Estimation of Within-Host Reproductive Number Distributions from HIV-1 Viral Load Data

Jennifer Lamb[1], Anibal Y. Lopez-Correa[2],
Kamau Romero[3], Xiaohong Wang[4],
Ariel Cintron-Arias[4]

[1] Department of Mathematics, Northern Kentucky University, Highland Heights, KY

[2] Department of Mathematics, University of Puerto Rico-Cayey, Cayey, PR

[3] Department of Mathematics, Brown University, Providence, RI

[4] Department of Mathematics, Arizona State University, Tempe, AZ

July 28, 2006

Abstract

The basic reproductive number $\mathcal{R}_0$ is defined as the expected number of secondary cases generated by a "typical" infected individual in a totally susceptible population. In the context of within-host viral dynamics, we study the reproductive number of virions and infected cells. In this study we utilize empirical data from ten anti-retroviral, drug-naive, patients infected with HIV-1, published by Stafford *et. al.*[18]. We apply a Bayesian procedure which results in the estimation of the full probability distribution of $\mathcal{R}_0$, due to Bettencourt and Ribeiro[5]. In addition, an uncertainty and sensitivity analysis of $\mathcal{R}_0$ is employed to assess the role played by variation of model parameters in within-host viral dynamics.

1 Introduction

In 1982 the United States Center for Disease Control officially recognized the term Acquired Immune Deficiency Syndrome (AIDS). This was a known disease that was thought to be a rare and deadly pneumonia found primarily in young homosexual men in the early 1980's. In 1981 it was then realized that the disease did not only infect homosexual males, but was capable of infecting the entire population. It was not until 1984 that a virus, later named the Human Immunodeficiency Virus (HIV), was discovered to be the cause of AIDS by Luc Montagnier of the Pasteur Institute and Robert Gallo of the National Cancer Institute [19].

HIV is a retrovirus that originated from chimpanzees. It is thought by many researchers that transmission of the virus from chimpanzees to humans most likely occurred from blood entering the wounds of humans in sub-Saharan Africa from killing the animals for food [10, 21]. From there the virus has been passed from human to human through three primary modes: as a sexually transmitted disease, through parental exposure, and through infected needles used in intravenous drug use [21].

Once HIV has entered the blood stream there are many immune responses that occur. The following is a very simple explanation of what happens within the body. After HIV, or any pathogen, enters the body it encounters dendritic cells. When dendritic cells encounter a foreign organism, they use their special receptors to decide whether the organism is nontoxic or pathogenic. If the dendritic cells decide that the organism is pathogenic, they then carry fragments of the pathogen to the lymph nodes where they stimulate the immune response from the $CD4^+$ helper T-cells. These cells then stimulate the B cells to produce antibodies that connect to the specific pathogen and immobilize it, stopping it

from causing infection. The B cells also serve the purpose of activating memory cells primarily named $CD4^+$ and $CCR5^+$. These cells insure that upon re-exposure to the same pathogen, a stronger and quicker immune response can be made. B cells also produce $CD8^+$ Cytotoxic T-cells that destroy infected cells [10].

HIV primarily infects helper T-cells and macrophages of the immune system [12]. Without a large supply of helper T-cells, the immune system cannot tell B cells to produce $CD8^+$ Cytotoxic T-cells, therefore leaving the body defenseless to HIV [10]. Another way that HIV defeats the human immune system is through mutation. With every replication of itself, there is roughly one mutation. Often times these mutations change the virus so much that antibodies and $CD8^+$ Cytotoxic T-cells can not recognize it, enabling the virus to infect other cells and continue replicating [12]. An additional way HIV survives is through hiding. HIV is capable of viral latency, this is when the virus can live in a cell, but does not reproduce itself until later, practically making it invisible to the immune system [3].

In order for HIV to reproduce itself it first needs to enter a cell. Once inside the cell, reverse transcriptase copies the viral RNA into viral DNA. This viral DNA then integrates into the cells DNA and becomes part of the cells genetic code. This new DNA then creates two strands of RNA. One of these strands is translated into subunits of HIV that will eventually become enzymes such as protease, integrase, and reverse transcriptase. The other strand is then converted into genetic material for the new viruses. The strand of subunits of HIV is then cleaved by the protease enzyme. After cleavage, the subunits combine to make new virions. These new virions then pinch off and enter the blood stream, looking for other cells to infect [15].

HIV progression is broken up into four stages. The first stage is what is usually called the primary HIV infection stage. During this stage HIV starts attacking the body, forcing an immune response which produces HIV antibodies and cytotoxic lymphocytes. This process is often called seroconversion, and usually lasts for a few weeks [4]. The second stage in HIV infection is the Clinically Asymptomatic Stage. During this stage a persons HIV virion count per microliter will drop to very low levels. The person still remains infectious, but no major symptoms are apparent. As the immune system fails, a new stage in HIV infection begins. This is the Symptomatic HIV Infection Stage. During this stage of the infection, many opportunistic infections start to emerge. These infections usually affect the respiratory system, gastro-intestinal system, the central nervous system, and the skin. The fourth and final stage is when HIV progresses into AIDS [4].

Usually the onset of AIDS occurs approximately 10 years after the initial HIV infection [8]. The average healthy individual usually has between 800 and 1,200 $CD4^+$ T-cells per micro-liter of blood [9]. Once the $CD4^+$ T-cells drop below 500 per micro-liter, a person is defined as having clinical AIDS. Eventually, after many $CD4^+$ T-cells have been infected or destroyed, and when the $CD4^+$ T-cells drop below 200 per micro-liter, the immune system becomes susceptible to opportunistic infections [21]. As of December 2005 AIDS has killed at least 25 million people since in was first recognized in 1981. Currently there are an estimated 40.3 million people living with HIV [1].

2 Classic Viral Dynamics Model

Following are our equations for the HIV model with x representing the uninfected cells, y representing the infected cells, and v representing the free HIV virus [13, 14].

$$\dot{x} = \Lambda - \mu x - \beta x v \tag{1}$$

$$\dot{y} = \beta x v - a y \tag{2}$$

$$\dot{v} = k y - s v \tag{3}$$

The model that we are using is a predator-prey model. The predator can be represented as the HIV virus and the prey as the uninfected cells. In reducing this system we suppose that x is constant. Since we are assuming that x is a constant then $\dot{x}$ is equal to 0.

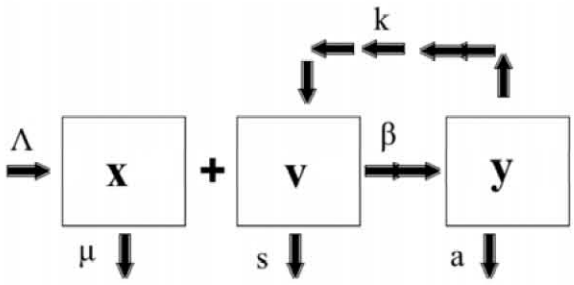

Figure 1: Sketch of the model of virus dynamics [13]

Thus we can look for the disease free equilibrium.

$$0 = \Lambda - \mu x^* - \beta x^* v \tag{4}$$

$$0 = \beta x^* v - ay \tag{5}$$

$$\rightarrow \beta x^* v = 0, \rightarrow ay = 0 \tag{6}$$

$$0 = ky - sv \tag{7}$$

$$\rightarrow ky = 0, \rightarrow sv = 0 \tag{8}$$

$$\Rightarrow 0 = \Lambda - \mu x^* \tag{9}$$

$$\rightarrow x^* = \frac{\Lambda}{\mu} \tag{10}$$

The disease free equilibrium is the equilibrium point where no disease exists. In order to solve for this point the infected cell and virion class must both be set to 0. Thus our disease free equilibrium point is ($\frac{\Lambda}{\mu}$,0,0).

The basic reproductive number represents the number of secondary infections caused by a virion in the uninfected cells. First we have to find the matrix that represents the new cases of infected cells:

$$F = \begin{pmatrix} \beta xv \\ 0 \\ 0 \end{pmatrix} \tag{11}$$

The Jacobian of this matrix evaluated at the disease free equilibrium is:

$$JF(\frac{\Lambda}{\mu}, 0, 0) = \begin{pmatrix} 0 & 0 & \frac{\beta\Lambda}{\mu} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{12}$$

Matrix A:

$$A = \begin{pmatrix} ay \\ -\Lambda + \mu x + \beta x v \\ -ky + sv \end{pmatrix} \tag{13}$$

The Jacobian of this matrix evaluated at the disease free equilibrium is:

$$JA^{-1} = (\frac{\Lambda}{\mu}, 0, 0) = \begin{pmatrix} \frac{1}{a} & 0 & 0 \\ \frac{-\beta k \Lambda}{a\mu^2 s} & \frac{1}{\mu} & \frac{-\beta\Lambda}{\mu^2 s} \\ \frac{k}{as} & 0 & \frac{1}{s} \end{pmatrix} \tag{14}$$

$$JF * JA^{-1} = \begin{pmatrix} \frac{\beta k \Lambda}{a\mu s} & 0 & \frac{\beta\Lambda}{\mu s} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{15}$$

Thus, $\mathcal{R}_0$=$\frac{\beta k \Lambda}{a\mu s}$. Our $\mathcal{R}_0$ is the basic reproductive number of the new virus in the host. If $\mathcal{R}_0 < 1$ then the virus is decreasing. If $\mathcal{R}_0 > 1$ then the virus is increasing, and the disease is killing the T-cells.

How Virions produce more Virions

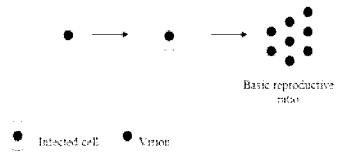

Figure 2: This figure shows how a virion infects a cell and in turn produces more virions [13].

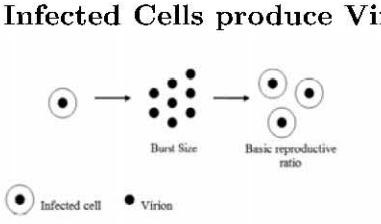

Figure 3: This figure shows how an infected cell produces virions who then go on to infect more cells [13].

3 Empirical Longitudinal Data on Patient Viral Loads

3.1 General statements about HIV diagnosis

Most HIV tests measure the level of HIV antibodies the body has produced to fight HIV. The two most common tests used are the Enzyme Immunoassay (EIA) and the Enzyme-linked Immunosorbent Assay (ELISA). A hindrance for these two tests is that sometimes it takes the body three to six months from the time of infection for enough HIV antibodies to be produced in order to be detected in the bloodstream [7].

Viral load testing is used mostly for patients that are already diagnosed with HIV. The reason for this is that viral load testing is expensive, as well as it is used as a tool for doctors to monitor their patients and decide whether a treatment is effective, or whether a change in treatments will be beneficial. Three viral load tests that are commonly used are the branched-chain DNA test (bDNA), the reverse transcriptase polymerase chain reaction test (RT-PCR), and the nucleic acid sequence based amplification test (NASBA) [2]. These tests are used to measure the HIV RNA levels in the blood [20].

