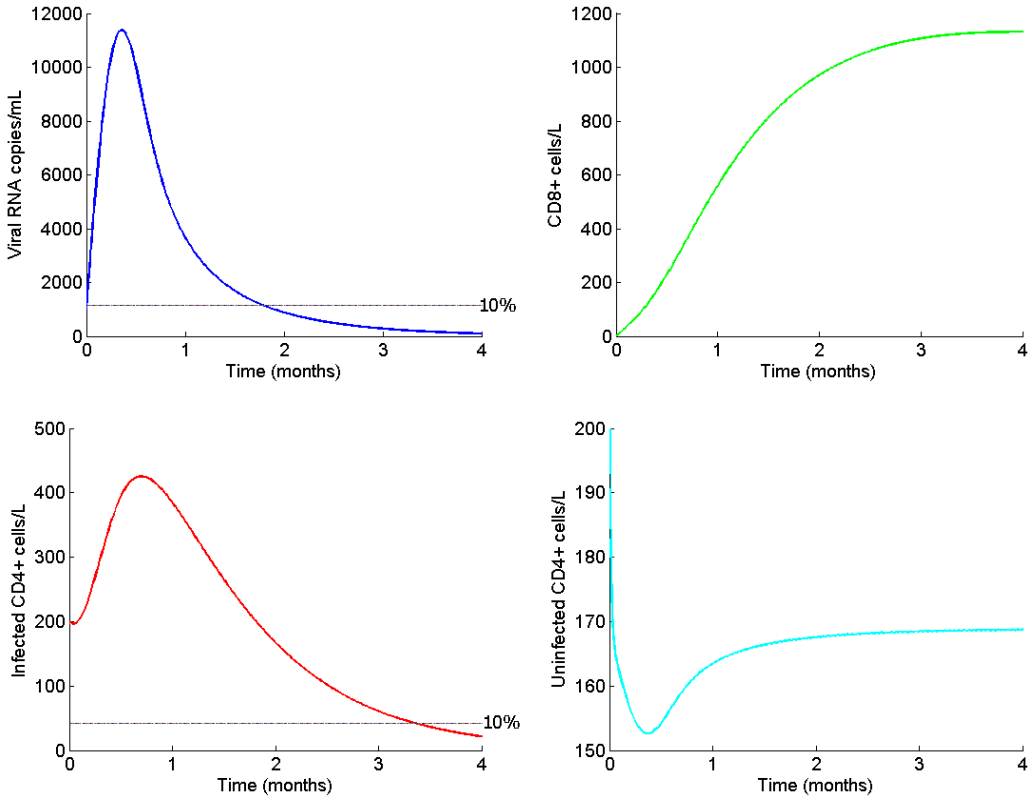


Figure 3.19: The solution curves to equations (37)-(40) with dimensions (using constants (77)-(88) and (93)-(96)).

When using these constants with dimension, we find the fixed point to occur at

$$V = 0 \; virions/ml \; blood \tag{116}$$
$$X = 400 \; CD8+ \; cells/l \; blood \tag{117}$$
$$T^* = 0 \; CD4+ \; cells/l \; blood \tag{118}$$
$$T = 168 \; CD4+ \; cells/l \; blood \tag{119}$$

(The curve for CD8+ cells decreases slowly until equilibrium is reached.) This can be artificially adjusted by choice of (96) and (100), but we regard the return to steady uninfected cell levels within about two months as merely sufficient to support a new outburst of viremia. It is interesting to note that without the infected CD4+ cells, the (imaginary) patient represented by Figure 3.19 falls into the category of AIDS.

3.5 Mean Viral Load and CD4+

The mean height of viral load in the ODE solution (with both $V(0) = 1$ and $T^*(0) = 1$) is $0.41h$, where h is the maximum viral load. Thus our parabola (mean=$0.67h$) overestimates the biological mean. In the case of infected CD4+ cells, the mean in Figure 3.3 is $0.49h$, where h is the maximum number of CD4+ cells. Therefore our parabola also overestimates the biological mean in this case. In both viral load and infected CD4+ counts, the profile was terminated at a minimum height 10% of the maximum, because the tails fall into and are hidden by the low-level population H_{min}. This result is discussed below.

4 Discussion

Recall that we use observed M and R (for overlap model) and M' and H_{max} (for intermittent model) to solve for n and h, and $\overline{\tau_L}$ respectively. Recall also that the overlap model must fail if $n < 1$ and the intermittent model must fail if $\overline{\tau_L} < 0$.

4.1 Deciding the Model for Viral Load and CD4+ Data

4.1.1 Viral Load Data

Using the MACS dataset [27], described above, we find the results displayed in Table 4.1 for viral load measurements.

	P	#	M	max	min	R	σ	n^{VL}	$\overline{\tau_L}^{VL}$
NoARV	138	6.64	5.84	18.0	0.69	17.3	6.63	0.75	1.04
NoARV A	127	6.57	6.11	16.2	0.46	17.7	6.76	0.77	1.00
NoARV B	114	6.60	6.55	10.0	0.18	18.3	7.12	0.78	0.95
NoARV AIDS	3	5.33	15.9	32.8	5.72	27.1	10.8	0.98	0.30
Pre92	380	9.39	4.89	18.6	0.60	18.0	5.76	0.77	1.15
Pre92 A	359	8.91	5.41	18.6	0.60	18.6	6.11	0.81	1.03
Pre92 B	358	8.92	5.43	18.6	0.60	18.7	6.12	0.81	1.03
All	754	18.3	5.04	29.6	0.22	29.4	7.99	0.54	3.30
All A	751	17.6	5.22	29.6	0.22	29.4	8.07	0.55	3.17
All B	698	14.3	6.16	29.6	0.22	29.1	8.44	0.62	2.19
All AIDS	50	6.88	10.6	34.6	0.92	33.7	12.6	0.72	1.30

Table 4.1: The statistics of viral load for MACS data, where P is the number of patients per group, # is the mean number of data points per patient in this group, M is the mean viral load, R is the range of viral load, and σ is the standard deviation of the viral load. All viral load data is measured in 10^4 copies/ml. A patient was only taken to belong to a particular category if they had more than 5 points in that category. Group A removed entries where viral load was 300 and group B removed entries where viral loads were 40 or 300. The intermittent model is clearly favoured for viral load data.

The minimum viral load in the group of all patients is lower than the other groups. This is probably because patients whose viral loads have decreased due to treatment are included in this group. However, the maximum viral load is also noticeably higher for this group. It is seen that there are more AIDS patients in this group than in the no treatment group, and this could account for the higher maximum. It is due to this that the range is much higher for this group, and this results in the lower value for n^{VL}. The higher maximum is the cause of the higher value for $\overline{\tau_L}^{VL}$.

The removal of special viral load values in groups A and B (see Section 1.4.3) resulted in an increased mean viral load in all groups. This is to be expected, as we are removing all successfully treated candidates from consideration. This results in the lower values for $\overline{\tau_L}^{VL}$ and slightly higher values for n^{VL}. The removal of these special values had no impact on maximum and minimum viral loads, except in the group of no treatment. In this case both maximum and minimum values dropped, possibly due to the conditions for consideration: a patient was only considered to be part of a group if they had 5 or more measurements in that group. It is reasonable to suppose that those with high viral loads might be put on treatment and the viral load would drop to a special value if treatment was successful. There may not have been enough remaining measurements once the special values had been removed for those with high initial viral loads to be considered.

