Markov process models of the dynamics of HIV reservoirs

Jane M. Hawkins*

Department of Mathematics, University of North Carolina at Chapel Hill, CB #3250 Chapel Hill, North Carolina 27599-3250, United States

ARTICLE INFO

ABSTRACT

While latently infected CD4+ T cells are extremely sparse, they are a reality that prevents HIV from being cured, and their dynamics are largely unknown. We begin with a two-state Markov process that models the outcomes of regular but infrequent blood tests for latently infected cells in an HIV positive patient under drug therapy. We then model the hidden dynamics of a latently infected CD4+ T cell in an HIV positive patient and show there is a limiting distribution, which indicates in which compartments the HIV typically can be found. Our model shows that the limiting distribution of latently infected cells reveals the presence of latency in every compartment with positive probability, supported by clinical data. We also show that the hidden Markov model determines the outcome of blood tests and analyze its connection to the blood test model.

1. Introduction

Up to now, eradicating the human immunodeficiency virus (HIV-1, which we denote by HIV) in an individual who has become infected has remained out of reach of current medical practice. However the advent of antiretroviral therapy (ART) changed HIV from a direct path to AIDS, a fatal disease, to a chronic disease [22,27,28]. As early as 1996 studies showed that even effective ART that brings the level of HIV below the level of detection in the blood does not successfully eliminate the virions from other cells or biological compartments in the body [7,13,21,26].

There remains a fundamental problem of understanding how and why the reservoirs are formed and maintained, and why the current drug therapy does not completely eradicate them. When ART is stopped, in all patients (with one possible exception known as the Berlin patient), the virus reappears in the blood after a period of time that varies from individual to individual [14], often as soon as a few weeks later [4]. In the case of the Berlin patient, he had developed acute myelogeneous leukemia, and was treated with a stem cell transplant from a donor harboring a mutated form of the CCR5 receptor that provides resistance to HIV infection. The patient's continued absence of detectable virions has led to exploration of a cure in this direction [32]. Another notable case is referred to as the "Mississippi baby" where HIV infection was discovered and treated within 2 days of birth, and ART

was administered for about 18 months. After stopping the drug, the virus remained undetectable for 27 months, but viral rebound occurred [19].

We approach this topic by presenting two models of the location of latently infected cells, referred to as HIV reservoirs, in an individual undergoing ART, as Markov processes. These models utilize a tool in dynamical systems that is advantageous for analyzing spatiotemporal phenomena in extremely complex systems, even those systems whose dynamics are far from random, as is the case here. Instead of following each individual virus particle or infected cell, since there can be millions of virions detected per milliliter of blood in a newly infected or untreated patient, and since it has been shown to be impossible to track and destroy a single virus, or more importantly, every virus particle, we study qualitative and probabilistic aspects of the dynamics of HIV reservoirs. Moreover since typically a blood sample is used to determine the virus levels of a patient, the actual location of source of the virus remains hidden. Therefore we are in fact dealing with a hidden Markov model; that is, there is the output that is viewed (a blood test showing the presence or absence of virus) and a hidden process that is governing the output. Our models incorporate this multilevel structure.

There are several prevailing theories about why HIV has remained incurable so far; we mention a few here. First, the host cells for the virions are CD4+ T helper cells, white blood cells essential to a functioning immune system. Their main role is to signal other T cells to destroy pathogens. Once a CD4+ T cell becomes actively infected by HIV, the immune cell usually is destroyed; it can also happen that the HIV enters the cell and produces viral DNA but does not complete the replication process [6].

* Tel.: +19199629611.
E-mail address: jmh@math.unc.edu, jhawkins@email.unc.edu

While the drug therapy appears to inhibit the replication of virtually all active susceptible virions in the blood, there seem to be virions hiding outside the path of the drugs [10]. In one recent review on latency of the virus under drug therapy [6], the authors offer an opinion that the drugs are not able to go effectively to all sites where the virions can be found:

> Recent studies have indicated that anti-retroviral drug- penetration is site- and compound-specific, and drugs that penetrate poorly may allow viral replication at that site even when plasma viral load is below 50 copies ml^{-1}.

Such locations are called sanctuary sites but are not the subject of this study.

Another theory as to why cures (and vaccinations) remain inaccessible is that the virus is known to mutate extremely quickly. Autopsies of patients with HIV show numerous variations (species) of the virus exist in a body, even within a single organ [2]. This aspect of HIV dynamics is treated mathematically using cellular automata models first introduced in [34], and developed rigorously in [5]. In a follow-up paper by Hawkins and Molinek [18] the authors show that the model has limiting values of healthy CD4+ T cells that can occur and the viral particles are never completely eliminated without "perfect drugs" that are 100% effective in all compartments of the body, coupled with an immune response that can eradicate all the HIV. Up to now this has not been achieved.

In addition to the possibility of active replicating viruses perhaps even under drug therapy, albeit at a low level, it is now generally accepted that there are reservoirs of latent viruses that can persist in the system of an HIV infected patient for years [6,7,14], which form the focus of this paper. A working definition is that a latently infected cell is a cell that does not produce infectious viral particles but is able to do so at some future time, behaving like a Trojan horse. The cell appears to be a healthy cell to the immune system, the drugs do not affect the hidden virions, so it can spring to action at any moment with the viruses replicating in the host cell and producing millions of new virions. The passive viral material (provirus) is passed to cells during normal cellular replication and can reactivate without warning. Therefore up to now, killing the cell containing the provirus seems to be one of the only ways to get rid of the provirus [4]. Many studies have shown that there remain latently infected resting CD4+ T cells in the lymph nodes and other organs; moreover these reservoirs can be long-lived, with a mean half-life of more than 3 years [13]. A clinical study of 36 HIV patients on ART showed the presence of viruses able to replicate in 34 of the patients, with a typical occurrence of .1 - 1.0 infectious units per million resting CD4+ T cells [13].