Parameters

Parameter	Meaning	Units
Λ	Growth rate of uninfected cells	$\frac{virion}{(mL)(day)}$
μ	Death rate of uninfected cells	$\frac{1}{day}$
β	Rate at which uninfected cell becomes infected	$\frac{mL}{(virion)(day)}$
a	Death rate of the infected cells	$\frac{1}{day}$
k	Rate of infected cells becomes virion	$\frac{1}{day}$
s	Death rate of the virus	$\frac{1}{day}$
$x(t)$	Number of Uninfected cells	$virion$
$y(t)$	Number of Infected cells	$virion$
$v(t)$	Number of Virus	$virion$

Table 1: Parameters used and their meanings

4 Recursive Formula of New Virions

Since we are assuming that x is constant, we need to find an expression for $v(t)$, that does not depend on v or y. We will find a recursive method in order obtain the number of new virions produced from the new virions of the previous time. To find $v(t)$ we can take (3).

$$\ddot{v} = k\dot{y} - s\dot{v} \tag{16}$$

$$\ddot{v} = k(\beta x^* v - ay) - s\dot{v} \tag{17}$$

$$\ddot{v} = k\beta x^* v - kay - s\dot{v} \tag{18}$$

Taking equation (3) we can then solve for y.

$$ky = \dot{v} + sv \tag{19}$$

Then if we multiply both sides by a:

$$aky = a\dot{v} + asv \tag{20}$$

Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10
22, 27.2	3, 469.8	0, 766.8	0, 153.0	0, 228.2	0, 939.26	0, 1350.6	0, 2217.7	0, 216.4	4, 8057.2
43, 210	11, 1600	7, 947.6	5, 284.0	2, 599.2	3, 1485.0	4, 2398.6	4, 2427.9	5, 355.2	9, 9622.8
78, 85.9	15, 42.8	9, 706.2	6, 216.0	6, 2617.4	8, 701.6	9, 337.2	7, 2200.4	8, 355.4	10, 7830.0
106, 81.1	43, 41.7	15, 14.4	14, 143.0	14, 169.6	10, 564.0	12, 340.6	11, 1134.3	12, 146.8	14, 715.81
- -	- -	29, 2.3	21, 30.2	21, 93.7	15, 106.5	16, 202.3	14, 705.9	19, 100.9	16, 213.79
146, 46.2	71, 12.22	36, 1.1	32, 6.4	42, 165.6	17, 11.2	19, 169.7	18, 447.8	29, 34.7	18, 121.03
183, 60.1	99, 14.17	50, 1.0	39, 4.1	- -	22, 87.3	23, 141.4	21, 412.7	57, 11.4	28, 16.36
230, 82.8	129, 18.2	57, 1.8	46, 5.85	98, 127.0	24, 20.6	26, 56.48	26, 302.1	- -	30, 11.79
268, 103.	197, 70.8	64, 2.1		203, 65.9	29, 14.78	30, 182.75	29, 118.8	121, 17.3	35, 31.75
358, 72.1	255, 16.3			329, 144.7	36, 27.5	50, 267.0	33, 248.8	197, 90.1	42, 24.05
435, 79.4	330, 81.2				- -	60, 182.7	36, 173.6	280, 68.2	51, 16.257
489, 70.4					64, 6.32	- -	40, 131.3	376, 55.3	- -
519, 207.					273, 2.27	213, 186.3	49, 259.1	525, 94.5	84, 19.59
534, 42.6					288, 5.64	551, 89.4	- -	604, 34.4	177, 41.17
584, 10.8					347, 14.55		56, 132.24	645 61.7	211, 61.95
610, 54.2					430, 13.6		63, 103.2	757, 55.9	239, 137.77
687, 22.3					478, 13.1		75, 117.1	776, 52.7	
778, 40.8							547, 5.62		
							659, 24.24		

Table 2: This table is taken directly from Stafford et. al. [18]. Data points are presented as ordered pairs with first number in each entry representing a relative time in days and the second number in each entry the virus concentration in thousands of HIV-1 RNA copies ml^1. A horizontal line in a column indicates only the data points above the line were used in parameter estimation. All points were used if there is no horizontal line. The times listed for patient 9 are from 35 days following initial infection (Borrow et al., 1997). Patients 1 and 2 are patient numbers 1019 and 1113 from University of Washington study. Patients 3-9 are JSW-DAAR, CMO-DAAR, HOBR-SHAW, SUMA-SHAW, BORI-SHAW, INME-SHAW, and WEAU-SHAW from Aaron Diamond AIDS Research Center, respectively. Data for patient 10 are from patient DR from the Cedars-Sinai Medical Center in Los Angeles, CA. [18]

	Patient 1	Patient 2	Patient 5	Patient 7	Patient 8	Patient 9
Parameter	μL, mL	μL, mL	μL, mL	μL, mL	μL, mL	μL, mL
Λ	1.3E-1, 1.30E+2	2.E-1, 2.E+2	1.70E-1, 1.70E+2	1.70E-1, 1.70E+2	8.50E-2, 8.50E+1	6.00E-2, 6.00E+1
β	4.6E-4, 4.60E-7	3.60E-4, 3.60E-7	6.30E-4, 6.30E-7	8.00E-4, 8.00E-7	6.60E-4, 6.60E-7	2.50E-3 , 2.50E-6
Init x	1.E+1, 1.E+4	1.00E+1, 1.00E+4	1.00E+1, 1.00E+4	1.00E+1, 1.00E+4	1.00E+1, 1.00E+4	1.00E+1, 1.00E+4
Init y	0.00, 0.00	0.00, 0.00	0.00, 0.00	0.00, 0.00	0.00, 0.00	0.00, 0.00
Init v	1.E-9, 1.E-6	1.E-9, 1.00E-6	1.00E-9, 1.00E-6	1.00E-9 , 1.00E-6	1.00E-9, 1.00E-6	1.00E-9, 1.00E-6
μ	1.3E-2, 1.3E-2	2.E-2, 2.00E-2	1.70E-2, 1.70E-2	1.70E-2, 1.70E-2	8.50E-3, 8.50E-3	6.00E-3 , 6.00E-3
a	4.E-1, 4.E-1	8.E-1, 8.E-1	3.90E-1, 3.90E-1	3.10E-1, 3.10E-1	1.70E-1, 1.70E-1	1.30E-1, 1.30E-1
k	980, 980	1800, 1800	870, 870	730, 730	830, 830	110, 110
s	3, 3	3, 3	3, 3	3, 3	3, 3	3, 3

Table 3: Patients used in Bayesian Estimation

Now we can substitute this in (18) and we obtain:

$$\ddot{v} = k\beta x^* v - (a\dot{v} + asv) - s\dot{v} \tag{21}$$

$$0 = \ddot{v} + \dot{v}(ka + s) + v(as - k\beta x^*) \tag{22}$$

Now we can find the characteristic equation.

$$0 = \rho^2 + (a + s)\rho + (as - k\beta x^*) \tag{23}$$

$$\rho_{1,2} = \frac{-a - s \pm \sqrt{(a + s)^2 - 4(as - k\beta x^*)}}{2} \tag{24}$$

$$\rho_{1,2} = \frac{-a - s \pm \sqrt{a^2 + 2as + s^2 - 4as(1 - \frac{k\beta x^*}{as})}}{2} \tag{25}$$

$$\rho_{1,2} = \frac{-a - s \pm \sqrt{(a + s)^2 - 4as(1 - \mathcal{R}_0)}}{2} \tag{26}$$

This is true if and only if the discriminant is greater than 0. This condition can be shown as:

$$0 > as - k\beta x^* \tag{27}$$

$$0 > as(1 - \frac{k\beta x^*}{as}) \tag{28}$$

$$\rightarrow 0 > 1 - \mathcal{R}_0 \tag{29}$$

$$\rightarrow \mathcal{R}_0 > 1 \tag{30}$$

Now we have the general solution for $v(t)$ and $v(0)$.

$$v(t) = c_1 e^{\rho_1 t} + c_2 e^{\rho_2 t} \tag{31}$$

$$v(0) = c_1 + c_2 \tag{32}$$

Now we must find an equation for $v(t + \tau_i)$

$$v(t - \tau_{i-1}) = c_1 e^{\rho_1 (t - \tau_{i-1})} + c_2 e^{\rho_2 (t - \tau_{i-1})} \tag{33}$$

$$v(t - \tau_{i-1}) = c_1 e^{\rho_1 t} e^{-\rho_1 \tau_{i-1}} + c_2 e^{\rho_2 t} e^{-\rho_2 \tau_{i-1}} \tag{34}$$

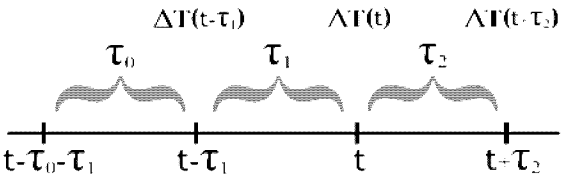

Figure 4: Diagram of New Cases in a Certain Period of Time

From (31) we know that $c_1 e^{\rho_1 t} = v(t) - c_2 e^{\rho_2 t}$. So we can substitute in (34):

$$v(t - \tau_{i-1}) = (v(t) - c_2 e^{\rho_2 t}) e^{-\rho_1 \tau_{i-1}} + c_2 e^{\rho_2 t} e^{-\rho_2 \tau_{i-1}} \tag{35}$$

$$\rightarrow c_2 e^{\rho_2 t} = \frac{v(t - \tau_{i-1}) - v(t) e^{-\rho_1 \tau_{i-1}}}{e^{-\rho_2 \tau_{i-1}} - e^{-\rho_1 \tau_{i-1}}} \tag{36}$$

Now we can write $v(t + \tau_i)$ as an equation that depends on the last two equations.

$$v(t + \tau_i) = c_1 e^{\rho_1 (t + \tau_i)} + c_2 e^{\rho_2 (t + \tau_i)} \tag{37}$$

$$v(t + \tau_i) = c_1 e^{\rho_1 t} e^{\rho_1 \tau_i} + c_2 e^{\rho_2 t} e^{\rho_2 \tau_i} \tag{38}$$

$$\rightarrow v(t + \tau_i) = (v(t) - c_2 e^{\rho_2 t}) e^{\rho_1 \tau_i} + c_2 e^{\rho_2 t} e^{\rho_2 \tau_i} \tag{39}$$

$$\rightarrow v(t + \tau_i) = v(t) e^{\rho_1 \tau_i} + \left(\frac{v(t - \tau_{i-1}) - v(t) e^{-\rho_1 \tau_{i-1}}}{e^{-\rho_2 \tau_{i-1}} - e^{-\rho_1 \tau_{i-1}}}\right)(e^{\rho_1 \tau_i} - e^{\rho_2 \tau_i}) \tag{40}$$

Using this equation that is function of $v(t)$, $v(t - \tau_{i-1})$, a, s and $\mathcal{R}_0$; we know that $\frac{dT}{dt} = ky$. Then we have to solve for y. Since we have the solution for $v(t)$ we can obtain $\dot{v}(t)$.

$$v(t) = c_1 e^{\rho_1 t} + c_2 e^{\rho_2 t} \tag{41}$$

$$\dot{v}(t) = \rho_1 c_1 e^{\rho_1 t} + \rho_2 c_2 e^{\rho_2 t} \tag{42}$$

From (3) we have can solve for y.

$$ky - sv = \rho_1 c_1 e^{\rho_1 t} + \rho_2 c_2 e^{\rho_2 t} \tag{43}$$

$$\Rightarrow y = \frac{sv + \rho_1 c_1 e^{\rho_1 t} + \rho_2 c_2 e^{\rho_2 t}}{k} \tag{44}$$

If we substitute v from (31) we obtain:

$$y = \frac{s(c_1 e^{\rho_1 t} + c_2 e^{\rho_2 t}) + \rho_1 c_1 e^{\rho_1 t} + \rho_2 c_2 e^{\rho_2 t}}{k} \tag{45}$$

$$\Rightarrow y = \frac{(\rho_1 + s) c_1 e^{\rho_1 t} + (\rho_2 + s) c_2 e^{\rho_2 t}}{k} \tag{46}$$

As $\frac{dT}{dt} = ky$ then we know that $\frac{dT}{dt} = (\rho_1 + s) c_1 e^{\rho_1 t} + (\rho_2 + s) c_2 e^{\rho_2 t}$.