The groups of AIDS patients (patients with CD4+ counts persistently lower than 200 cells/l) had large increases in all viral load markers. This is to be expected, and the increased n^{VL} and decreased $\overline{\tau_L}^{VL}$ as a result shows that viral outburst occurs more frequently during the end phase. This is a very important result, discussed below.

We expected good agreement between the groups of patients not on treatment and measurements taken before 1992, but the pre92 group has slightly lower means and smaller standard deviations. This could be because many of the patients were healthy before 1992 but remained in the trial in the no treatment group as they got sicker and their viral loads increased. The other markers were in fair agreement between the two groups (without special values removed).

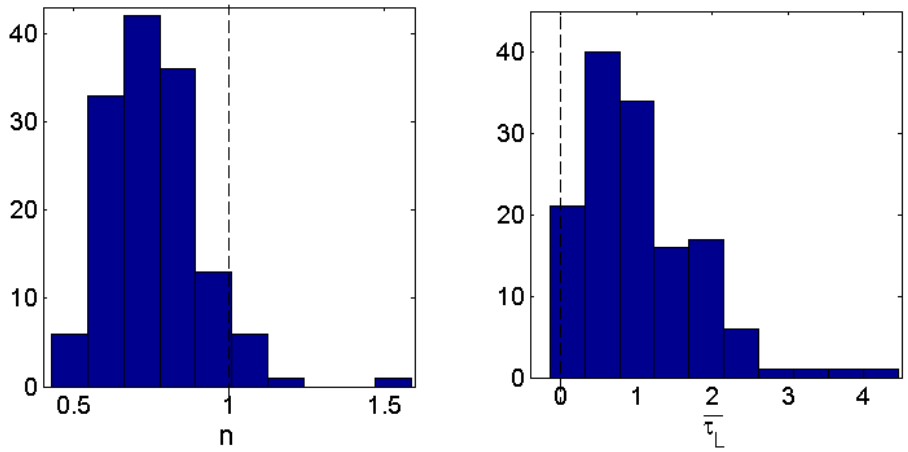


Figure 4.1: Histograms of the n^{VL} and $\overline{\tau_L}^{VL}$ values for viral load data of MACS patients not on treatment. $n^{VL} < 1$ for 93% of patients and $\overline{\tau_L}^{VL} > 0$ for 98% of patients.

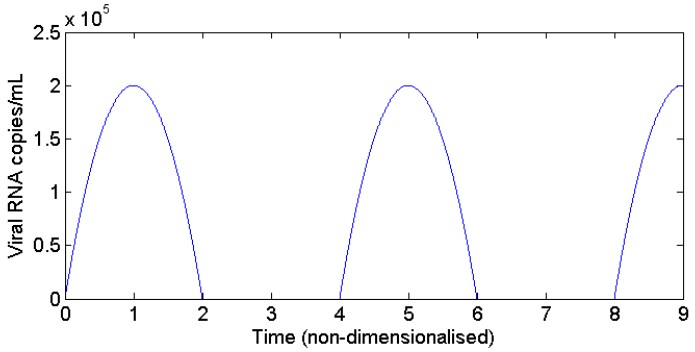


Figure 4.2: For untreated HIV-1 MACS data, observed mean and range gives $\overline{\tau_L}^{VL} = 1$.

As seen in Figure 4.1, $\overline{\tau_L}^{VL} > 0$ for viral load in 98% of cases. The intermittent model holds when $\overline{\tau_L} > 0$ and we find that it applies to viral load data such that $\bar{\tau_L}^{VL} \cong \tau'^{VL}$

(Table 4.1). $n^{VL} < 1$ for viral load in 93% of cases. The overlap model does not, in significant numbers, apply to VL data.

We refer to the measurement errors discussed in Section 1.4.2. In the case of the 10^2 copies/ml error in PCR amplification, only our formulae for n and $\overline{\tau_L}$ would be affected by these errors, and we find that compared to our values for mean, range and maximum viral load ($\sim 10^4$,$\sim 10^5$ and $\sim 10^5$ copies/ml respectively), the error of 10^2 is ignorable.

In the case of 11.4% variation in terms of the mean (Section 1.4.2), we find that if we consider this error for every data point and recalculate $\overline{\tau_L}^{VL}$ and n^{VL} for each patient, we have 91% of patients without treatment with $\overline{\tau_L}^{VL} > 0$ (instead of 98% without error) and 83% of patients in the same group with $n^{VL} < 1$ (instead of 93% without error). So it is still very clear that viral outburst occurs.

There is no patient in the considered MACS dataset not on treatment where both models fail, however there is a contradiction in 7 out of 138 (5%) of patients for viral load data where neither model seems to fail. As seen in Table 4.2, these contradictions happen when n or $\overline{\tau_L}$ are close to the limit, since in 6 of these patients $1 < n \leq 1.09$ and in the remaining patient (patient 5 in Table 4.2) $\overline{\tau_L} = 0.1$.

Contradicting patients:	1	2	3	4	5	6	7
n^{VL}	1.058	1.090	1.001	1.031	1.127	1.006	1.076
$\overline{\tau_L}^{VL}$	0.269	0.143	0.129	0.235	0.117	0.234	0.054

Table 4.2: There are 7 patients in the MACS viral load data of patients not on treatment where both the overlap and intermittent model apply. For these patients, either $\overline{\tau_L}^{VL}$ is close to 0 or n^{VL} is close to 1.

The solution to $\overline{\tau_L}^{VL}$ using MACS viral load data is depicted in Figure 4.2. This is a simplified version of what we envisage the dynamics of viral populations to look like. These dynamics result in large variation in viral load as observed in the MACS data.

4.1.2 CD4+ Data

Note that, for the parabola models, we refer to *infected* CD4+ populations. In order to use MACS data, we assume that by using total CD4+ data for our model we are not significantly altering the results. This is reasonable as follows. If we consider the uninfected CD4+ cells to be a long-lived CD4+ population co-existing with shorter-lived infected populations, the long-lived population may be taken to have a constant value over the lifetime of the shorter-lived population. The numbers of uninfected and abortively infected CD4+ cells are known to exceed the numbers of infected CD4+ cells by two orders of magnitude [1]. In this case, the mean sum, M^s, of the short-lived populations will decrease, but the range, R^s, will remain the same. The maximum, H^s_{max}, of the sum over the short-lived populations will decrease by the same amount as the mean. In the case of the overlap model, this will result in a smaller value for n. Since M and H_{max} decrease by the same amount, $\overline{\tau_L}$ should remain similar for the intermittent model. Thus, even for a population of uninfected CD4+ cells included in MACS data, $\overline{\tau_L}$ does not change, and the overall conclusions are not affected.

Using MACS data [27], we find the results displayed in Table 4.3 for CD4+ measurements.