To summarize, while ART is effective in suppressing HIV replication indefinitely, it does not eradicate all the virions in the system and HIV seems to return in virtually all patients, and sometimes quite quickly when treatment is stopped. The reason for the term "HIV reservoir" is that the genome of the virus is securely protected by a seemingly healthy cell until something activates it to continue to HIV production. While the definition of reservoir has other interpretations, evidence shows this is the most likely so we use that here [12] and do not work with sanctuary sites, except indirectly. We make some simplifying assumptions throughout; one is that the latently infected cells are CD4+ T cells, even though there is evidence that other types of cells may serve as reservoirs for HIV [1]. The other assumption is that latently infected CD4+ T cells circulate, though our model allows for the existence of resident cells as well [25]. This is discussed in Section 2.3.

In this paper we construct a mathematical model based on limited data about the reservoirs that have been observed and analyzed in clinical studies. The virions, as well as most actively and latently infected cells circulate throughout the body through biological pathways, and the location of a latently infected cell at any given moment is extremely uncertain. While most studies indicate that CD4+ T cells circulate throughout a body, we incorporate the observation that some CD4+ T cells remain resident in certain locations [24,25]. Our use of stochastic methods to understand and model the location of the latently infected cells, and how they impact a blood test, does not assume that the migration of the cells is random. Indeed the movement of T cells throughout a body is subject to severe physical stresses (fast blood flow) as well as pressure to remain near lymph nodes. Our model uses the stochastic features of the movement, bolstered by some theorems that estimate the rate of spread, and the limiting distributions. We conclude that after a fairly short period of time (perhaps measured in days), latently infected cells have spread sparsely throughout the body. This indicates that treatment ought to follow the same pathways and dynamical processes in order to reach the hidden virus.

We introduce and analyze the mathematical models in Sections 2.1 and 2.3. We give a simple two-state Markov process model of the outcome of blood tests for the presence of viral particles in Sec. 2.1, and in Sec. 2.3 we show the existence of a hidden Markov model that more accurately reflects the dynamics of the latently infected cells, and show it is a lifting of the two-state model. In Section 2.2 we give an overview of the location of the reservoirs based on the scientific literature. In Section 3 we assign some specific sample numbers to the entries in the matrices and compare the resulting measures on both models.

2. Theory and Models

2.1. The two-state model

When a blood sample is taken from an HIV positive patient under drug treatment, it is expected that the viral presence will be below the detectable levels [3,8]. More elaborate tests can be performed to assess the presence of latently infected CD4+ T cells [3,8] in the blood, however there is no simple method for measuring the presence and level of latently infected cells [16]. Therefore, after each sample is tested, we can think of the blood as being in one of two states, either the test is positive for the presence of such cells (state $P = 1$) or negative (state $N = 0$). We assume this test is performed once a month and the outcome is recorded. We also assume the blood is sampled monthly for an indefinite period of time, and we construct a Markov chain from the results.

This yields the following 2×2 incidence matrix B, where $b_{ij} = 1$ if and only if you can get from state i to state j in one time step. Clearly a negative test can be followed by a positive one, and the same result can occur twice in a row; our assumption is that once the presence of latently infected CD4+ T cells is established with a positive blood test result, the latently infected cells cannot be destroyed easily though the blood test could come back negative if the quantity of latently infected cells, believed to be very sparse, is not seen on a subsequent test. The adjacency graph in Figure 1 shows the possible connections between the nodes of Positive and Negative; even though the likelihood of each arrow is different (as shown in (2.2)), the graph shows it is possible for each blood test outcome to be followed by either outcome.

$$B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \tag{2.1}$$

The matrix is equivalent to the directed graph in Fig. 1, where $B = b_{ij}$, and $b_{ij} = 1$ if and only if there is an arrow from state i to j and 0 otherwise.

We start with a simple one-step Markov process associated to B; we review the mathematical underpinnings of a Markov process first. We begin with the space of all possible infinite sequences of outcomes, $\Omega = \{0, 1\}^{\mathbb{N}}$; a point $x \in \Omega$ is a one-sided sequence of
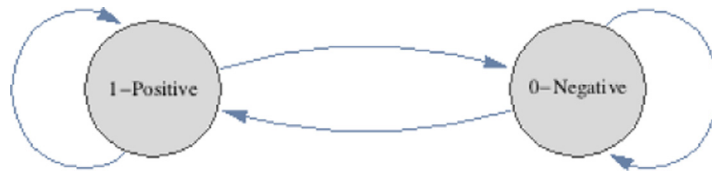
Fig. 1. Labelled graph showing the connections between the peripheral blood samples that test positive for latently infected CD4+ T cells and the samples that come back negative.

0's and 1's, written for example $x = 01101\cdots$. All strings of 0's and 1's can occur for B given in (2.1) and $x_j \in \{0, 1\}$ denotes the jth coordinate of x. The dynamical system we study on Ω is the left shift σ, given by $\sigma(x)_j = x_{j+1}$, showing the passage of time.