Then we can make the following two assumptions if τ is very small:

$$\Delta T(t - \tau_1) = (\rho_1 + s)\tau_0 c_1 e^{\rho_1 (t - \tau_1)} + (\rho_2 + s)\tau_0 c_2 e^{\rho_2 (t - \tau_1)} \tag{47}$$

$$\Delta T(t) = (\rho_1 + s)\tau_1 c_1 e^{\rho_1 t} + (\rho_2 + s)\tau_1 c_2 e^{\rho_2 t} \tag{48}$$

Using a system of two equations we can solve for the two unknown values:$(\rho_1 + s) c_1 e^{\rho_1 t}$ and $(\rho_2 + s) c_2 e^{\rho_2 t}$.

$$(\rho_1 + s) c_1 e^{\rho_1 t} = \frac{\Delta T(t)}{\tau_1} - \frac{\Delta T(t - \tau_1) - \frac{\tau_0}{\tau_1}\Delta T(t) e^{-\tau_1 \rho_1}}{\tau_0 (e^{-\rho_2 \tau_1} - e^{-\rho_1 \tau_1})} \tag{49}$$

$$(\rho_2 + s) c_2 e^{\rho_2 t} = \frac{\Delta T(t - \tau_1) - \frac{\tau_0}{\tau_1}\Delta T(t)}{\tau_0 (e^{-\rho_2 \tau_1} - e^{-\rho_1 \tau_1})} \tag{50}$$

Now we can obtain a recursive equation for $\Delta T(t + \tau_2)$.

$$\Delta T(t + \tau_2) = (\rho_1 + s)\tau_2 c_1 e^{\rho_1 (t + \tau_2)} + (\rho_2 + s)\tau_2 c_2 e^{\rho_2 (t + \tau_2)} \tag{51}$$

$\Delta T(t + \tau_2) =$

$$\frac{\tau_2 (\tau_0 \Delta T(t)(e^{\rho_1 \tau_2 - \rho_2 \tau_1} - e^{\rho_1 \tau_2 - \rho_1 \tau_1} + e^{\rho_2 \tau_1 - \rho_1 \tau_1} - e^{\rho_2 \tau_2 - \rho_1 \tau_1}) + \tau_1 \Delta T(t - \tau_1)(e^{\rho_2 \tau_2} - e^{\rho_2 \tau_1}))}{\tau_1 \tau_0 (e^{-\rho_2 \tau_1} - e^{-\rho_1 \tau_1})} \tag{52}$$

5 Bayesian Estimation of $\mathcal{R}_0$ Probability Distributions

5.1 Current Estimation Techniques

The basic reproductive number is defined as the average number of new infections generated by a typical previous infection in a mostly uninfected population. When speaking in the terms of the HIV virus inside a host, $\mathcal{R}_0$ will be referred to in this manuscript as the average number of new virions created by a typical infected cell in a population of mostly uninfected cells. In other words $\mathcal{R}_0$ would be the reproduction ratio of the virus.

In June of 2006 a new method for estimating $\mathcal{R}_0$ was created by Luis M. A. Bettencourt and Ruy M. Rubeiro in their paper, Detecting early human transmission of H5N1 avian influenza. In this method they use data from WHO global surveillance of new avian flu cases, and interpret it in real time to evaluate changes in transmissibility with measured uncertainty and to perform predictions of new cases. They also predict the probabilistic progression of new cases using a recursive equation from statistical analysis of prior disease progression. More specifically they succeeded in developing a Bayesian procedure that estimates the full probability distribution of $\mathcal{R}_0$.

Using this method we estimated the probability distribution of $\mathcal{R}_0$ for the reproduction of HIV-1 virions in untreated patients at the cellular level. Virus concentration measurements were obtained for 10 different patients [18]. The model we used to describe the dynamics of HIV is from the classic model of viral dynamics [13].

This research will give clinicians another tool for monitoring patients as well as help them tailor-fit disease regiments for individual patients. This method could assist in fine tuning the analysis of parameters given our sensitivity analysis.

5.2 Bayesian Estimation

With the recurrence relation created in the previous section we now have a description for the new cases at time 't+τ_2' defined completely in terms of the previous cases. Additionally, as we had originally intended, $\mathcal{R}_0$ is a part of the expression and is formulated to be the only part of our deterministic expression which varies. Stated differently, variation in $\mathcal{R}_0$ causes corresponding changes in the number of new cases at successive time steps. In this way, as new data is incorporated at each iteration through τ and ΔT the increase and decrease of the expression is due to $\mathcal{R}_0$. This can be stated more concisely by the following conditional probability,

$$P[R_0|\Delta T(t+\tau_2) \leftarrow \Delta T(t), \Delta T(t-\tau_1)] = \frac{P[\Delta T(t+\tau_2) \leftarrow \Delta T(t), \Delta T(t-\tau_1)|R_0]P[R_0]}{P[\Delta T(t+\tau_2) \leftarrow \Delta T(t), \Delta T(t-\tau_1)]} \tag{53}$$

Our deterministic expression for new cases acts in this way as a detector for statistical variation. The key is using this deterministic expression as the mean for the distribution. This makes sense because the average of a stochastic process corresponds to the deterministic model for the same process. Whereas a purely deterministic approach fails to capture the variation in real data the power of this approach lies in its ability to interpret data as it comes in. This approach to $\mathcal{R}_0$ estimation was motivated by trying to marry both approaches and use the advantages of both to interpret and analyze data in real time.

The data from the aforementioned Stafford et. al. paper was used to create estimations of both $\mathcal{R}_0$ and $\mathcal{R}_{eff}$ during Stage 1 HIV-1 infection. The basic replacement number ($\mathcal{R}_{eff}$) is defined as the average number of secondary infected created by an average in-

fectious individual during its lifetime [8, 6]. Their study uses the same model that was presented earlier and nonlinear least squares estimation to attain parameter values for each of its patients. Using these values, $\mathcal{R}_0$ can be calculated for each patient and compared to the estimates achieved by applying Bayesian forward estimation directly to the data. In addition, observed patient data is used from the study to estimate $\mathcal{R}_{eff}$ after the initial growth phase. By definition, $\mathcal{R}_{eff}$ is equal to $\mathcal{R}_0$ at the time of invasion into a completely susceptible population [8]. However, the $\mathcal{R}_{eff}$ is then subsequently always smaller than $\mathcal{R}_0$, because the susceptible population has been reduced to only a fraction of its original size.

Using this formulation, it is possible to calculate $\mathcal{R}_0$ directly from the sample points. In order to estimate $\mathcal{R}_0$, we chose a sample from within the initial growth period as shown in Figures 5 and 6. You can estimate $\mathcal{R}_0$ directly from the sample points. It is remarkably accurate, and does not require individual parameter estimation through curve fitting or other means. With the simulation data, the values approach the calculated value for $\mathcal{R}_0$ rather quickly.

Predicted data was generated for each patient using a numerical solver and the parameter estimates found in another manuscript [18]. A sample was chosen from the initial growth period for each patient as shown in Figures 5 and 6. These graphs show the number of new cases between sample times. The sample size was constrained to about 50 data points for the population between 1 and 10000 virions per nanoliter. Although most of the data and parameters we encountered were given in milliliters, we converted our data to virions per nanoliter due to our computer constraints. Because the numbers were simultaneously very large and very small, we would often hit the limits of computer memory for storage of our numerals. Some of our expressions would then give impossible

values, such as zero. In order to avoid this problem, we scaled our problem to look at change on a nanoscale over a short period of time. One motivation for further work would be to somehow quantify the sensitivity of this methodology to min/max numeral storage.

The observed data was also taken from the study [18]. They obtained their results for patients 1 and 2 from a University of Washington study and the rest from the Aaron Diamond AIDS Research Center. All of the data is from Stage 1 of HIV-1 infection from the peak of the initial growth period to the beginning of the quasi-steady state that follows. This quasi-steady state period can last up to 10 years for many patients and we were able to get data for our estimation procedure well into the second year for many patients. As with the predicted patients, we used the number of new cases between sample times as shown in Figures 7 and 8, although here it is represented in micro-liters. In addition, the data was originally a time series of viral load over time. In order to convert this into a time series for change in new cases we applied the following expression:

$$\Delta T(t) \approx \Delta V(t) + \tau_1 s V(t - \tau_1) \tag{54}$$

which gives us a rough estimation for what occured between the samples. This result was attained by substituting ΔT(t) into the expression for ΔV(t). This also makes some intuitive sense because the difference in viral load alone does not represent the number of new cases unless we add the number of deaths that occured over that same time. Of course, as with any estimation, we cannot be exactly sure how far we are from the true dynamics and this may be another place for further work.

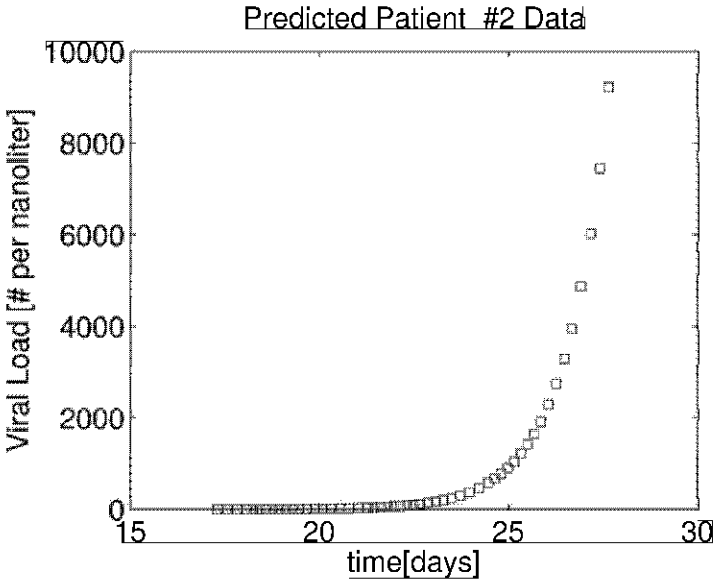

Figure 5: Initial growth sample from numerical solution of Patient 2.

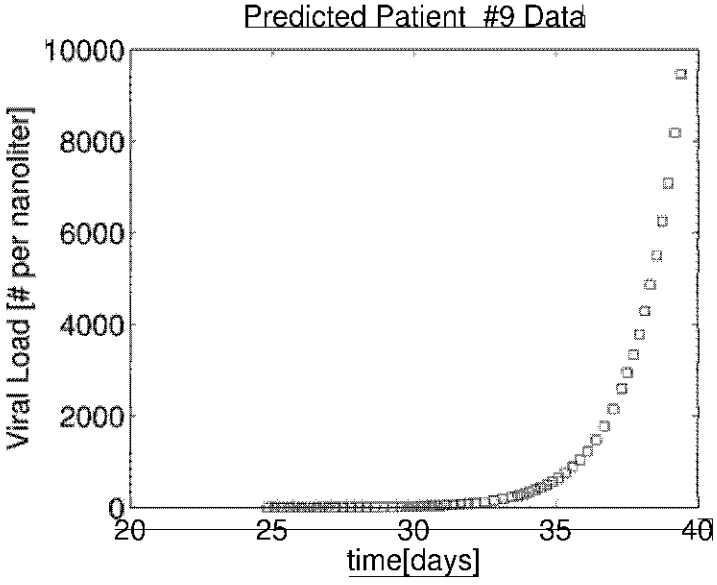

Figure 6: Initial growth sample from numerical solution of Patient 9.