The mean and maximum CD4+ counts in the group with no treatment is lower than the other groups. This must be due to healthier patients in the other two groups - the pre92 group because they are possibly more recently infected, and the group of all patients because

	P	#	M	max	min	R	σ	n^{CD}	$\overline{\tau_L}^{CD}$
NoARV	138	6.64	339.49	500.26	214.70	285.56	105.52	1.40	0.05
NoARV A	127	6.57	336.64	729.67	261.41	282.17	105.54	1.40	0.04
NoARV B	114	6.60	334.49	822.51	296.67	279.31	104.79	1.41	0.03
NoARV AIDS	3	5.33	112.34	167.33	67.33	100.00	39.50	1.38	0.03
Pre92	380	9.39	584.90	923.93	344.58	579.35	188.10	1.31	0.09
Pre92 A	359	8.91	554.34	923.93	344.58	517.26	172.80	1.35	0.07
Pre92 B	358	8.92	553.64	923.93	344.58	513.86	171.39	1.35	0.07
All	754	18.27	510.23	926.45	222.26	704.19	195.33	1.08	0.29
All A	751	17.56	501.30	926.45	222.26	659.27	186.72	1.10	0.25
All B	698	14.57	473.48	926.45	222.26	620.49	186.78	1.11	0.26
All AIDS	50	6.88	103.66	172.32	46.06	126.26	47.79	1.20	0.18

Table 4.3: The statistics of CD4+ count for MACS data, where P is the number of patients per group, # is the mean number of data points per patient in this group, M is the mean CD4+ count, R is the range of CD4+ count, and σ is the standard deviation of the CD4+ count. All CD4+ data is measured in 10^6 copies/l. A patient was only taken to belong to a particular category if they had more than 5 points in that category. Group A removed entries where viral load was 300 and group B removed entries where viral loads were 40 or 300. The overlap model is clearly favoured for CD4+ data.

some of the patients are probably on treatment. This results in lower n^{CD} and higher $\overline{\tau_L}^{CD}$. In this case, the overlap model may be replaced by the intermittent model in some patients.

The removal of special viral load values had no effect on the maximum and minimum CD4+ counts, except in the group with no treatment. In that group, the maximum and minimum CD4+ counts increased, while the mean decreased slightly. This parallels an increase in mean viral load in the same group. The change in maximum and minimum must be due to the conditions for consideration, as discussed in the case of viral load. In the other groups, the removal of special values resulted in a slightly decreased mean, in the case of all patients due to the removal of healthier patients on treatment. In the pre92 group, it is possible that treatment exists, but patients with effective immune systems could possibly bring the virus to low levels without treatment and we would again be excluding healthier patients.

The groups of AIDS patients had substantially lower CD4+ markers, as is to be expected. In the no treatment group, where only 3 patients are classified as AIDS patients, we do not see a substantial change in n^{CD} and $\overline{\tau_L}^{CD}$. But in the group of all patients which includes more AIDS patients, there is an increase in n^{CD} and a decrease in $\overline{\tau_L}^{CD}$, which parallels in the more frequent outburst in viral load populations during end phase as discussed above.

Unlike the viral load data, there is no agreement in the biological markers between the groups of no treatment and measurements before 1992. This is most likely due to the difference in time since infection. Unfortunately it is extremely difficult to find dates of infection for most HIV-1 infected patients and so this is not possible to test. However, the values for n^{CD} and $\overline{\tau_L}^{CD}$ are comparable, which is interesting since it shows that the same or similar behaviour holds throughout the aymptotic stage, regardless of time since infection, and these dynamics only change in the end phase.

As seen in Figure 4.3, $n^{CD} > 1$ for CD4+ data in 85% of cases. The overlap model holds when $n > 1$ and we conclude that it applies to CD4+ data with $n^{CD} = 1.4$ (Table 4.3). $\overline{\tau_L}^{CD} < 0$ for CD4+ data in 61% of cases. The intermittent model does not, in significant numbers, apply to CD4+ data.

Using the error in CD4+ measurements of 11.9% discussed in Section 1.4.2, we find that

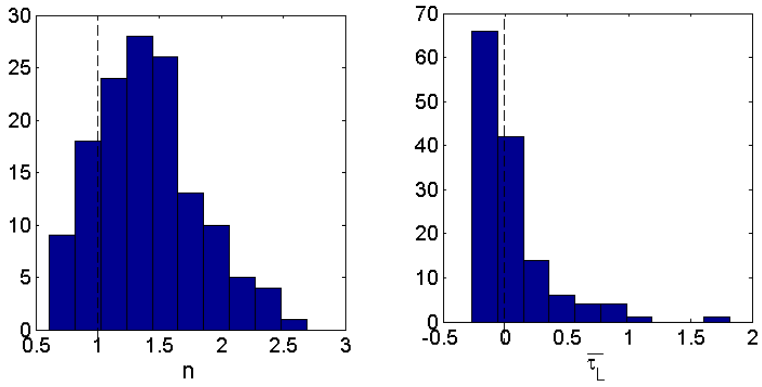


Figure 4.3: Histograms of the n^{CD} and $\overline{\tau_L}^{CD}$ values for CD4+ count data of MACS patients not on treatment. $n^{CD} > 1$ in 85% of patients and $\overline{\tau_L}^{CD} < 0$ in 61% of patients.

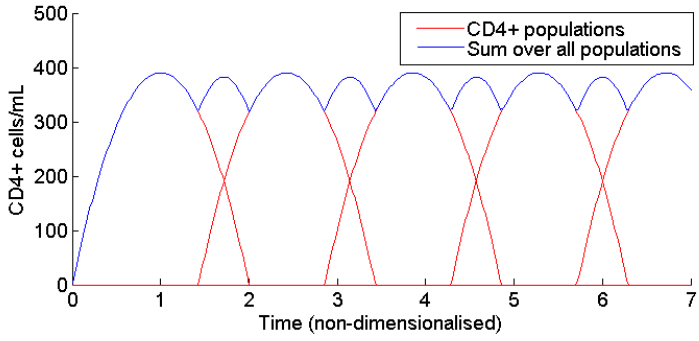


Figure 4.4: For untreated CD4+ MACS data, observed mean and range show $n^{CD} = 1.4$ overlapping parabolas.

for patients not on treatment, 67% had $n^{CD} > 1$, but 91% of untreated patients also had $\overline{\tau_L}^{CD} > 0$. Thus there is contradiction between the models in 57% of untreated patients. The remaining 43% of patients favour one or other model, showing that we are on the borderline between the two models. The reason for the contradiction is that the $\overline{\tau_L}^{CD}$ values were all grouped closely around zero initially, so any error in the data creates problems within the model for CD4+ count. We thus suppose that the contradiction is owing to errors in data measurement.

The solution to n^{CD} using MACS CD4+ count data is depicted in Figure 4.4. This is what we imagine the dynamics of viral populations to look like. These dynamics give M and R as observed.

A principle conclusion of this dissertation is that the untreated disease progresses intermittently. It follows immediately that HIV-1 mutations are only successful on time scales of a month or so. In the case of infected CD4+ cells, there is an overlap in populations, but the phasing with HIV-1 is of great importance.

4.2 Time Scales and Phasing of HIV-1 and Infected CD4+ Cells

The actual durations τ, τ' cannot be determined without knowledge of the date of each patient's hospital visit. Such an investigation must await a better dataset. However, the $\sim$ 1month time scale is certainly possible, as discussed in Section 2.1.