We then associate a stochastic matrix V (a 2×2 matrix with each row j, with $j = 0, 1$, summing to 1), obtained from B with probabilities assigned by $v_{ij} \geq 0$ representing the probability of moving from state i to state j in one time step. Therefore $v_{ij} > 0$ if and only if $b_{ij} = 1$. For this model, for some $\alpha, \epsilon > 0$, V is:

$$V = \begin{pmatrix} \alpha & 1 - \alpha \\ \epsilon & 1 - \epsilon \end{pmatrix} = \begin{pmatrix} v_{00} & v_{01} \\ v_{10} & v_{11} \end{pmatrix} \qquad (2.2)$$

The value of ϵ will be small (e.g., $\epsilon = .1$) but not tiny, because a positive test for latently infected cells would typically be followed by another positive test; however due to the sparsity of the latently infected cells in the bloodstream [8], there is a significant probability that a subsequent test would come back negative. We somewhat arbitrarily assign for now the value $v_{10} = .1$. The value of α is even harder to determine precisely; it is likely to be large since the reservoirs do not show up in the blood immediately and measurement errors occur, so we use $v_{00} = .7$ as a test value. Recall that v_{00} represents the probability that a negative blood test is again negative a month later, and v_{10} is the probability that a positive test is negative after one month. It is easy to calculate by hand that the matrix V has a left eigenvector $w = (w_0, w_1)$ for the eigenvalue 1, with positive entries adding to 1, of the form:

$$w = \left(\frac{\epsilon}{(1 - \alpha + \epsilon)}, \frac{1 - \alpha}{(1 - \alpha + \epsilon)} \right). \qquad (2.3)$$

Then $wV = w$ and $w_0 + w_1 = 1$; i.e., w is a stationary probability vector for V.

We mention the significance of the entries of w from (2.3); no matter what the initial conditions are that might impact the expected results of the first blood test, applying Theorem 2.1 below about Markov processes, called Doeblin's theorem (see [29], Thm. 2.2.1) we have that, upon repeating the test monthly, the system will approach the distribution given by (2.3).

We define a sequence of $\{0, 1\}$-valued random variables $\{Y_n : n \geq 1\}$ as follows: $Y_1 = 0$ if the first blood test shows are no latently infected cells in the peripheral blood, and 1 if positive. Then Y_2 is the test result at the next observation, and Y_n is the test result at the nth observation. We assume that Y_n depends on Y_{n-1} and not on earlier Y_j's. (This assumption perhaps should be dropped, and we change the measure on the model in the next section, but for now it simplifies the model and illustrates the idea.)

Moreover, Doeblin's theorem includes a rate of convergence to the limiting distribution. To explain this, we first note that if q is a row vector with i^{th} entry $q_i = \mathbb{P}(Y_1 = i)$, then q is called the *initial distribution* of the Markov chain and if P is any $k \times k$ stochastic matrix, then $qP^n_j = \mathbb{P}(Y_n = j)$ for each $n \geq 0$, $j \in \mathcal{A}$, where $P^n = P \cdot P \cdots P$ (n-fold matrix product). Given any row vector q with k entries, we can assign a length or size to the vector q as follows:

$$\|q\| = \sum_{i=0}^{k-1} |q_i|. \qquad (2.4)$$

One can easily check that $\|qP\| \leq \|q\|$ since P is stochastic. We are now ready to state the theorem.

Theorem 2.1 (Doeblin's Theorem [29])**.** *If P is a stochastic matrix for a Markov process, and for some state j_o there exists some $\gamma > 0$ such that $P_{ij_0} \geq \gamma$ for all $i = 0, 1, \ldots, k - 1$, then there exists a unique stationary probability vector p, with $p_{j_0} \geq \gamma$, and for any initial distribution vector q,*

$$\|qP^m - p\| \leq 2(1 - \gamma)^m, \quad \forall m \geq 0. \qquad (2.5)$$

We note that with the choices we made for entries in the matrix V in (2.2), we obtain the limiting proportions as:

$$w = \left(\frac{1}{4}, \frac{3}{4} \right), \qquad (2.6)$$

which means that even if 99% of the time we expect the first test for latent virions to be positive, subsequent tests will likely yield a negative test result eventually (25% of the time). Doeblin's Theorem shows that, under our assumptions, after one year the test should come back positive more than 70% of the time. Moreover, under our assumptions on V, the test should come back negative 3 months in a row only 12% of the time, given our matrix entries for V, which are based on the data that shows that virtually all HIV positive patients have some latent reservoirs [6,7]. The model then helps in determining if a cure is effective in the sense that if a blood test comes back negative for 6 months in a row, then it is quite unlikely the latent virus is present, since if there is latent virus present this result should only occur 4.2% of the time according to the model. This percentage is obtained by taking $w_0 v_{00}^5 \approx .042$.

The analysis above shows that if a monthly blood test is negative three or more times in a row, then it is likely there are no latent reservoirs present which would be useful in a clinical setting. However, whether or not a blood test for the presence of latent viral particles comes back positive or negative depends on many more factors than the outcome of the previous test; in particular the location of the reservoir could impact the time it takes for it to show up in a blood sample. This model could be improved to contain more information; we would like to understand the hidden Markov process behind this one, in order to understand what the appropriate measure should be on the resulting Markov process. In other words, the outcome of a blood test is far from random and depends on a variety of factors. Moreover, if a blood test were developed that is more accurate than testing for active viral presence, but less elaborate than the current complicated assays that involve collecting a large blood sample, its accuracy would need to be analyzed. We turn to a more detailed mathematical model, but one whose probabilities remain concealed from most medical practice, so it is termed a hidden Markov process.