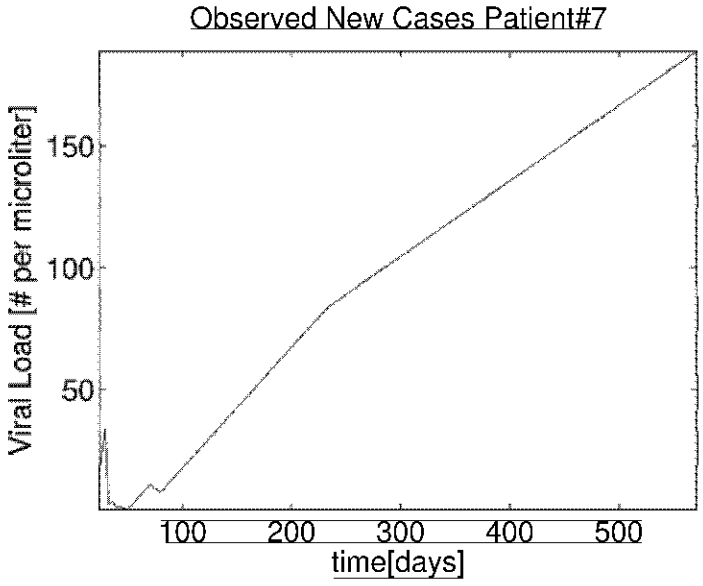

Figure 7: Observed data from Patient 7 Stage 1 HIV.

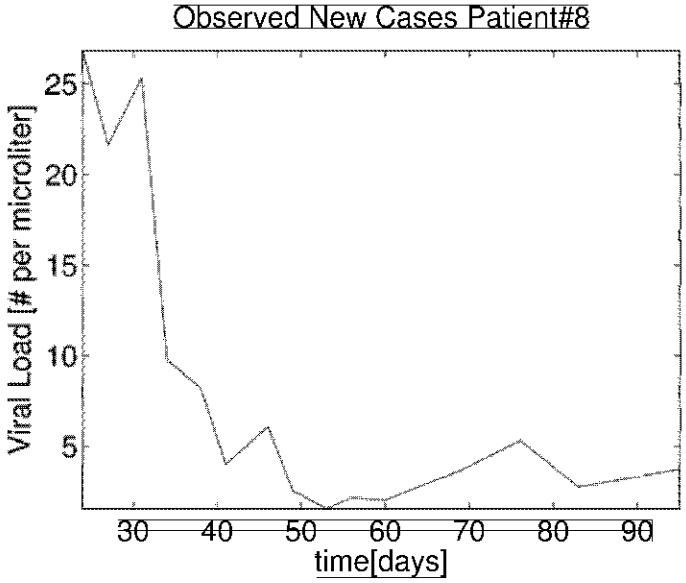

Figure 8: Observed data from Patient 8 Stage 1 HIV.

5.3 Results

In Figure 9 the estimated distributions for the predicted and observed data is represented for Patient 1. The $\mathcal{R}_0$ distribution for the predicted data converged very quickly to a mean of 3.8 +/- 0.144 which is precisely the number calculated by Stafford et. al. for the same patient [18]. Although the predicted data was generated with these parameters, those values were not used explicitly in the expression for the distribution. This value for $\mathcal{R}_0$ was elicited purely from the variation in data. The same $\mathcal{R}_0$ estimate shown on the right as a graph of max values over time shows that the estimate converged to 3.8 as the confidence interval tightened. The observed data also converged to a value of 1. This would imply that the data saturated at an equilibrium, which is confirmed by the viral load time series for this patient.

Patient 2 data looks very similar to Patient 1 with convergence to a value of 2.6+/-.086 compared to a value of 2.8 calculated by Stafford et. al. Here the estimated $\mathcal{R}_0$ converged very smoothly because the initial guess for the distribution was very close to the saturation value. The observed data also converged very quickly to a value of .94+/-3x10^{-6} for the estimate of the effective number ($\mathcal{R}_{eff}$)after the period of initial growth.

Patient 5 data, again, looks very similar to the previous patients in that the data smoothly converges to the expected $\mathcal{R}_0$ value and remains there while continuously decreasing the variance. Here the initial guess for the $\mathcal{R}_0$ distribution was very close to the final estimate so the graphs do not show very much movement. In the estimates for the observed data, the similar pattern of stabilizing near 1 is observed again. Viral load data confirms that this patient has reached an endemic state for this period.

As before, the pattern for all the remaining patients shows a smooth progression from

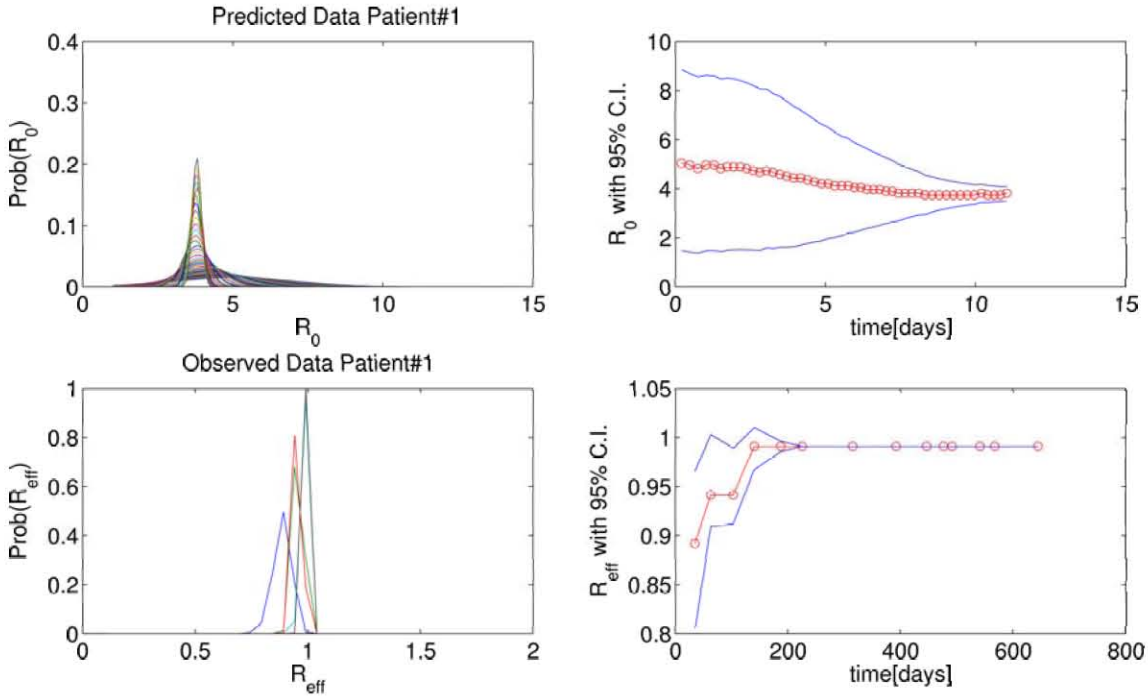

Figure 9: Distribution of $\mathcal{R}_0$ vs time for Patient 1.

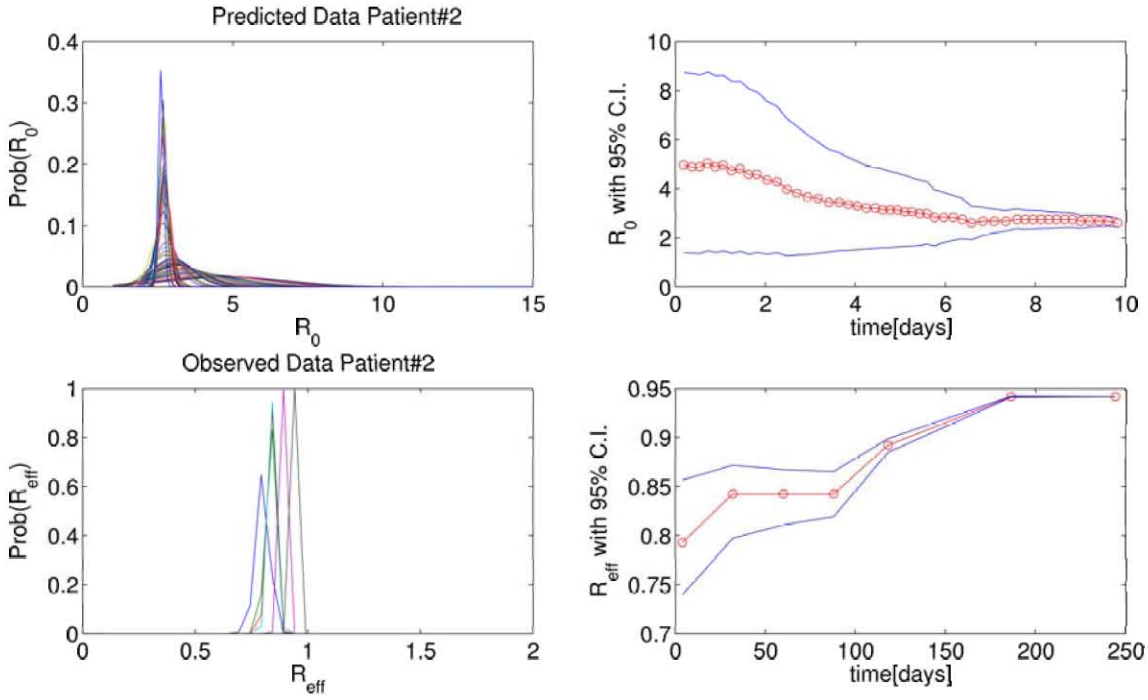

Figure 10: Distribution of $\mathcal{R}_0$ vs time for Patient 2.

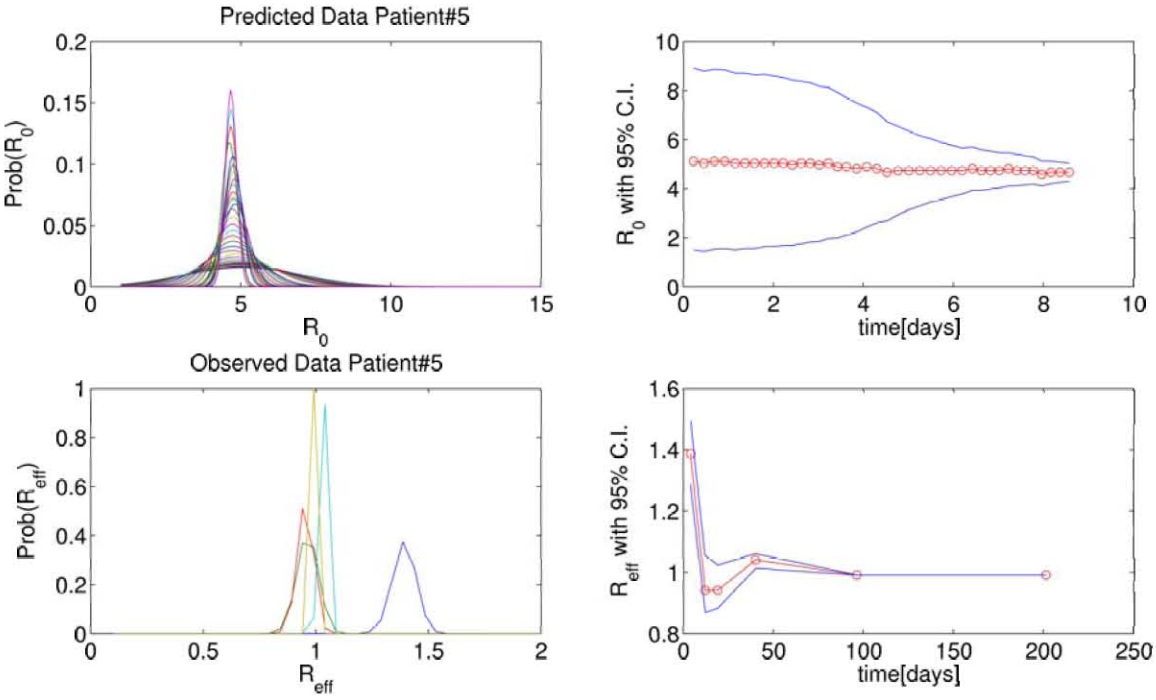

Figure 11: Distribution of $\mathcal{R}_0$ vs time for Patient 5.