Since different models apply to viral load and CD4+ count data, the question of phasing must be addressed. It is a reasonable assumption that HIV-1 and CD4+ must be in phase, in the sense $\tau \approx \tau' + \tau_L$, if we suppose that they are related in any way at all. Because, if they are not in phase, then all possible relationships occur between the two datasets, contradicting any relationship between them. Biologically, there is certainly a link, and the mathematical models reviewed in Section 1.3 all interconnect the two datasets. An example for the relationship in our ODE is given in Figure 4.5. This scenario can only hold for $\tau \neq \tau'$, in fact, owing to total viral timescale $\tau' + \tau_L$ and $\tau' \approx \tau_L$ we find $\tau \approx 2 \times \tau'$. During the asymptomatic stage, we then suggest that $\tau = 2 \times \tau'$, but observe that there are various possible phasings within this scenario, some of which are displayed in Figure 4.6.

The phasing between viral load and CD4+ count solution graphs for the ODE model (Figure 4.5) corresponds to the left hand graph of possible phasings in Figure 4.6. We note approximate agreement with our conclusion that $\tau \approx \tau' + \tau_L$. Biologically, this means that viral load and infected CD4+ cells grow at roughly the same rate initially, as virus infects the cells. But, because the infected cells live longer than the free virions, they decay at a slower rate. This rate is calculated by substituting the solutions for V and T^* at the time points corresponding with the respective half-heights of each graph (Figure 4.5), into the differential equations (37) and (39). We find the slopes $\frac{dV}{dt} \approx -17$ and $\frac{dT^*}{dt} \approx -8$, which further supports the conclusion $\tau \approx \tau' + \tau_L$ from the intermittent model, using MACS data.

Alternate phasings, shown in Figure 4.6 could represent a delay in CD4+ cells releasing free virus. We assume the time scale of initial viral growth to plateau in Figure 4.5 to be approximately a month, to match time scales of the window period and viral blips (mentioned previously). Since an infected CD4+ cell has a half life of $\sim$ 2 days [19], the delay is not significant for our assumed time scales. For one of the other phasings in Figure 4.6 to hold, the life time of each viral population would have to be on the order of days. We do not favour these latter phasings.

Statistical correlation of CD4+ and HIV-1 fluctuations on time scales of a month or

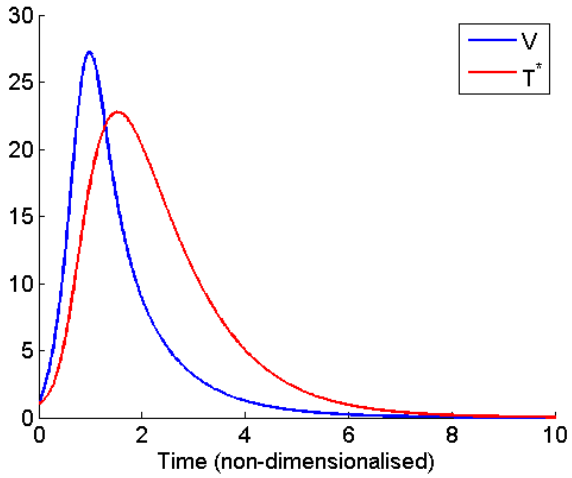


Figure 4.5: The phasing between viral load and infected CD4+ cell solutions to ODE (37)-(40). (Using non-dimensionalised variables.)

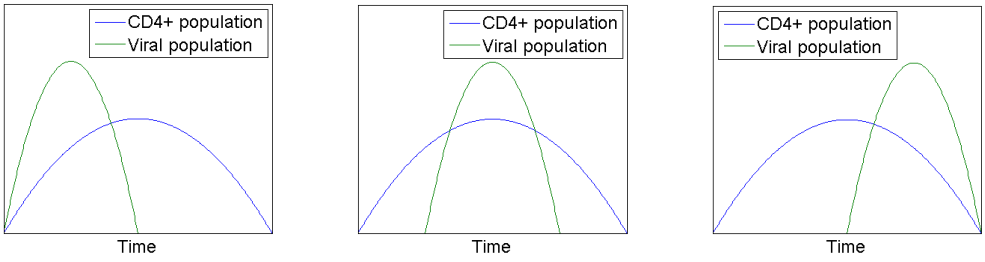


Figure 4.6: Some possible phasings between viral load and CD4+ data. Time is non-dimensional and the heights of each parabola are arbitrary.

so, has not, to our knowledge, been successfully reported in the literature. We note the higher frequency of oscillation of CD4+ of Figure 4.4 as compared with Figure 4.2 for viral load. The more parabolas that overlap, the higher the frequency of oscillation of the CD4+ component. It is clear that correlations between the two components will be hard to detect.

4.3 Our Model vs Henrard's Results

As discussed in Section 1.2, Henrard *et al* [18] discovered many patients with a stable viral load. This differs from our findings from the MACS trial [27]. However, Henrard *et al* did find some patients (14% of the trial) with large fluctuations (greater than $1log_{10}$ fluctuation from baseline viral load) [18], and we did find some patients from the MACS trial (7% of patients) with stable viral load ($n \geq 1$ from the overlap model).

In terms of our parabola models, we find that the intermittent model applies to MACS [27] data, whereas Henrard *et al* might find that the overlap model is more applicable to his data. However there is some overlap between these two datasets, so it is possible that neither conclusion is wrong.

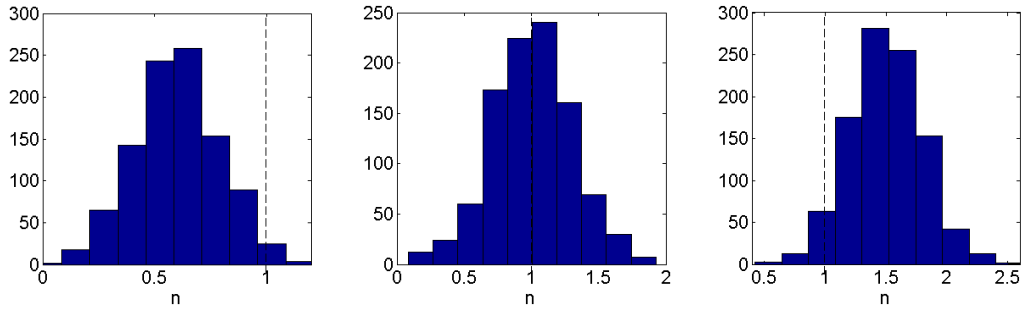


Figure 4.7: Toy histograms for n values in various possible scenarios. The left hand graph would represent a dataset such as MACS viral load data. The right hand graph would represent a dataset such as Henrard *et al*'s [18] trial. The middle graph would be some combination between the two datasets.

As seen in Figure 4.7, there are various possible distributions for n values over a trial. The left hand graph simplistically displays the type of results we find from MACS [27] data, where most patients have large fluctuations in viral load, and only a few display a set point. The opposite extreme is displayed on the right hand graph, which simplistically depicts trials like Henrard *et al*'s [18] where most patients display a set point, with few exceptions.

Then it is important to note that MACS-type data confirms that sequential outburst of virus is possible and that Henrard-type data is simply an example of more frequent outburst. The sequential outburst model is new in mathematical modelling of HIV-1. Some mathematical models, for example in Nowak and May [29], may display synchronous fluctuation of HIV-1 and infected CD4+ (see Figure 1.4), thus modelling our conclusions for the MACS data. But such models cannot model rapid sequential outburst in Henrard-type data, without examining individual populations, as we have done. Our approach is fundamentally "deeper" than standard models (that consider a single population). As an example, Nowak-type models that show oscillating outburst of a single viral strain (in Figure 1.4) cannot adequately describe Henrard-type data.