2.2. Locations of the latent reservoirs

It is a widely held belief, based on in vivo observations, that most of the latent viral cells are located in lymph nodes, or more precisely, in the resting CD4+ T cells in lymph nodes and lymphatic tissue [3]. Resting CD4+ T cells are cells that have been activated

by an antigen, and then have made a transition back to a resting state that are then able to respond to the same antigen in the future; they are a type of long-lived memory T cells. Typically the reservoirs consisting of resting memory CD4+ T cells carry integrated viral genomes. The assumption is that such a reservoir of latently infected, long-lived, drug resistant cells re-establishes infection once ART is stopped and therefore these reservoirs prevent HIV from being cured. Understanding the properties of these cells is therefore important to eradicating HIV. We describe the movement of latently infected cells circulating in the blood and lymph, and other locations outside the blood and lymph vessels where latent or active HIV has been found.

Since the cells in reservoirs are not actively infected, the drug therapy does not kill them or prevent them from infecting their host cells at a later time. These latently infected cells travel to different compartments of the body where they can either productively infect nearby cells at a slow rate or do nothing for extended periods of time and then become active [3]. We review the major repositories of latently infected cells. The lymphatic system is the second part of the circulatory system (the cardiovascular being the first); this system is connected to, but quite different from the cardiovascular system insofar as the fluid, consisting of lymph and the immune cells called lymphocytes, circulates much more slowly and is primarily moved by surrounding muscles. There are associated lymphatic tissue and organs that filter pathogens from the lymph before returning it to the bloodstream.

1. The peripheral blood system is the pool of circulating blood in the cardiovascular system; it contains red blood cells, white blood cells (including latently infected CD4+ T cells), and platelets suspended in plasma. The blood is rapidly circulated through a closed circulatory system pumped by the heart. Red blood cells do not escape the cardiovascular system in a healthy individual, though white blood cells, slightly smaller, do.
2. Gut associated lymphatic tissue (GALT) is the site of many CD4+ T cells; indeed it is the largest immune organ in the body. Most GALT CD4+ T cells are in an activated (not resting), memory state. Clinical tests have shown the presence of significant numbers of latently infected cells in GALT, around 36 copies per million cells [17,23].
3. Lymph nodes are (some of the) filtering compartments in the lymphatic system that have an elaborate structure through which lymph flows in order to attack pathogens. The lymph nodes in a human (there are likely to be 500–600 of them per person) also house a large percentage of CD4+ T cells, including a large percentage of the latently infected ones [3].
4. Cerebrospinal fluid (CSF) and central nervous system (CNS) house reservoirs of HIV as well. Infected cells in the CSF and CNS appear from the early stages of infection, with or without symptoms [15,30], and the infected cells harboring the viruses are believed to be CD4+ T cells though the concentration is lower than in other parts of the body.
5. Bone marrow hosts HIV reservoirs as well. Clinical studies seem to indicate that the HIV DNA is latently present in CD4+ T cells found in bone marrow, but is not likely to be in the stem cells that eventually become lymphocytes [9,11,33].

The are some factors to consider when modeling HIV reservoirs. Even in an individual with no disease, the circulation of CD4+ T cells, as well as the location of resident (non circulating) cells is somewhat of a mystery, and testing for the latently infected ones among them is difficult [16]. However some data collected in cooperation with the New York Organ Donor Network (NYODN), reveals some T-cell distribution information [24]. In general one cannot sample the areas listed above in vivo, so the presence or absence of latently infected cells is inferred from blood samples taken from patients with HIV and undergoing ART. Moreover the latently

Table 1
The states of the Markov process represent compartments in the body where latent HIV infected cells are found

State k	Circulatory compartment
0	Lymph nodes (node)
1	Peripheral blood (blood)
2	Bone marrow (bone)
3	Cerebral-Spinal Fluid (CSF)
4	Gut-associated lymphatic tissue (GALT)

infected CD4+ T cells comprise a small percentage of the resting CD4+ T cells; estimates range from .02% to about .065%, with the mean among patients undergoing ART to be around .038% [8]. The number also varies below and above these values and there are studies showing how difficult it is to detect and quantify the reservoirs [16]. A peripheral blood sample is typically used to test for the presence of latently infected CD4+ T cells, and only 2% of the body's CD4+ T cells can be found in the blood at any given time. Moreover typically if someone is HIV positive and on ART then they do not have their blood tested more often than once a month.

2.3. The Hidden Markov Process Model of Latent Cell Dynamics

A model that captures the dynamics of a latently infected CD4+ T cell circulating among or resident in the different anatomical compartments is desirable, but it is largely unknown how such a model should look. Studies have been done on the movement of CD4+ T cells throughout a body and the observation that some CD4+ T cells remain resident in certain locations [24,25]. While the data is still sparse, we can improve the model given in the previous section by setting up the hidden Markov model. If it were known precisely where each latently infected cell was hiding out, targeted therapy could cure HIV; however that is still impossible to determine so we describe a math model which can nonetheless capture some of the important features of the migration of these cells. The reader needs to keep in mind that we have moved from the "real" (testable) Markov process to an idealized one.