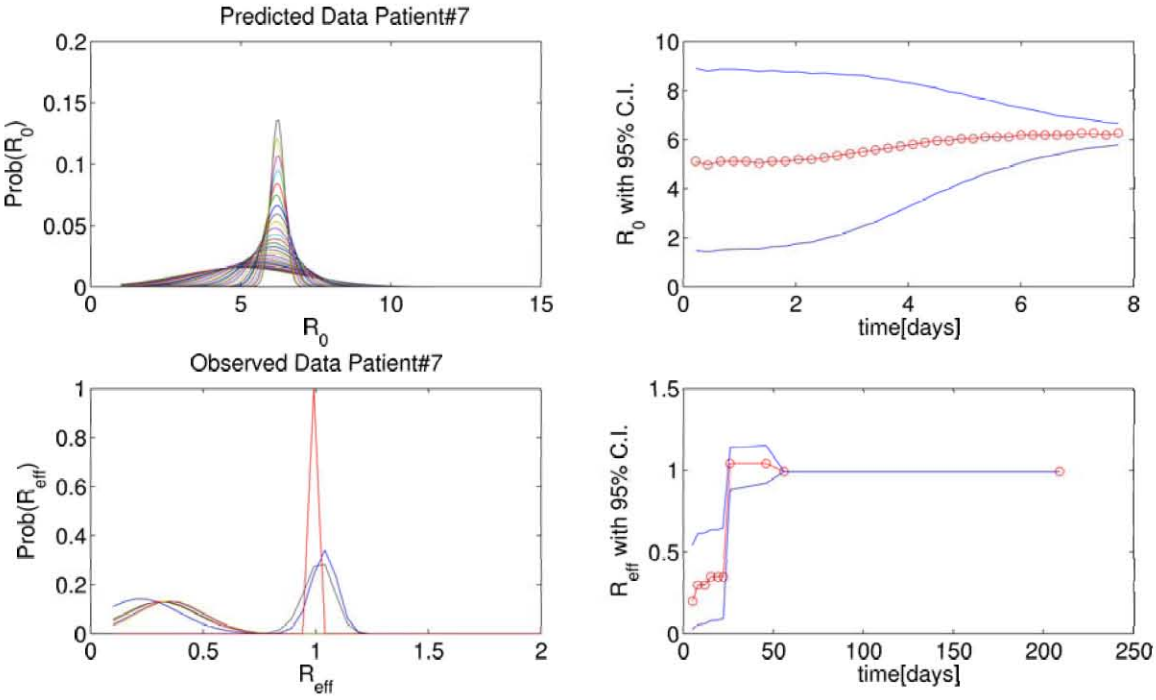

Figure 12: Distribution of $\mathcal{R}_0$ vs time for Patient 7.

Patient	Expected $\mathcal{R}_0$ [18]	Bayes $\mathcal{R}_0$	Bayes $\mathcal{R}_{eff}$
1	3.8	3.805+/-0.144	0.991+/-0.0
2	2.8	2.609+/-0.086	0.941+/-0.001
5	4.7	4.669+/-0.189	0.991+/-0.0
7	6.2	6.231+/-0.221	0.991+/-0.0
8	11.0	10.53+/-0.374	0.651+/-0.114
9	6.6	6.482+/-1.117	0.991+/-0.001

Table 4: Expected $\mathcal{R}_0$ values vs. Bayes Estimations

the initial estimate to the expected $\mathcal{R}_0$ value calculated for the patient. Patient 8 data is particularly representative of the progression of the mean towards the expected value while the variance becomes smaller as shown in Figure 13. The observed data for Patient 8 also had the highest resolution from all selected patients. The estimate for Patient 8 $\mathcal{R}_{eff}$ stays below 1.0 for all estimates because this data comes from the period just after initial growth when the viral load is dipping down to its set point level. After this period of decline the viral load is sometimes so low that it is undetectable. Over years, the immune system slowly deteriorates and the T-cell count gets so low that the immune system eventually collapses into what is called Acquired Immune Deficiency Syndrome or AIDS. The patients selected for our study have positive viral load values during this period as shown in Figure 12. A summary of the estimates given by Stafford et. al. and the values estimated by the Bayes Forward Iteration are summarized in Table 4. All the values estimated by our procedure match those predicted by Stafford et. al. along with quantified uncertainty for each patient which was not provided there. Furthermore, all of the patients showed a higher estimate for $\mathcal{R}_0$ than for their $\mathcal{R}_{eff}$, which agrees with theoretical expectations for these values.

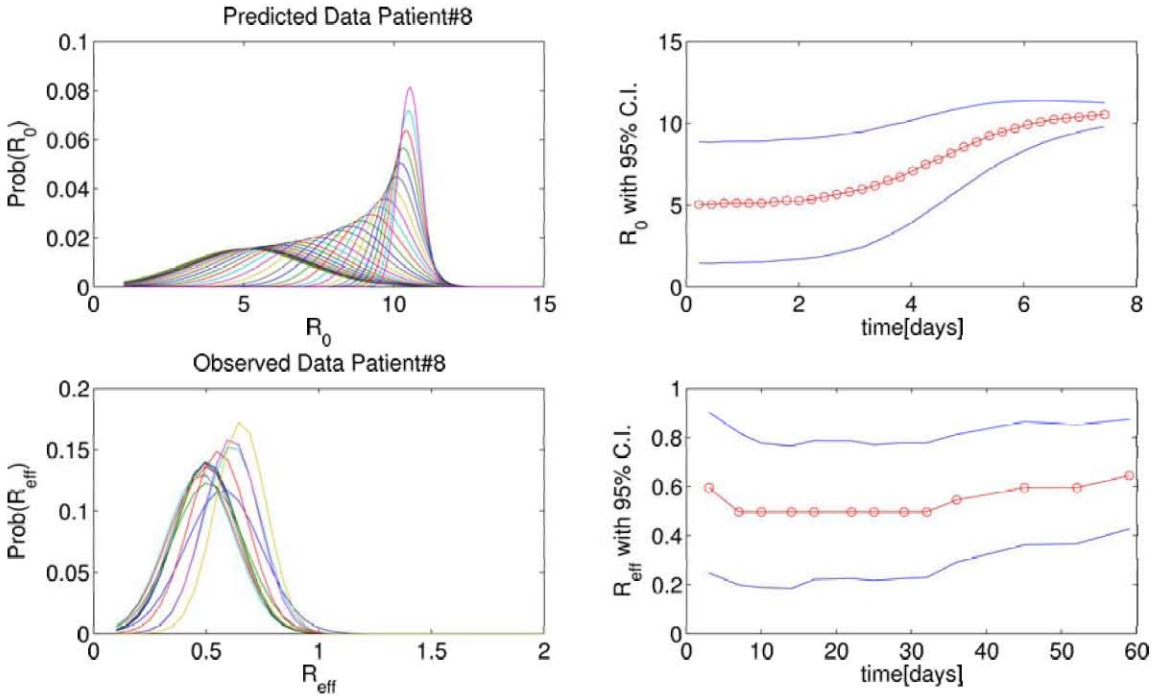

Figure 13: Distribution of $\mathcal{R}_0$ vs time for Patient 8.

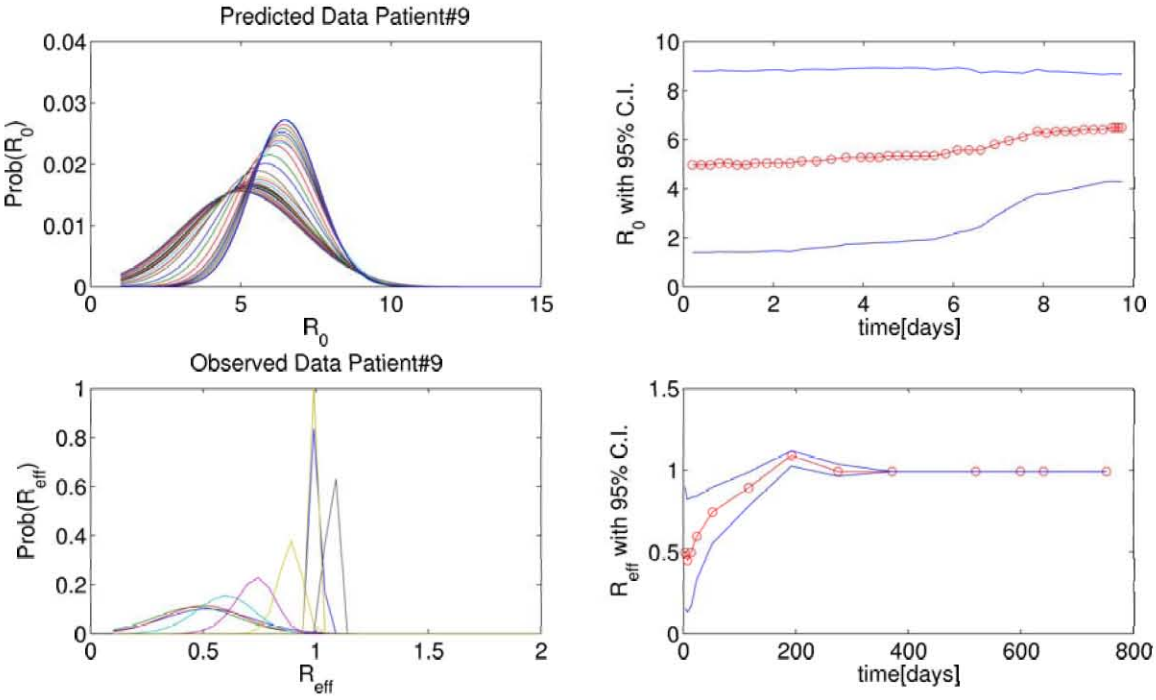

Figure 14: Distribution of $\mathcal{R}_0$ vs time for Patient 9.

6 Uncertainty and Sensitivity Analysis of $\mathcal{R}_0$

In order to determine if our distributions are an accurate representation of HIV-1 viral load within the host, we performed uncertainty and sensitivity analysis. To do so, we decided to create probability distributions for all of the parameters in our model. All of our parameter estimations were taken from a previous paper [18]. From this we know that $\Lambda = \mathcal{X}_{init}\mu$. Due to Λ and μ being related to each other, we chose to use a new variable, $\mathcal{T}_0$, instead as a parameter. Thus:

$$\mathcal{X}_{init} = \frac{\Lambda}{\mu} = \mathcal{T}_0$$

$$\mathcal{R}_0 = \frac{\Lambda\beta k}{\mu a s} = \frac{\mathcal{T}_0\beta k}{as} = \mathcal{T}_0\beta k\frac{1}{a}\frac{1}{s}$$

The average healthy person has an average of 800 to 1200 $CD4^+$ T-cells per micro-liter [9]. The density of target cells before infection are found to be 1% of the $CD4^+$ T-cell density in peripheral blood. This value was found in a paper using nuclear antigen Ki-67 [16, 18]. Our uninfected cell count will range between 8000 and 12000 cells per milliliter. We assumed a uniform distribution for the parameter $\mathcal{T}_0$. For β we chose to use a triangular distribution, with a minimum of .00000019, a peak of .00000065, and a maximum of .0000048. Also, k was assumed to be a triangular distribution. This was created using a minimum of 98, a peak of 850, and a maximum of 7100. An exponential distribution was chosen for both $\frac{1}{a}$ and $\frac{1}{s}$. The mean for $\frac{1}{a}$ was chosen to be 2.56 and the mean for $\frac{1}{s}$ was chosen as .33. All of these distributions can be seen in Figure 16. Using these five parameter distributions we ran our program 10 times, creating 50,000 values of $\mathcal{R}_0$ per simulation. The code used to create these distributions can be viewed in the Appendix.