In the case of CD4+ data, the division between our models for patients was unambiguous. In the case of Henrard's data, we expect $\overline{\tau_L} < 0$ always. However, for this to hold, from equation (33)

$$Max_H < \frac{3}{2} Mean_H, \tag{120}$$

which, using the mean (for the group that progressed to AIDS at an average rate) of 7×10^3 RNA copies/ml [18], the maximum viral load must be $< 1.05 \times 10^4$. Since 86% of the trial had increases less than 10-fold [18], we may assume that the maximum of each patient is $< 7 \times 10^4$ copies/ml. We do not know the spread of the data, so the maximum is most likely less than that. However, we do have possible contradictions for Henrard's data.

4.4 Window Period and End Phase

The window period is defined to be the time between initial infection and before seroconversion (see Figure 1.1). This window period is, by definition of our model, the first viral load parabola, simply of greater height than those that follow. What would cause this greater height? It is known that there is a delay in the immune system [38], and this could explain why the first parabola is brought under control and does not escape, and also why the following populations would reach lower heights. The decrease in parabola height may be an instantaneous change, or the heights may decrease slowly, as simulated in Figure 4.8.

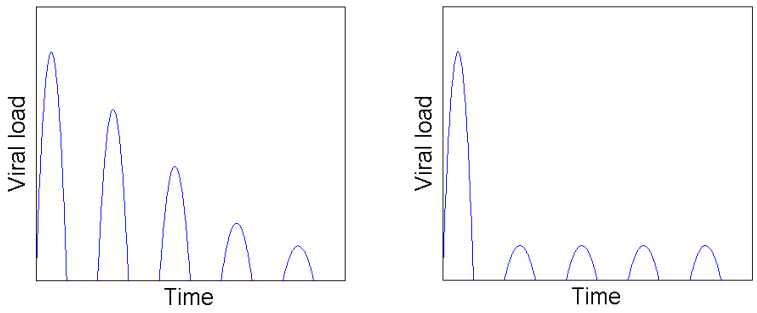


Figure 4.8: A series of parabolas depicting the window period. A series of parabolas with declining heights would represent a steady decay to the plateau, or one parabola of large height followed by regular parabolas of smaller height would represent a sudden change in "set point".

The end phase is defined as the point where a patient's CD4+ count drops below 200 copies/ml and the viral load begins to rise, until death. Like the window period, this may result from a rising accumulation of factors (linear or exponential), or simply because a new equilibrium in these factors has been reached. This could only be determined by more frequent measurements during end phase, which are unavailable in MACS data. Some of the factors that may cause the viral load to rise are: an increase in the height of parabolas (Figure 4.9), an increase in the lifetime of parabolas (Figure 4.10), and a decrease in the gap between parabolas (Figure 4.11).

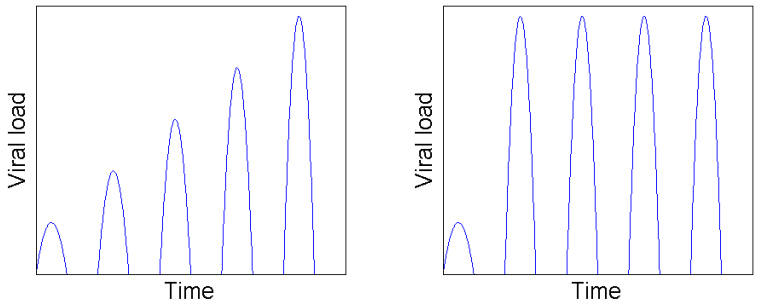


Figure 4.9: A series of parabolas depicting possible viral load end phase. Increasing heights of parabolas, in (a) growth to represent an ever-increasing viral load, or (b) step increase to represent a new "set point" occurring due to some biological change.

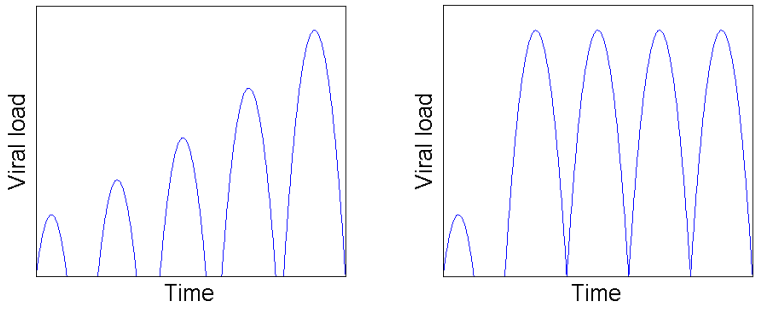


Figure 4.10: A series of parabolas depicting possible viral load end phase. Increasing baselines of parabolas, in (a) growth to represent an ever-increasing viral load, or (b) step increase to represent a new "set point" occurring due to some biological change.

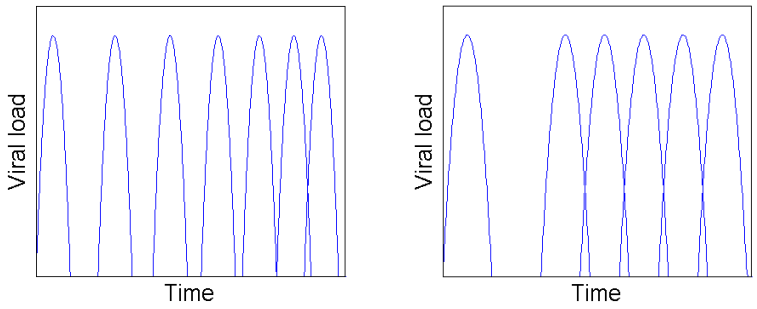


Figure 4.11: A series of parabolas depicting possible viral load end phase. Decreasing distances between starting times of parabolas, in (a) growth to represent an ever-increasing viral load, or (b) step increase to represent a new "set point" occurring due to some biological change.

These factors would be as a result of certain biological events. An increase in the height of the parabolas, with the lifetime remaining the same, would mean that the virus is able to replicate faster, while the immune system responds in the same amount of time. An increase in the lifetime of the parabola would mean that the immune system is taking longer to respond to the threat, allowing the virus more time to replicate. A decrease in the gap between parabolas would mean that new virus populations are escaping the low level reservoir more frequently, possibly due to new, fitter mutations. All of these are plausible.

As seen in Table 4.1, the heights of parabolas are increasing, and the gap between parabolas is decreasing. Our model is independent of τ' so we are unable to determine this change. However, if τ' was to increase, but the total time of $\tau' + \tau_L$ remained the same, the observed changes in the heights of parabolas and gaps between parabolas could occur, since increased time τ' for the virus to replicate would result in a larger height of the parabola. Thus, the rise in viral load seen in the end phase could be due to an increase in τ', or it could be due to a combination of all above-mentioned factors. In all cases we note that from Figures 4.9(b),4.10(b) and 4.11(b) the viral load is adjusting to a higher "set point".