We model the viral spread within an individual human patient using a Markov process on a finite state space $\mathcal{A} = \{0, 1, \ldots k - 1\}$, where k is the number of compartments of the body in which CD4+ T cells with latent virus have been found. We are interested in the dynamics of a randomly chosen latently infected resting CD4+ T cell from the peripheral blood; since this fluid circulates quickly, we choose a time increment for taking measurements, assigning a time increment to be a day is a reasonable choice. With this model we do not concern ourselves with precisely which cell we are considering since the information obtained is statistical. In other words, we study the statistical properties of a generic latently infected cell located in one of the compartments.

We define the family of $\mathcal{A}$-valued random variables $\{X_n : n \geq 0\}$ as follows: X_0 is the compartment where the cell resides in the body at the beginning of our observations. Then X_1 is the compartment where the cell resides at the first (time 1) observation, and X_n is the compartment at the nth observation. We assume that X_n depends on X_{n-1} and not on earlier X_j's; this is a reasonable observation given that some of the compartments allow flow from one to another, but not all. We define the states 0, 1, 2, 3, 4 in Table 1.

We next construct a $k \times k$ incidence matrix, which is given by $B = b_{ij}$, with $b_{ij} = 1$ if in one time step a CD4+ T cell can move from compartment i to j, and 0 otherwise. Using this notation, if B^n is the product of n copies of the matrix, writing $B^n = b_{ij}^n$, it follows that $b_{ij}^n = 1$ if a CD4+ T cell can move from compartment i to j in n steps, and is 0 otherwise. In the current model we use the matrix given by Eqn. (2.7), obtained by the graph given in Fig. 2.
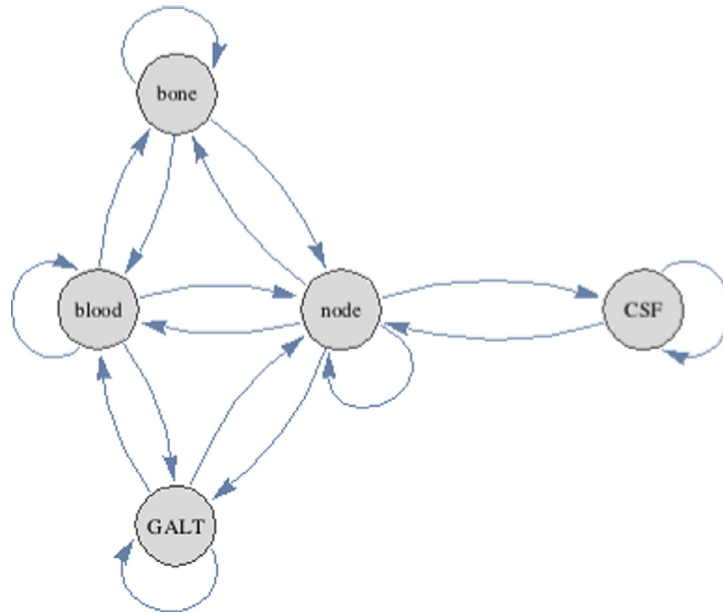
Fig. 2. Labelled graph showing the fluid flow connections in one day among the compartments listed in Table 1.

We next need a stochastic matrix $P = p_{ij}$ with the property that

$$p_{ij} = \mathbb{P}(X_n = j | X_{n-1} = i),$$

where $\mathbb{P}$ denotes probability. It is evident that $p_{ij} > 0$ if and only if $b_{ij} > 0$ for the corresponding incidence matrix B_0. We consider the meaning of these probabilities in the first example by setting the entries as shown in Table 1.

We can fine-tune the model of the presence or absence of latently infected CD4+ T cells within an individual human patient using a Markov process with more states; i.e., $\mathcal{A} = \{0, 1, \ldots, 4\}$. We split the blood system and lymphatic system into smaller compartments that manifest different dynamics, using Section 2.2 and the data from the references given there. In Figure 2 we show the directed graph of the connections among these compartments, over one day, where latently infected cells have been found. We label the new states as Peripheral blood, GALT, Bone marrow, Cerebral-Spinal fluid, and Lymph nodes, assigning numbers $0 - 4$ respectively as shown in Table 1. (The term peripheral refers to the circulating pool of blood.)

The incidence matrix with a time step of a day corresponding to the graph in Figure 2 is given by the matrix B, where $b_{ij} = 1$, where $i, j = 0, 1, 2, 3, 4$ if and only if a latently infected CD4+ T cell in compartment i can enter j.

$$B_{\text{day}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix} \tag{2.7}$$

If we had perfect information then we would be able to associate a stochastic matrix to the incidence matrix by following where the CD4+ T cells move over time. There have been studies done on the migration of T-cells [31], the fact that there are resident (non-migratory) T-cells [24,25], and the location of HIV reservoirs. First we synchronize the time step to match the blood test time step of one month (four weeks). Since B_{day}^2 has only positive entries, in one month, a CD4+ T cell could in theory travel from any one compartment to any other. The model does not imply that they will

migrate, as some do not move, rather it allows for either possibility. Therefore the monthly graph is as in Figure 3 and the incidence matrix is as in (2.8).