Parameter Distribution

Parameter	Distribution	Min.	Max	Mean	Peak
$\mathcal{T}_0$	Uniform	8,000	12,000	-	-
β	Triangular	0.00000019	0.0000048	-	0.00000065
k	Triangular	98	7,100	-	850
$\frac{1}{a}$	Exponential	-	-	2.56	-
$\frac{1}{s}$	Exponential	-	-	.33	-

Table 5: Distributions and values used

Uncertainty Analysis

Run	Min. $\mathcal{R}_0$	Max. $\mathcal{R}_0$	Median $\mathcal{R}_0$	Mean $\mathcal{R}_0$	Std. Dev.	$\Pr[\mathcal{R}_0 < 1]$	$\Pr[\mathcal{R}_0 \geq 1]$
1	0.000045	3174.858832	11.201395	42.903161	103.195643	.13756	.86244
2	0.000029	5698.776718	10.935308	42.895685	108.835796	.1366	.8634
3	0.000017	2726.865301	10.957302	42.57981	104.225365	.13418	.86582
4	0.000026	5099.867277	10.855994	42.568557	106.460654	.1383	.8617
5	0.000004	3899.840203	11.070551	42.535354	104.257277	.138	.862
6	0.000012	3867.021795	11.231353	42.763078	105.715445	.1375	.8625
7	0.000034	2949.600973	10.983479	42.636431	103.073843	.1382	.8618
8	0.000005	2375.636386	10.989205	42.879221	103.436334	.13608	.86392
9	0.000055	4273.549069	11.051081	43.208162	110.193039	.1364	.8636
10	0.000008	3209.269798	11.060117	42.688299	105.604199	.13646	.86354
Mean	0.0000235	3727.528635	11.0335785	42.7657758	105.4997595	0.136928	0.863072

Table 6: Results of the Uncertainty Analysis

$\mathcal{R}_0$ Quartiles

First Quartile	Second Quartile	Third Quartile	Fourth Quartile
2.728798	10.957172	38.471825	4273.549069

Table 7: Quartiles of the Probability Distribution of $\mathcal{R}_0$

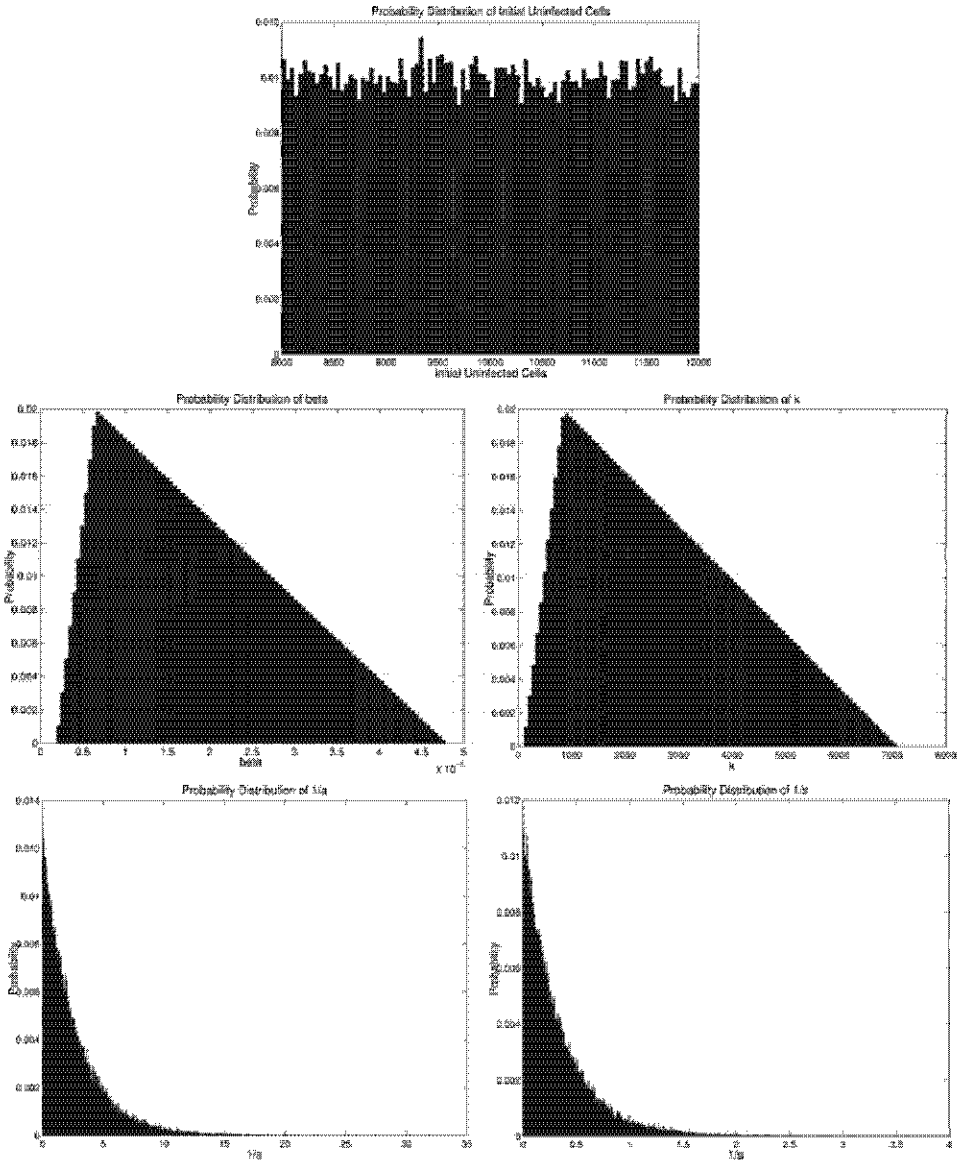

Figure 15: In order from top to bottom and left to right are the Probability Distributions of: $\mathcal{T}_0$, β, k, $\frac{1}{a}$, and $\frac{1}{s}$

.

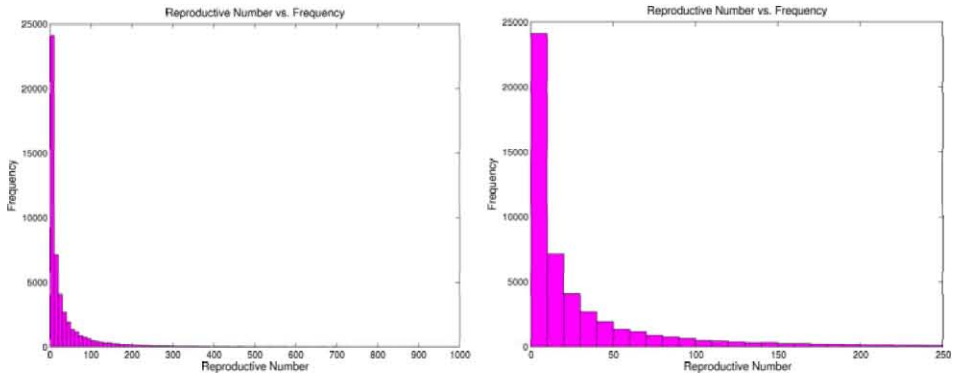

Figure 16: To the left is a histogram of the simulated $\mathcal{R}_0$ distribution for patient 3. To the right is the same histogram truncated and enlarged. Since all 10 of the replications are comparable, any of the replications is a reasonable representation of the other nine.

6.1 Uncertainty Analysis

From Table 6 it can be shown that the range for $\mathcal{R}_0$ ran from .000004 to 5698.776718. This is an extremely large range, but acceptable given the parameter ranges we used. The estimated $\mathcal{R}_0$ values that we were hoping for ranged from about 2 to 20 [18, 11]. Despite the large variation in the range of $\mathcal{R}_0$, the median $\mathcal{R}_0$ was 11.0335785. This number falls directly in our estimated $\mathcal{R}_0$ values. The mean $\mathcal{R}_0$ given was 42.7657758, with a standard deviation of 105.4997595. This shows that our $\mathcal{R}_0$ range is very large, and that our frequency histogram of $\mathcal{R}_0$ is strongly skewed to the right. Given that our mean is 42.7, and our median is 11.033, it can be concluded that there exist very large values of $\mathcal{R}_0$ per simulation that increase the mean tremendously. Due to the large standard deviation values, it is easy to notice that there is a large fluctuation in $\mathcal{R}_0$. Also, since $\mathcal{R}_0$ has such a large range, there is a strong probability that $\mathcal{R}_0$ has the capability to explode into large numbers. When looking at the distribution of $\mathcal{R}_0$ values, we did find that 62.3% of $\mathcal{R}_0$ values were below 20, 70.4% below 30, 75.8% below 40, 79.7% below 50, and 89.3% below

100. Using the distribution from run 9, we calculated the quartiles. Run 9 was chosen since all other runs were comparable to this one. From Table 7 you can see that 75% of the distribution is under 38.471825; while the last quartile spans the range of over 4,000. This is also a strong indication that our distribution is skewed to the right. One surprising observation in our simulations, is that the probability of $\mathcal{R}_0 \geq 1$ was .863072. Given our knowledge of HIV dynamics, we would have expected a larger probability of $\mathcal{R}_0 \geq 1$.

6.2 Sensitivity Analysis

In the sensitivity analysis of our parameters, partial rank correlation coefficients (PRCC) were used. A parameter is influential in determining the magnitude of $\mathcal{R}_0$ if $|PRCC| > .5$. Also, if the $p-value$ associated with a parameter PRCC is less than .05 we say that this correlation is statistically significant [17]. We found a positive correlation for four parameters, namely: $\mathcal{T}_0$, β, k, and a. As you can see from Table 8 below, there was a small correlation (almost 0) between $\mathcal{T}_0$ and $\mathcal{R}_0$ and between a and $\mathcal{R}_0$. The correlation which is most influential is with respect to the parameter s. As s is increased, $\mathcal{R}_0$ is decreased, and vice versa.

The $p-value$ lets us know about the statistical significance of the correlations. Every correlation between $\mathcal{R}_0$ and the five parameters showed that they are statistically significant. The $p-values$ for all of the parameters were found to be 0.000 (i.e.$<.05$). Hence we are rejecting the null hypothesis, and the five parameters are considered as statistically significant. From these $p-values$ we can conclude that there is a correlation between all the parameters and $\mathcal{R}_0$. Thus the correlation between $\mathcal{R}_0$ and the parameters could not have arisen by chance. These $p-values$ show that there is approximately a 0% chance that these correlations occurred at random.

Sensitivity Index of $\mathcal{R}_0$ with respect to Parameters

Parameters	PRCC	$p-value$
$\mathcal{T}_0$	0.054	0.000
β	0.276	0.000
k	0.297	0.000
a	0.175	0.000
s	-0.541	0.000

Table 8: Table of Sensitivity Analysis results

7 Conclusion

We estimated the distributions of the basic reproductive number for all the patients selected in our study. As has been demonstrated by our graphs and previous discussion, our estimation progresses smoothly from the prior to the basic reproductive number while simultaneously decreasing the variance. Although we used predicted values for our parameters in our estimation of $\mathcal{R}_0$ this still serves as an effective control for checking the accuracy of the algorithm and its expressions, which should be employed whenever this algorithm is applied to a different model or context. Unlike, applications to larger time scales, like a flu season, the sample data used for this estimation resided wholly within the initial growth phase. This means that our estimation should stabilize at the basic reproductive number and remain there until the end of the sampled time. This is confirmed by our data, particularly in the graphs of the max $\mathcal{R}_0$ estimates for each patient. These graphs can be easily compared to the estimates given by Stafford et. al. using Table 4.

Bettencourt et. al. suggested a method for calculating the basic reproductive number from a time series of the effective number by taking the maximum value. The theoretical

reason behind this result pertains to the distinction between the effective reproductive number, or replacement number, and the basic reproductive number. Under normal conditions the effective reproductive number never gets any higher than in the initial growth phase because the population is completely susceptible in the initial growth phase. Due to different assumptions about the data and its context, such as different stages of growth, this method could not be directly applied to our samples. However, if our model and methodology, as well as computing constraints, were improved sufficiently this estimation procedure could encompass a larger time period which may allow more longer term analysis such as this to be employed.