4.5 The Effect of HAART

Viral blips are observed in many patients treated with HAART (two thirds of Di Mascio *et al*'s trial displayed blips [9]). This is viral outburst from an undetectable viral population and is thought to occur from latently infected CD4+ cells [12]. These dynamics of outburst followed by a rest period has clear similarities to our intermittent model. We believe that HAART does not change the essential dynamics of HIV-1: the frequency of outburst and the amplitude is lowered, but the outburst still occurs.

If the dynamics between treated and untreated are unchanged, then it is reasonable to assume the same trigger for outburst. In patients on HAART, the trigger is found to be owing to latently infected cells [12] and so the same source should be considered for untreated patients. It is interesting to consider that outburst may not originate from the low-level reservoir that we assume present during asymptomatic untreated HIV-1 infection.

This brings up questions on latency and the trigger that causes virus to reproduce in a latently infected cell. If this trigger could be found and blocked, HIV-1 could be eliminated without need for HAART, presuming that the mechanisms are adequately determined by our model.

The viral outburst under treatment is found to last approximately a month and is then controlled [9]. Since HAART is designed to stop viral replication, this must be some mutation of the virus impervious to HAART. The viral load is reduced to undetectable levels after approximately a month, and since HAART did not stop the initial replication, this control must be achieved by the immune system. This too is qualitatively similar to our model.

The frequency of outburst is inversely correlated with the CD4+ count at treatment initiation and a positive correlation was found between blip frequency and viral load at start of treatment [9]. This is extremely interesting, since a higher frequency was observed for lower CD4+ counts, correlating with our intermittent model having a smaller time between outburst during end phase (Table 4.1).

In datasets that show less variability in the viral load, where perhaps the overlap model applies instead of the intermittent model, more frequent outburst may be as a result of baseline viral load and CD4+ count.

4.6 The Effect of the Parabolic Shape

The statistical conclusions of Section 4.1 are very strong. A remaining concern is our choice of parabolic profile. Section 3.4 suggests on biological grounds that the mean viral load is overestimated by parabolas. This is not satisfactory because it biases our hypothesis that VL is intermittent. Further, the long tails confuse our conclusion that there is "no overlap" of independent populations. We recall that there is an observed viral baseline in MACS VL data, given by $H_{min} \sim 10^3$ copies/ml. It is then important to show that the baseline does not affect our earlier conclusions.

Next, we may recall that each population is "switched on". These sequential dynamics are certainly built into the model and the shape of the profile does not affect this. However, we essentially model VL on that part of the profile above H_{min}. Then for biological profiles at H_{max}, we must truncate possible biological profiles at H_{min} and ask if the remaining shape gives reasonable $\overline{\tau_L}$. Also, we must reasonably show that long tails can at most add to give H_{min}. Similarly for CD4+ count data.

4.6.1 On the Overlap model

The equations to determine n and h in the overlap model depend on the shape used to model the virus and CD4+ - here we consider CD4+. Shapes other than parabolas may be used, and it is important that our conclusions are not sensitive to the exact shape used. Figures in Section 3 give biological profiles with long tails. Some variants on a parabola are a gaussian curve, $e^{\frac{-t^2}{\delta}}$, and an exponential type curve, $te^{\frac{-t}{\epsilon}}$.

In order to adequately compare these shapes, those shapes with tails are only considered above some height C_{min}, since the number of co-existing viral populations is equal to the number of all viral populations that have ever existed if the tails do not terminate.

Using the same time normalisation of $\frac{\tau}{2}$ as in the overlap model, the parabola is represented by $P(t) = -ht(t-2)$, the gaussian is represented by $G(t) = he^{\frac{-t^2}{\delta}}$, and the exponential type curve is represented by $E(t) = h\frac{t}{\epsilon}e^{\frac{-t}{\epsilon}+1}$. All these curves reach a maximum height of h. Since some of the equations following cannot be solved analytically, we will take this C_{min} from data and assume the height $h = 340$ as seen in Table 4.3, row 1. We are using the CD4+ data here since the overlap model only applies to this data in general. We assume a baseline height $C_{min} = 0.1 \times h = 34$.

For all the curves to have the same area between m and h, we must first determine the times at which C_{min} occurs. For assumed h and C_{min} we find

$$\begin{aligned} P(1 \pm \frac{3}{\sqrt{10}}) &= C_{min} \\ G(\pm\sqrt{\delta ln(10)}) &= C_{min} \\ E(0.03823\epsilon) = E(4.8897\epsilon) &= C_{min} \end{aligned}$$

where the equation $xe^x = y$ does not have an analytical solution for x, and so the times in E(t) had to be numerically approximated.

With the above equations giving the interval of integration, we find the areas under the curves and above height C_{min}, and by setting all areas equal, we find

$$\delta = 0.5991$$
$$\epsilon = 0.5116$$

The relation between all three graphs is seen in Figure 4.12.

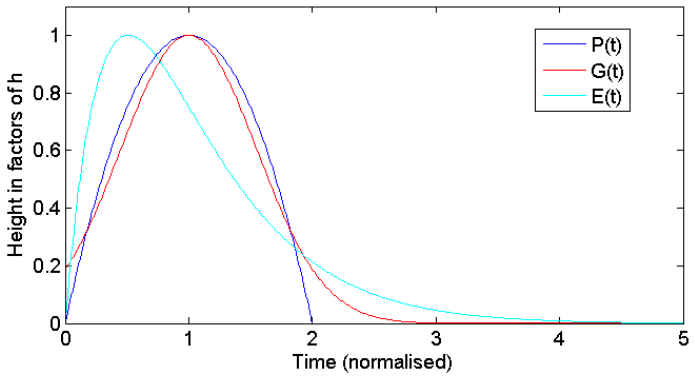


Figure 4.12: The graphs of the parabola, gaussian and exponential-type curve for CD4+. Each have the same area under the curve when terminated at height 10% of the maximum. The agreement between these curves is good.

To match the mean $M = 340$ and range $R = 285$, approximations from Table 4.3, row 1 (the CD4+ MACS data for patients not on treatment), we find that for gaussians, $n = 1.14$ (when terminated at a height of 10% of the maximum height of the gaussian). To match the same mean and range, we find that for exponential type functions, $n = 1.62$ (when terminated at a height of 10% of the maximum height of the profile). These are comparible to the calculated solution for parabolas, $n = 1.4$ (for the same data in the overlap model), and show that the overlap model still applies even when tails are terminated at a minimum height. Figure 4.13 shows how the two profiles $he^{\frac{-t^2}{\delta}}$ and $h\frac{t}{\epsilon}e^{\frac{-t}{\epsilon}+1}$ (with $\delta = 0.5991$ and $\epsilon = 0.5116$ as above) sum to yield similar results to the parabola overlap model.

Looking at the formula $n = \sqrt{\frac{3M}{2R}}$, where the factor $\frac{3}{2} = \frac{1}{\alpha}$ is owing to the parabolic shape. We find overlap for $\alpha < 1.12h$ (where h is the maximum height of the profile) using the M and R in MACS CD4+ untreated data. As seen in Figures 2.1-2.5, this is always true. Recall that for the ODE discussed in Section 3.4, $\alpha = 0.49h$ for infected CD4+ cells. Thus the overlap model is applicable to biological profiles.