$$B_{\text{month}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \tag{2.8}$$
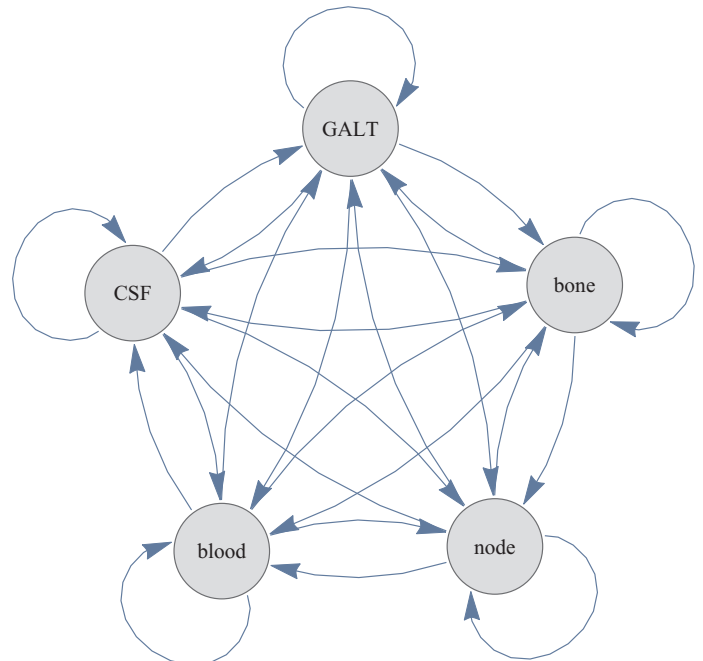

Fig. 3. Labelled graph showing the fluid flow connections in one month among the compartments listed in Table 1.

3. Results and Discussion

The precise mechanism by which the latently infected cells move through the circulatory and lymphatic systems remains unknown. Therefore while we can form the incidence matrix and the factor Markov shift with some confidence, we have incomplete information about the stochastic matrix P of hidden Markov shift given by (3.1):

$$P_{\text{month}} = \begin{pmatrix} p_{00} & p_{01} & p_{02} & p_{03} & p_{04} \\ p_{10} & p_{11} & p_{12} & p_{13} & p_{14} \\ p_{20} & p_{21} & p_{22} & p_{23} & p_{24} \\ p_{30} & p_{31} & p_{32} & p_{33} & p_{34} \\ p_{40} & p_{41} & p_{42} & p_{43} & p_{44} \end{pmatrix} \quad (3.1)$$

Each choice of parameters will yield a stochastic left eigenvector for the eigenvalue 1, and some choices might yield the same vector p, so the problem of determining the values is ill-posed. However we consider an example for the purposes of illustration. We consider the following associated stochastic matrix by using data given in the references in Section 2.2:

$$P_{\text{day}} = \begin{pmatrix} .5 & .19 & .005 & .005 & .3 \\ .35 & .4 & .05 & 0 & .2 \\ .3 & .2 & .5 & 0 & 0 \\ .2 & 0 & 0 & .8 & 0 \\ .4 & .1 & 0 & 0 & .5 \end{pmatrix} \quad (3.2)$$

Since each row sum in P_{day} is 1, every power of the matrix is also stochastic so we can calculate $P_{\text{month}} = P_{\text{day}}^{28}$ easily by computer. Using the entries of P_{day}, the left eigenvector for the eigenvalue 1 is approximately:

$$p = (0.428, 0.200, 0.024, 0.011, 0.337\}. \quad (3.3)$$

Since $P_{\text{day}}^2 > 0$, it is called primitive, and the following result holds (see e.g., [20], Chap. 4).

Proposition 3.1. *Let P be a primitive stochastic matrix. Then there exists a unique strictly positive probability vector, p, with pP = p, and for any probability vector q,*

$$\lim_{n \to \infty} qP^n = p.$$

Therefore, since $P_{\text{month}} = P_{\text{day}}^{28}$, we have that:

$$\lim_{k \to \infty} qP_{\text{month}}^k = \lim_{k \to \infty} qP_{\text{day}}^{28k} = p$$

as well, and indeed just computing, we see that rounded to the nearest .001,

$$P_{\text{month}} = \begin{pmatrix} .428 & .200 & .024 & .011 & .337 \\ .428 & .200 & .024 & .011 & .337 \\ .428 & .200 & .024 & .011 & .337 \\ .428 & .200 & .024 & .012 & .336 \\ .428 & .200 & .024 & .011 & .337. \end{pmatrix} \quad (3.4)$$

We note that while the assigned probabilities are somewhat arbitrary, the existence of a vector p with positive entries, is guaranteed and the entries of p will depend continuously on the entries of P. As in the simpler model from Section 2.1, we can interpret the jth vector entry in p as the probability that a latently infected cell can be found in compartment j. With this refinement of the compartments in which latent CD4+ T cells are found in patients, the model shows that a "random" latently infected CD4+ T cell is almost three times as likely to be found in the lymphatic system (GALT, 34%, or lymph nodes, 43%, giving 77%) as the blood flow (20% likelihood), over time and independently of where the latent infection started. We also see that the model shows what is supported by clinical data, that there is a positive probability that there will be a steady state of latently infected cells found in the bone marrow and cerebral spinal fluid [9,11,30]. Applying Theorem 2.1 to the matrix P, we see that after about four days, we are quite close to the limiting distribution in the sense that each entry is within about .01 of the limiting value, so at one month we see we are very close to the limiting distribution.