The effective reproductive number calculated for each patient after the initial growth phase agreed with expectations about its size relative to $\mathcal{R}_0$. More so, estimated $\mathcal{R}_{eff}$ values seemed to support intuitive expectations from the slope of the viral load curve during those periods. Because the observed data was low resolution and highly dynamic it was less than ideal for this estimation procedure given the assumptions we made, such as constant uninfected cell population and poisson distributed new cases. For future work, these assumptions would have to be reevaluated and compared to more relevant alternatives, in which more care would be taken choosing data and applying the method in context.

Using the probability density functions, the uncertainty analysis showed that the simulated median for $\mathcal{R}_0$ was 11.0335785. The mean $\mathcal{R}_0$ given was 42.7657758, with a standard deviation of 105.4997595. Due to the large standard deviation values, it is concluded that there is a large fluctuation in the $\mathcal{R}_0$ distribution. Since there exist $\mathcal{R}_0$ values exceeding 5000, this indicates that $\mathcal{R}_0$ has exponential growth. Also, since $\mathcal{R}_0$ has such a large range, there is a strong probability that $\mathcal{R}_0$ has the capability to explode into large numbers.

Also, it was shown through the sensitivity analysis that s was the only parameter that was statistically significant. This means that there is a large negative correlation between $\mathcal{R}_0$ and s. This suggests that viral death rate is a crucial factor in varying the threshold $\mathcal{R}_0$. Therefore, in order for $\mathcal{R}_0$ to be decreased, s should be increased.

7.1 Improvements and Future Work

Some ideas for future work include finding a way to solve the whole system. This way we could assign a solution to x(uninfected cells) that is a function of t so that our deterministic process will be closer to the data. One thing we would like to do is improve our estimation of dynamics between sample points. Another idea is to choose a different distribution with a larger variance to minimize statistical anomalies. On thought is to possibly use truncated normal distributions. Another thought is to try to apply the method on a more accurate model for HIV, one that is more complex. Our hope is that this can one day be extended to real time analysis of medicated patients.

References

[1] *AIDS Epidemic Update; UNAIDS and WHO.* http://www.who.int/hiv/epiupdate2005_en.pdf, December, 2005.

[2] Aids.Org, *Viral Load Tests.* San Francisco AIDS Foundation, http://www.aids.org/Factsheets/125-Viral-Load-Tests.html, July, 2006.

[3] *Antiviral Immunity.* Microbiology @ Leicester, http://www-micro.msb.le.ac.uk/MBchB/8a.html, July,2006.

[4] Avert.Org, *The Different Stages of HIV Infection.* http://www.avert.org/hivstages.htm, June, 2006.

[5] Bettencourt, Luis M.A. and Ruy M. Ribeiro, *Detecting early human transmission of H5N1 avian influenza.* June, 2006.

[6] Del Valle, S. et. al., *Effects of behavioral changes in a smallpox attack model.* Mathematical Biosciences, **195**, 228-251, 2005.

[7] EhealthMD, *How is HIV Infection Diagnosed?* http://www.ehealthmd.com/library/aids/ADS_diagnosis.html, January, 2004.

[8] Hethcote, Herbert W., *The Mathematics of Infectious Diseases.* SIAM Review, Vol. 42, No. 4, pp. 599653, 2000.

[9] *How HIV Causes Aids.* National Institute of Allergy and Infectious Diseases, http://www.mtholyoke.edu/dept/biol/documents/How_HIV_Causes_Aids.pdf, October, 2001.

[10] *Immune Response: Overview.* University of Arizona, http://www.biology.arizona.edu/immunology/tutorials/AIDS/HIVimmune.html, July, 2006.

[11] Little, Susan J. et. al., *Viral Dynamics of Acute HIV-1 Infection.* The Journal of Experimental Medicine, 841-850, **190**, 6, 1990.

[12] McMichael, Andrew, *How HIV Fools the Immune System.* Institute of Molecular Medicine, Oxford University, http://www.tulane.edu/∼dmsander/WWW/335/McMichael/McMichael.html, July, 2006.

[13] Nowak, Martin A. and Robert M. May, *Virus Dynamics: Mathematical Principles of Immunology and Virology.* Oxford University Press, 2000.

[14] Perelson, Alan S. et. al., *HIV-1 Dynamics in Vivo: Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time.* Science, 1582-1586, **271**, December, 2005.

[15] Pieribone, David, *The HIV Lifecycle.* Aids Community Research Initiative of America, http://www.thebody.com/cria/winter03/hiv_lifecycle.html, Winter, 2003.

[16] Sachsenberg, Nicolas et. al., *Turnover of $CD4^+$ and $CD8^+$ T Lymphocytes in HIV-1 Infection as Measured by Ki-67 Antigen.* The Journal of Experimental Medicine, 1295-1303, **187**, 8, 1998.

[17] Sanchez, M and Sally Blower, *Uncertainty and Sensitivity Analysis of the Basic Reproductive Rate.* American Journal of Epidemiology, 1127–1137, **145**, 1997.

[18] Stafford, Max A. et. al., *Modeling Plasma Virus Concentration during Primary HIV Infection.* Journal of Theoretical Biology, 285-301, **203**, 2000.

[19] The Henry J. Kaiser Family Foundation, *The Global HIV/AIDS Timeline.* http://www.kff.org/hivaids/timeline/index.cfm, July, 2006.

[20] *Viral Load Testing.* San Francisco AIDS Foundation, http://www.sfaf.org/aids101/viral_testing.html, December, 1998.

[21] WebPath, *WebPath: The Internet Pathology Laboratory for Medical Education, Florida State University, College of Medicine.* http://www-medlib.med.utah.edu/WebPath/TUTORIAL/AIDS/HIV.html, July, 2006.

8 Acknowledgments

We would like to express our gratitude to the numerous people who offered their time and effort in contributing to this project. Foremost, Dr. Carlos Castillo-Chavez, who gave us the opportunity to participate in MTBI 2006. We would also like to commend our

two advisors: Dr. Xiaohong Wang and Dr. Ariel Cintron-Arias for their support and patience. We would also like to recognize Dr. Linda Gao, Anuj Mubayi, and Mustafa Erdem for their contributed thoughts and ideas. We would especially like to thank Alan S. Perelson for his help and guidance.

The Mathematical and Theoretical Biology Institute would like to thank the following grant sources for the Summer 2006 REU program: The National Science Foundation (NSF), award number DMS-0502349. The Alfred P. Sloan Foundation. The National Security Agency(NSA), award number H98230-06-1-0097 and Arizona State University.

9 Appendix

9.1 MATLAB code for the HIV Model

```
function dx=xyv_para(t,x,lambda,beta,mu,a,k,s)


%patient #4


% lambda = 46;          % cells/((mL)(day))
% beta   = 4.8*10^-6;   % mL/((virion)(day))
% mu     = .0046;       % 1/day
% a      = .18;         % 1/day
% k      = 98;          % 1/day
% s      = 3;           % 1/day



dx=zeros(5,1); dx(1)= lambda - mu*x(1) - beta*x(1).*x(3); dx(2)=
beta*x(1).*x(3) - a*x(2); dx(3)= k*x(2) - s*x(3);
dx(4)= k*x(2);  %Total new cases
dx(5)= s*x(3);  %Total virion deaths
```

9.2 MATLAB code used to create Predicted Data

```
function patient_sim(index_vector) set(0,'DefaultAxesFontSize',16)


initX  = 10000;      % cells/microL
initY  = 0;          % cells/microL
initV  = 1*10^-6;    % virions/mL
```

```
init_conds = [initX,initY,initV,0,0];

num_patients = 10; patient = zeros(1,num_patients);

%initialize parameter arrays
lambda = zeros(1,num_patients); beta = zeros(1,num_patients); mu =
zeros(1,num_patients); a = zeros(1,num_patients); k =
zeros(1,num_patients); s = zeros(1,num_patients);

% tspan = [0:3:99];
tspan = [0 49]; %number of samples

%patient #1
patient(1)=1;
lambda(1) = 1.3*10^2;  % cells/((mL)(day))
beta(1)   = 4.6*10^-7; % mL/((virion)(day))
mu(1)     = 1.3*10^-2; % 1/day
a(1)      = .4;        % 1/day
k(1)      = 980;       % 1/day
s(1)      = 3;         % 1/day

%patient #2
patient(2)=2;
lambda(2) = 2.0*10^2;  % cells/((mL)(day))
beta(2)   = 3.6*10^-7; % mL/((virion)(day))
mu(2)     = 2.0*10^-2; % 1/day
```

```
a(2)      = .8;         % 1/day
k(2)      = 1800;       % 1/day
s(2)      = 3;          % 1/day

%patient #3
patient(3)=3;
lambda(3) = 65;         % cells/((mL)(day))
beta(3)   = 6.4*10^-7; % mL/((virion)(day))
mu(3)     = 6.5*10^-3; % 1/day
a(3)      = .43;        % 1/day
k(3)      = 960;        % 1/day
s(3)      = 3;          % 1/day


%%%%%%%
%patient #4
patient(4)=4;
lambda(4) = 46;         % cells/((mL)(day))
beta(4)   = 4.8*10^-6; % mL/((virion)(day))
mu(4)     = .0046;      % 1/day
a(4)      = .18;        % 1/day
k(4)      = 98;         % 1/day
s(4)      = 3;          % 1/day
%%%%%%%


%%%%%%%
```

```
%patient #5
patient(5)=5;
lambda(5) = 170;         % cells/((mL)(day))
beta(5)   = 6.3*10^-7; % mL/((virion)(day))
mu(5)     = .017;      % 1/day
a(5)      = .39;        % 1/day
k(5)      = 870;         % 1/day
s(5)      = 3;          % 1/day
%
%%%%%%%
%patient #6
patient(6)=6;
lambda(6) = 1.2*10^-4;         % cells/((mL)(day))
beta(6)   = 7.5*10^-7; % mL/((virion)(day))
mu(6)     = .012;      % 1/day
a(6)      = .39;        % 1/day
k(6)      = 790;         % 1/day
s(6)      = 3;          % 1/day
%
%%%%%%%
%patient #7
patient(7)=7;
lambda(7) = 170;         % cells/((mL)(day))
beta(7)   = 8.0*10^-7; % mL/((virion)(day))
mu(7)     = .017;      % 1/day
a(7)      = .31;        % 1/day
```

```
k(7)      = 730;         % 1/day
s(7)      = 3;          % 1/day
%
%%%%%%%
%patient #8
patient(8)=8;
lambda(8) = 85;         % cells/((mL)(day))
beta(8)   = 6.6*10^-7; % mL/((virion)(day))
mu(8)     = .0085;      % 1/day
a(8)      = .17;        % 1/day
k(8)      = 830;         % 1/day
s(8)      = 3;          % 1/day
%
%%%%%%%
%patient #9
patient(9)=9;
lambda(9) = 60;         % cells/((mL)(day))
beta(9)   = 2.5*10^-6; % mL/((virion)(day))
mu(9)     = .006;      % 1/day
a(9)      = .13;        % 1/day
k(9)      = 110;         % 1/day
s(9)      = 3;          % 1/day
%
%%%%%%%
%patient #10
patient(10)=10;
```

```
lambda(10) = 43;          % cells/((mL)(day))
beta(10)   = 1.9*10^-7; % mL/((virion)(day))
mu(10)     = .0043;       % 1/day
a(10)      = .46;         % 1/day
k(10)      = 7100;          % 1/day
s(10)      = 3;           % 1/day


%solve system with given parameters
%index_vector;
for i=1:length(index_vector)
    j = index_vector(i);
    t = []; y = []; indx = []; data = [];%should null x and y
    figure;
    [t,y]=ode45(@xyv_para,tspan,init_conds,[],lambda(j),beta(j),mu(j),a(j),k(j),s(j));
%     semilogy(t,y(:,3),'s');
%     plot(t,y(:,3),'s');
    graph_title = strcat('Predicted New Cases Patient  #',int2str(patient(j)));
    title(graph_title);

    %data storage
    indx = find((1<y(:,3))&(y(:,3)<1e4));
%     indx = 1:length(y(:,3));

%     fprintf('(%i) %f\n',indx, y(indx,3));
```

```
    data(:,1) = diff(t(indx));
    data(:,2) = diff(y(indx,3));
    %fprintf('%1.15e \n', y(1:(end-1),3) );
    %fprintf('%1.15e \n', s(i)*y(1:(end-1),3) );
    tmp=s(j)*y(indx,3);
    data(:,3) = data(:,2) + data(:,1).*tmp(1:end-1);

    plot(t(indx),y(indx,3),'s');
    xlabel('time[days]')
    ylabel('Viral Load [# per nanoliter]');
    title(graph_title);

    this_f=strcat('data\simdata_patient',int2str(patient(j)),'.txt');
    mt=[data(:,1) data(:,3)];
    save(this_f,'mt','-ascii');

end
```