4.6.2 On the Intermittent Model

In the intermittent model, $\overline{\tau_L} = \frac{2}{3}\frac{H_{max}}{M'} - 1$, and the factor $\frac{2}{3}$ is owing to the parabolic shape. The mean height of a parabola is $\frac{2}{3}h$. If we assume the mean height of some general profile to be α, we find that a gap still exists between viral populations for $\alpha \geq \frac{1}{3}h$, if we use the MACS average H_{max} and M' for patients not on treatment.

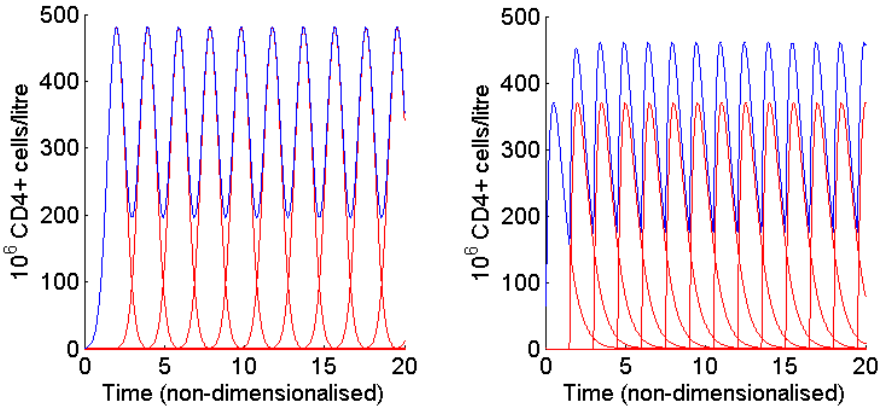


Figure 4.13: Gaussians and exponential-type curves sum to reach observed mean and range for MACS CD4+ data for patients not on treatment. The agreement between these profiles and the parabola overlap model is good.

In Figures 2.1-2.5 in Section 2.3, we examined various possible profiles and their resulting values for α. For the viral load ODE solution discussed in Section 3.4, we find $\alpha = 0.4h$. This will result in a smaller value $\overline{\tau_L} \approx 0.2$, but the intermittent model still holds in 98% of MACS patients examined in this dissertation. Thus even for profiles with tails, we still find that the intermittent model applies to MACS viral load data. The general principle of outburst from a low-level population still holds for all above profiles.

4.6.3 On Ignoring Tails

Ignoring tails of viral profiles will result in a constant error on our heights, that is taken to be a low level viral population of height insignificant to the maximum height of the parabolas. That tails do not accumulate significantly over time is shown below. If we assume that the tail dominates over the rise, we will simply sum exponential tails, such as

$$S(t) = e^{\frac{-t}{\delta}} + e^{\frac{-(t-x)}{\delta}} + e^{\frac{-(t-2x)}{\delta}} + ...$$
$$S(t+x) = e^{\frac{-(t+x)}{\delta}} + e^{\frac{-t}{\delta}} + e^{\frac{-(t-x)}{\delta}} + e^{\frac{-(t-2x)}{\delta}} + ...$$

where x is a small positive time step. Using the forward difference formula, the gradient is

$$S'(t) = \frac{1}{x} e^{\frac{-(t+x)}{\delta}} \geq 0$$

$$\forall \delta \text{ since } x > 0.$$

Now say

$$\frac{1}{x} e^{-(t+x)} \delta \neq 0$$

This implies

$$e^{\frac{-(t+x)}{\delta}} \neq 0, \text{ since } x << \infty.$$

Supposing that the exponential significantly decays by e^{-3}, we take

$$\frac{t+x}{\delta} < 3,$$

or

$$t_{crit} < 3\delta - x.$$

Since $\delta > 0$, there is an upper bound for t_{crit} where a positive gradient in the sum over tails occurs. Thus summing over many tails does not result in an increase in the sum.

5 Conclusions

In the MACS data we find high levels of variation in viral load, and thus the idea of an equilibrium viral load does not apply. Instead, we find that viral load in the asymptomatic stage is well-modelled by sequential outburst of independent viral populations with stable mean, each with a life time $\sim$ a month. The notion of "set point" can refer only to stable mean. We have argued that viable viral populations are born on one month time scales. The success of the intermittent model implies that if outburst is owing to mutations, viable mutations occur on similar time scales. We have noted that outburst frequency and amplitude increase for a small sample of AIDS patients relative to levels in the asymptomatic phase.

If the outburst is not triggered by a mutation, then mutation *per se* is not dangerous to patients. Treatment by HAART is often characterised by viral outburst, and so we find that behaviour on treatment mimics dynamics in untreated patients in the sense of outburst. The effect of HAART is to increase the time between such outbursts and to decrease the amplitudes.

Outburst during HAART is known to come from latently infected CD4+ cells, thus it is reasonable to believe that outburst during untreated infection has the same source, and is not outburst from a low-level reservoir of free virions. The qualitative agreement between untreated asymptomatic dynamics, AIDS dynamics and HAART dynamics is striking.

In Henrard-type data, with small variation in viral load, we suggest that viral outburst simply occurs more frequently. In HAART, blip frequency is found to correlate with CD4+ counts at start of treatment and this suggestion could be tested by looking at a similar correlation in untreated patients. Although Henrard-type viral loads can be modelled by ODE models with equilibrium as reviewed in Section 1.3, these models cannot account for outburst dynamics. The overlap model better fits the CD4+ count data, which means that there are co-existing populations of infected CD4+ cells.

In the case of CD4+, the overall statistics remain consistent with $n > 1$ and overlap, and in this sense are unaffected by measurement error. In 43% of patients, both models are implied, which is a contradiction. We have concluded that this is owing to the marginal nature of overlap.

Since viral load and CD4+ count are biologically related, viral outburst must be in phase with the dynamics of CD4+ count. We argue that the time scale of viral load outburst is half the time scale of an infected CD4+ cell population during asymptomatic phase, for MACS data, with our assumptions. Mathematical models suggest that the starting times of each are the same or similar. Correlation of viral load and CD4+ will be difficult to detect (see Figures 4.2 and 4.4). AIDS patients in the MACS trial displayed markedly shorter gaps between intermittent viral populations. This emphasises the importance of the trigger for such outburst, which should be investigated.

Since the latent CD4+ population is determined to be genetically stable [12] and that memory T-cells are long-lived [38], we must have *new* viable mutations occurring to result in the outburst. Once this mutation occurs, there may be a favourable environment for further mutation of the reproducing virus.

Our findings are robust under changes in profile shape, assuming the existence of a low-level viral population that hides long tails of virus. We have argued that the immune system memory will prevent secondary outburst from a population that has already been eliminated and that the tails will not add up to significant values over the patient life-time.

Further work needs to be done on the time scales of populations. This could be done with a dataset with exact measurement dates, where an optimisation problem could be

designed to determine the best fit to the data. Using the mean and range of the data as before, we are interested in determining n for the overlap model and $\overline{\tau_L}$ for the intermittent model. Thus with only one variable to determine in each case, the baseline τ would simply be adjusted between reasonable values (as discussed previously) and the least-squares error between observed data points and the nearest point on the model would determine the best fit.

The ODE model needs to be explored as well. For this dissertation it was sufficient to see that the phasing between viral load and CD4+ count was possible. If we could fit the model to accurate data for all variables, we could determine some constants that could prove interesting, such as the rate d, at which the immune system kills off free virions.