We note that other sample matrices P_{day} would likely yield different limiting distributions, but for the same reasons as above, these distributions would be apparent already in P_{month}.

3.1. Connections between the two Markov models

If we consider the space $\Lambda = \mathcal{A}^{\mathbb{N}}$, then we have a continuous shift commuting map from Λ onto Ω as follows. For any $a \in \Lambda$, $a = a_0 a_1 \ldots a_k \ldots$, with $a_k \in \mathcal{A}$. Define $\phi(a) = 1$ if $a_0 = 1$, and $\phi(a) = 0$ otherwise. In other words, we only map the point a to a "positive blood test" if the coordinate $a_0 = 1$ signifying the latently infected cell is currently in the blood. Letting σ_Ω and σ_Λ denote the shift on the spaces Ω and Λ respectively, then clearly for all points $a \in \Lambda$,

$$\phi \circ \sigma_\Lambda(a) = \sigma_\Omega \circ \phi(a).$$

For simplicity of notation, we write $Q = P_{\text{month}}$ and let q_{ij} denote the ij^{th} entry of Q, $i, j = 0, \ldots, 4$. From the pair (p, Q) we we obtained in (3.4), we define a measure ν on Λ by setting, for each finite sequence $c_0, c_1, \ldots, c_m$, with $c_j \in \mathcal{A}$

$$\nu(\{a \in \Lambda : a_k = c_k, \, k = 0, \cdots, m\}) = p_{c_0} \cdot q_{c_0 c_1} \cdots q_{c_{m-1} c_m}. \quad (3.5)$$

We call (σ_Λ, ν) the (p, Q)-Markov shift, and note that ν is invariant for σ_Λ. The measure ν in turn induces a measure $\tilde{\nu}$ on the space Ω, which we view as a factor space. In particular, for any measurable set $U \in \Omega$, $\tilde{\nu}(U) = \nu(\phi^{-1}(U))$.

The pair (w, V) from (2.2) and (2.3) induces a measure μ on Ω exactly as shown in Eqn (3.5) but using the vector w and matrix V, so it is useful to compare them. We calculate each one on the same set, using for example the set of outcomes that come back "Negative, Negative, Positive" on 3 consecutive blood tests; i.e.,

$$U_{001} = \{x \in \Omega : x_0 = 0, x_1 = 0, x_2 = 1\}. \quad (3.6)$$

Using the Markov measure μ on Ω,

$$\mu(U_{001}) = \frac{\epsilon}{(1 - \alpha + \epsilon)} \cdot \alpha \cdot (1 - \alpha) \approx .053,$$

(using the sample values of $\epsilon = .1$ and $\alpha = .7$ as chosen in Sec. 2.1); and with respect to the hidden Markov measure we compute

$$\tilde{\nu}(U_{001}) = p_0 q_{00} q_{01} + p_0 \left(\sum_{j=2}^{4} q_{0j} q_{j1} \right)$$
$$+ \sum_{k=2}^{4} \left(p_k q_{k0} q_{01} + \sum_{i=2}^{4} p_k \left(\sum_{j=2}^{4} q_{kj} q_{j1} \right) \right)$$
$$\approx .128$$

(using the numbers given in (3.4)). Clearly these two measures can yield quite different answers in general, and we hypothesize that the measure $\tilde{\nu}$ is more accurate, so solving the hidden Markov model problem is worthwhile. Specifically, given enough blood test data, a subject of ongoing study is to deduce the matrix in (3.2) or (3.4).

3.2. Conclusions

While latently infected CD4+ T cells are extremely sparse, they are a reality that prevents HIV from being cured, so we use a probabilistic model to follow their location, migration and subsequent detection by a blood test. We model the dynamics of a random latently infected CD4+ T cell in an HIV positive patient using a Markov process, and show there is a limiting distribution, which indicates in which compartments the latently infected cells can be found. Our model shows that the limiting distribution of latent cells contains cells in every compartment with positive probability, supported by clinical data. We also show that a hidden Markov model determines the outcome of blood tests and describe its structure in general terms. From the analysis given, while efforts are made to remove the latently infected cells, these models provide information helpful to interpret the likelihood that for some n, and n could be 3 or 4, n blood tests with negative results for the presence of viral DNA indicate with high probability that the virus has been eradicated. This analysis allows for less intrusive and less costly, and therefore possibly less accurate, blood tests for the latent virus than are currently being used, but a larger data set can be obtained for future research.

References

[1] A. Alexaki, Y. Liu, B. Wigdahl, Cellular reservoirs of HIV-1 and their role in viral persistence, Current HIV Research 6 (5) (2008) 388–400.

[2] J. Ball, E. Holmes, H. Whitwel, U. Desselberger, Genomic variation of human immunodeficiency virus type 1 (HIV-1): molecular analyses of HIV-1 in sequential blood samples and various organs obtained at autopsy, Jour. of Genl. Virology 75 (1994) 867–879.

[3] D. Barouch, S. Deeks, Immunologic strategies for HIV-1 remission and eradication, Science 345 (2014) 169–174.

[4] J. Blankson, J. Siliciano, R. Siliciano, Finding a cure for human immunodeficiency virus-1 infection, Infect Dis Clin N Am 28 (2014) 633–650.