9.3 MATLAB code used to Graph Observed Data

```
function graph_obs(index_vector)
set(0,'DefaultAxesFontSize',16);

for i=1:length(index_vector)
    j = index_vector(i);
    f_name = strcat('data\obsdata_patient',int2str(j),'.txt');
    x = load(f_name, '-ascii');
```

```
    figure;
    plot(cumsum(x(:,1))+20, x(:,2));
    graph_title = strcat('Observed New Cases Patient# ',int2str(j));
    title(graph_title);
    xlabel('time[days]');
    ylabel('Viral Load [# per microliter]');
    axis tight;
end
```

9.4 MATLAB code to perform Bayes Forward Estimation

```
function dy=bayes_frwd_estmtn(simdata, p_num)

a = zeros(1,10); s = zeros(1,10);
a(1)      = .4;       % 1/day
a(2)      = .8;       % 1/day
a(3)      = .43;       % 1/day
a(4)      = .18;       % 1/day
a(5)      = .39;       % 1/day
a(6)      = .39;       % 1/day
a(7)      = .31;       % 1/day
a(8)      = .17;       % 1/day
a(9)      = .13;       % 1/day
a(10)      = .46;       % 1/day
s(:) = 3.0;    %per day

%test for loading simulated data or observed data
```

```
if(simdata)
    this_file=strcat('data\simdata_patient',int2str(p_num),'.txt');
    R0=linspace(1,20,250);
    R0_dist=normpdf(R0,5,2)';
    R0_dist=R0_dist./sum(R0_dist);
else
    this_file=strcat('data\obsdata_patient',int2str(p_num),'.txt');
    R0=linspace(1e-1,5,100);%change this per patient
    R0_dist=normpdf(R0,.5,.2)';
    R0_dist=R0_dist./sum(R0_dist);
end

x=load(this_file,'-ascii');

tau= x(:,1); %times
DeltaT= round(x(:,2)); %number of new cases at each time

data_length = length(x); data_cursor = 2; iterations =
data_length-data_cursor;

% R0_dist = normpdf(R0, 2, 12);
% R0_dist = R0_dist / sum(R0_dist);
%R0_dist=R0.*exp(-R0); %first prior: positive because of condition

%R0_dist=ones(length(R0),1);
% R0_dist=normpdf(R0,5,2)';
```

```
% R0_dist=R0_dist./sum(R0_dist);

lambda=zeros(iterations,length(R0_dist));
vals=zeros(length(R0_dist),iterations);
maxvals=zeros(iterations,1); cond_prob=zeros(1,length(R0));

for i=data_cursor:data_length-1
    rho_1 = ((-1*a(p_num)-s(p_num)) + sqrt((a(p_num)+s(p_num))^2
    -4*a(p_num)*s(p_num)*(1-R0)))/2; %needs to be positive for stability condition
    rho_2 = ((-1*a(p_num)-s(p_num)) - sqrt((a(p_num)+s(p_num))^2
    -4*a(p_num)*s(p_num)*(1-R0)))/2;

    A = (DeltaT(i)./tau(i))-(tau(i).*DeltaT(i-1)-tau(i-1).*exp(-1*rho_1.*tau(i)).
    *DeltaT(i)).*(tau(i).*tau(i-1).*(exp(-1.*rho_2.*tau(i))-exp(-1.*rho_1.*tau(i))));
    B = DeltaT(i)/tau(i) - A;

    lambda(i,:)= tau(i+1).*(A.*exp(rho_1.*tau(i+1))+B.*exp(rho_2.*tau(i+1)));

    if sum(lambda(i,:))~=0
        cond_prob = poisspdf(DeltaT(i+1),lambda(i,:)');
        if(cond_prob == 0)
            fprintf('ANOMALY!!! (culprit: new cases fall outside the distribution)\n');
        end
        q=cond_prob.*R0_dist;
        q=q/sum(q);
        R0_dist=q;
```

```
        vals(:,i-(data_cursor-1))=q;
    else
        fprintf('ANOMALY!!!  (culprit: lambda)\n');
    end
    hold on
    plot(R0,q)
    graph_title = strcat('Patient  #',int2str(p_num));
    title(graph_title);
    pause(.75)

end
% hold off

%save data to MATLAB .mat file
if(simdata)
    name1=strcat('data\r0_dstn-sim_patient',int2str(p_num),'.mat');
else
    name1=strcat('data\r0_dstn-obs_patient',int2str(p_num),'.mat');
end save(name1,'vals','R0','tau');
```

9.5 MATLAB code to graph and display patient data and confidence intervals

```
function run_datagraph(index_vector)
set(0,'DefaultAxesFontSize',16) sim_data = false; for
i=1:length(index_vector)
    j = index_vector(i);
```

```
subplot_indx = 1;
figure;
for w=1:2  %loop for patient data and sim data
    R0=[]; vals = []; tau = []; q = [];%null
    sim_data = ~sim_data;
    if(sim_data)
        name = strcat('data\r0_dstn-sim_patient',int2str(j),'.mat');
    else
        name = strcat('data\r0_dstn-obs_patient',int2str(j),'.mat');
    end
    %load variables from file
    load(name);
    R0 = R0';

    %%%%%(1)
    %plot distribution
    subplot(2,2,subplot_indx);
    subplot_indx = subplot_indx+1;
    plot(R0,vals);

    if(sim_data)
        xlabel('R_{0}');
        ylabel('Prob(R_{0})');
        graph_title=strcat('         Predicted Data Patient#',int2str(j));
        v = axis; %get axis data
        axis([0 15 v(3) v(4)]);
```

```
else
    xlabel('R_{eff}');
    ylabel('Prob(R_{eff})');
    graph_title=strcat('Observed Data Patient#',int2str(j));
    v = axis; %get axis data
    axis([0 2 v(3) v(4)]);
end
title(graph_title);
%%%%%%%%%%%%%

%%%%%(2)
width_vals = length(vals(1,:));
length_vals = length(vals(:,2));
%i could store the first prior in vals so i don't have to index tau from 2
graph_data = zeros(width_vals, 3);
timeline = cumsum(tau(2:(end-1)));
[blah, indx_m] = max(vals);
graph_data(:,2) = R0(indx_m);

for i=1:width_vals
q = vals(:,i);
av = sum(R0.*q);
st  = sqrt(sum((R0-av).^2.*q));
graph_data(i,1) = av - (1.96)*st;
graph_data(i,3) = av + (1.96)*st;
end
```

```
        subplot(2,2,subplot_indx);
        subplot_indx = subplot_indx+1;
        plot(timeline, graph_data(:,2), 'ro-', timeline,
        graph_data(:,1), 'b-', timeline, graph_data(:,3), 'b-');
        %axes tight;
        if(sim_data)
            ylabel('R_{0} with 95% C.I.');
            final_str = strcat('Patient#',int2str(j),' final estimate for R_{0} = ');
        else
            ylabel('R_{eff} with 95% C.I.');
            final_str = strcat('Patient#',int2str(j),' final estimate for R_{eff} = ');
        end
        xlabel('time[days]');
        fprintf('%s %f +/- %f\n',final_str, av, st);

        %print out final estimation

    end
end
```

9.6 MATLAB code for Estimated Parameter Distributions and Probability Density Functions

```
function Sampling;
%% Choose distributions for each parameter in the R_0 expression
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
NUM = 50000

T0=unifrnd(8000,12000,NUM,1);
%creates a uniform distribution between 8000 and 12000

figure; [T0_freq,bins] = hist(T0,100);
bar(bins,T0_freq/length(T0)); xlabel('Initial Uninfected Cells');
ylabel('Probability'); title('Probability Distribution of Initial
Uninfected Cells');

beta=triangular(.00000019,.00000065,.0000048,NUM);
%creates a triangular distribution with min=1.9e-7, peak=6.5e-7, and max=4.8e-6

beta=beta'; figure; [beta_freq,bins] = hist(beta,100);
bar(bins,beta_freq/length(beta)); xlabel('beta');
ylabel('Probability'); title('Probability Distribution of beta');

k=triangular(98,850,7100,NUM);
%creates a triangular distribution with min=98, peak=850, and max=7100

k=k'; figure; [k_freq,bins] = hist(k,100);
bar(bins,k_freq/length(k)); xlabel('k'); ylabel('Probability');
title('Probability Distribution of k');

a=exprnd(2.56,NUM,1);
%creates an exponential distribution with mean=2.56
```

```
a=1./a; figure; [a_freq,bins] = hist(1./a,1000);
bar(bins,a_freq/length(a)); xlabel('1/a'); ylabel('Probability');
title('Probability Distribution of 1/a');

s=exprnd(.33,NUM,1);
%creates an exponential distribution with mean=.33
s=1./s; figure; [s_freq,bins] = hist(1./s,1000);
bar(bins,s_freq/length(s)); xlabel('1/s'); ylabel('Probability');
title('Probability Distribution of 1/s');

R0=zeros(NUM,1);

%% Calculate R_0 using sets of the parameters generated above
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

for i=1:NUM
    R0(i,1)=(beta(i)*k(i)*T0(i))/(a(i)*s(i));
end result=[R0, T0, beta, k, a, s]; save -ASCII
sampling_result_1.txt; x_axis=(5:10:2005); figure;
hist(R0,x_axis); h=findobj(gca,'Type', 'patch'); set(h,
'FaceColor','m','EdgeColor','k')
fprintf('data are on the following line: \n %f %f %f %f %f %f %f \n',
min(R0),max(R0),median(R0),mean(R0),std(R0),sum(R0<1)/NUM,sum(R0>=1)/NUM);

quartile = floor(NUM/4); sorted_R0 = sort(R0);
```

```
fprintf('First quartile(25 percent): %f\n', sorted_R0(quartile));
fprintf('Second quartile(50 percent): %f\n', sorted_R0(2*quartile));
fprintf('Third quartile(75 percent): %f\n', sorted_R0(3*quartile));

%This function generates random numbers from a triangular distribution
%a=min val, b=top value and c=max value
function  t= triangular(a,b,c,N);

x=zeros(1,N+1); m=zeros(1,N); t=zeros(1,N); s=zeros(1,N);

p=(c-a)*(b-a); %area of min value
q=(b-a)/(c-a); %area of top value
r=(c-a)*(c-b); %area of max value

x(1)=a; %initializing min value

for i=2:N+1
    y=(i-1)/N;
    if y<q
        x(i)=sqrt(y*p)+a;
    else
        x(i)=c-sqrt((1-y)*r);
    end
end

for i=1:N
```

```
    u=unifrnd(0,1);
    m(i)=x(i)+u*(x(i+1)-x(i));
end

%Since the random numbers generated above are generated
%by default in increasing order we need to permute so that they are
%generated randomly in the given interval
s=randperm(N);

for i=1:N t(i)=m(s(i)); end
```