Specific counter of latently infected CD4+ cells is of interest, but there is pessimism that this can be achieved, owing to a small latent CD4+ fraction, the small amount of integrated viral RNA in the host chromosome and of course absence of replication [37].

The trigger for outburst should be investigated. If the mechanisms for the trigger could be countered, it is possible that more effective treatments can be achieved.

6 References

[1] Bajaria *et al*. Dynamics of Naive and Memory CD4+ T Lymphocytes in HIV-1 Disease Progression. *JAIDS*, 30:41–58, 2002.

[2] Ball CL, Gilchrist MA, Coombs D. Modelling within-host evolution of HIV: mutation, competition and strain replacement. *Bulletin of Mathematical Biology*, 69:2361–2385, 2007.

[3] Ball SC *et al*. Comparing the ex vivo fitness of CCR5-tropic human immunodeficiency virus type 1 isolates of subtypes B and C. *J Virol*, 77:1021–1038, 2003.

[4] Bartlett JG. Ten Years of HAART: Foundation for the Future. In *13th Conference on Retroviruses and Opportunistic Infections*, 2006.

[5] Callaway DS & Perelson AS. HIV-1 infection and low steady state viral loads. *Bulletin of Mathematical Biology*, 64:29–64, 2002.

[6] Chun TW *et al*. Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature*, 387:183–187, 1997.

[7] Coombs RW, Reichelderfer PS & Landay AL. Recent observations on HIV type-1 infection in the genital tract of men and women. *AIDS*, 17:455–480, 2003.

[8] de Wolf F *et al*. AIDS prognosis based on HIV-1 RNA, CD4+ T-cell count and function: markers with reciprocal predictive value over time after seroconversion. *AIDS*, 11:1799–1806, 1997.

[9] Di Mascio M, Ribeiro RM, Markowitz M, Ho DD, Perelson AS. Modelling the long-term control of viremia in HIV-1 infected patients treated with antiretroviral therapy. *Mathematical Biosciences*, 188:47–62, 2003.

[10] Eriksson N *et al*. Viral population estimation using pyrosequencing. *PLoS Comput Biol*, 4(5):e1000074, 2008.

[11] Feinburg MB. Changing the natural history of HIV disease. *The Lancet*, 348:239–246, 1996.

[12] Finzi *et al*. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science*, 278:1295–1300, 1997.

[13] Fraser *et al*. Variation in HIV-1 set-point viral load: Epidemiological analysis and an evolutionary hypothesis. *PNAS*, 104(44):17441–17446, 2007.

[14] Giorgio JV *et al*. Quality Control in the Flow Cytometric Measurement of T-Lymphocyte Subsets: The Multicenter AIDS Cohort Study Experience. 55:173–186, 1990.

[15] Goulder PJ *et al*. Substantial differences in specificity of HIV-specific cytotoxic T cells in acute and chronic HIV infection. *J Exp Med*, 193:181–194, 2001.

[16] Hahn S, Gehri R, Erb P. Mechanism and Biological Significance of CD4-mediated Cytotoxicity. *Immunol Review*, 146:57–79, 1995.

[17] Heckerman D, Kadie C, Listgarten J. Leveraging information across HLA alleles/supertypes improves epitope prediction. *J Comput Biol*, 14(6):736–746, 2006.

[18] Henrard *et al*. Natural history of HIV-1 cell-free viremia. *JAMA*, 274(7):554–558, 1995.

[19] Ho DD. Dynamics of HIV-1 Replication in Vivo. *J. Clin. Invest*, 99(11):2565–2567, 1997.

[20] Ho *et al*. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature*, 373:123–126, 1995.

[21] Hockett R *et al*. Constant mean viral copy number per infected cell in tissues regardless of high, low, or undetectable plasma HIV RNA. *J Exp Med*, 189:1545–1554, 1999.

[22] Langford SE, Ananworanich J, Cooper D. Predictors of disease progression in HIV infection: a review. *AIDS Research and Therapy*, 4:11, 2007.

[23] Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund o, Nielsen M. Large scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics*, 8:424, 2007.

[24] Lindbäck *et al*. Viral dynamics in primary HIV-1 infection. *AIDS*, 14(15):2283–2291, 2000.

[25] Mellors J *et al*. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science*, 272:1167–1170, 1996.

[26] Mellors *et al*. Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion. *Annals of Internal Medicine*, 122(8):573–579, 1995.

[27] Multicenter AIDS Cohort Study, Public Data Set: Release P16.

[28] Nowak M. HIV mutation rate. *Nature*, 347:522, 1990.

[29] Nowak M, May R. *Virus Dynamics: Mathematical principles of immunology and virology*. Oxford University Press, 2000.

[30] Nowak MA & May RM. Mathematical biology of HIV infections: antigenic variation and diversity threshold. *Math Biosci*, 106:1–21, 1991.

[31] Nowak MA, May RM, Anderson RM. The evolutionary dynamics of HIV-1 quasispecies and the development of immunodeficiency disease. *AIDS*, 4:1095–1103, 1990.

[32] Nowak MA *et al*. Antigenic diversity thresholds and the development of AIDS. *Science*, 254:963–969, 1991.

[33] Oguntibeju OO, van den Heever WMJ & Van Schalkwyk FE. A review of the epidemiology, biology and pathogenisis of HIV. *Journal of Biological Sciences*, 7(8):1296–1304, 2007.

[34] Overbaugh J, Bangham CRM. Selectional forces and contraints on retroviral sequence variation. *Science*, 292:1106–1109, 2001.

[35] Perelson AS, Nelson PW. Mathematical analysis of HIV-1 dynamics in vivo. *SIAM review*, 41(1):3–44, 1999.

[36] Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 Dynamics in Vivo: Virion Clearance Rate, Infected Cell Lifespan, and Viral Generation Time. *Science*, 271:1582–1586, 1996.

[37] Persaud D *et al*. Latency in Human Immunodeficiency Virus Type 1 Infection: No Easy Answers. *Journal of Virology*, 77(3):1659–1665, 2003.

[38] Prescott LM, Harley JP, Klein DA. *Microbiology*. McGraw-Hill Higher Publication, Boston, 2005.

[39] Quer J, Hershey CL, Domingo E, Holland JJ, Novella IS. Contingent neutrality in competing viral populations. *Journal of Virology*, 75:7315–7320, 2001.

[40] Schacker *et al*. Biological and virologic characteristics of primary HIV infection. *Annals of Internal Medicine*, 128(8):613–620, 1998.

[41] UNAIDS. Questions and Answers II, Section I, 2004.

[42] Vidal *et al*. Lack of evidence of a stable viral load set-point in early stage asymptomatic patients with chronic HIV-1 infection. *AIDS*, 12:1285–1289, 1998.

[43] Wei *et al*. Viral dynamics in human immunodeficiency virus type 1 infection. *Nature*, 373:117–122, 1995.

[44] Wu J & Zhang J-T. The study of long-term HIV dynamics using semi-parametric non-linear mixed-effects models. *Statistics in medicine*, 21:3655–3675, 2002.

[45] Yu XG *et al*. Consistent patterns in the development and immunodominance of HIV-1 specific CD8+ T-cell responses following acute HIV-1 infection. *Journal of Virology*, 76:8690–8701, 2002.