[5] E. Burkhead, J. Hawkins, D. Molinek, A dynamical study of a cellular automata model of the spread of HIV in a lymph node, Bull. Math. Biol. 71 (1) (2009) 25–74.

[6] C. Chan, I. Dietrich, M. Hosie, B. Willett, Recent developments in human immunodeficiency virus-1 latency research, Journal of General Virology 94 (2013) 917–932.

[7] T. Chun, A. Fauci, HIV reservoirs: pathogenesis and obstacles to viral eradication and cure, AIDS 26 (2012) 1261–1268.

[8] T. Chun, J. Justement, J. Lempicki, R. Yang, Gene expression and viral production in latently infected, resting CD4 t cells in viremic versus aviremic HIV-infected individuals, PNAS 100 (4) (2003) 1908–1913.

[9] B. Davis, D. Schwartz, J. Marx, C. Johnson, Absent or rare human immunodeficiency virus infection of bone marrow stem/progenitor cells in vivo, Journal of Virology 65 (4) (1991) 1985–1990.

[10] P. Denton, J. Long, S. Wietgrefe, C. Sykes, Targeted cytotoxic therapy kills persisting HIV infected cells during ART, PLOS Pathogens 10 (1) (2014) 1–9. E1003872.

[11] C. Durand, G. Ghiaur, J. Siliciano, S. Rabi, HIV-1 DNA Is Detected in Bone Marrow Populations Containing CD4+ T Cells but Is not Found in Purified CD34+ Hematopoietic Progenitor Cells in Most Patients on Antiretroviral Therapy, Infect Dis. 205 (6) (2012) 1014–1018.

[12] E. Eisele, R. Siliciano, Redefining the viral reservoirs that prevent HIV-1 eradication, Immunity 37 (2012) 377–388.

[13] D. Finzi, J. Blankson, J. Siliciano, J. Margolick, Latent infection of CD4+ t cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy, Nature Medicine 5 (5) (1999) 512–517.

[14] F. Garcia, M. Plana, C. Vidal, A. Cruceta, Dynamics of viral load rebound and immunological changes after stopping effective antiretroviral therapy, AIDS 13 (1999) F79–F86.

[15] F. Gonzalez-Scarano, J. Martin-Garcia, The neuropathogenesis of AIDS, Nature Reviews, Immunology 5 (2005) 69–81.

[16] Y. Han, M. Wind-Rotolo, H. Yang, J. Siliciano, Experimental approaches to the study of HIV-1 latency, Nature Rev. Microbiology 5 (2007) 95–106.

[17] A.T. Hasse, Population biology of HIV-1 infection: Viral and $CD4^+ T$ cell demographics and dynamics in lymphatic tissues, Annu. Rev. Immunol. 17 (1999) 625–656.

[18] J. Hawkins, D. Molinek, Markov cellular automata as models for chronic disease progression, Intl. J. of Biomath. (2015), doi:10.1142/S1793524515500850.

[19] H. Ledford, HIV rebound dashes hope of 'Mississippi baby' cure, Nature 10 (2014), doi:10.1038/nature.2014.15535.

[20] D. Lind, B. Marcus, An Introduction to Symbolic Dynamics and Coding, Cambridge Univ. Press, Cambridge, UK, 1995.

[21] G. Pantaleo, A.S. Fauci, Immunopathogenesis of HIV infection, Ann. Rev. of Microbiology 50 (1996) 825–854.

[22] A. Perelson, S. Deeks, Drug effectiveness explained: The mathematics of antiviral agents for HIV, Sci. Transl Med 3.91 (2013) 1–7.

[23] M. Poles, J. Boscardin, Lack of decay of HIV-1 in gut-associated lymphoid tissue reservoirs in maximally suppressed individuals, J Acquir Immune Defic Syndr 43 (1) (2006) 65–68.

[24] T. Sathaliyawala, M. Kubota, N. Yudanin, D. Turner, Distribution and compartmentalization of human circulating and tissue-resident memory t cell subsets, Immunity 38.1 (2013) 187–197.

[25] J. Schenkel, D. Masopust, Tissue-resident memory t cells, Immunity 41 (2014) 886–897.

[26] S. Schnittman, M. Psallidopoulos, H. Lane, L. Thompson, The reservoir for HIV-1 in human peripheral blood is a T cell that maintains expression of CD4, Science 245 (4915) (1989) 305–308.

[27] L. Shen, S. Rabi, A. Sedaghat, L. Shan, A critical subset model provides a conceptual basis for the high antiviral activity of major HIV drugs, Sci Transl Med. 3 (91) (2011) 91ra63.

[28] L. Shen, S. Rabi, R. Siliciano, A novel method for determining the inhibitory potential of anti-HIV drugs, Trends Pharmacol Sci. 30 (12) (2009) 610–616.

[29] D. Stroock, An introdcution to markov processes, Springer GTM 230 (2005).

[30] V. Valcour, T. Chalermchai, N. Sailasuta, M. Marovich, Central nervous system viral invasion and inflammation during acute HIV infection, Journal of Infectious Diseases 206 (2) (2012) 275–282.

[31] U.V. Adrian, C. Mackay, T-cell function and migration: Two sides of the same coin, NEJM 343 (14) (2000) 1020–1034.

[32] S.A. Yukl, E. Boritz, M. Busch, C. Bentsen, Challenges in detecting HIV persistence during potentially curative interventions: A study of the berlin patient, PLoS Pathog 9 (5) (2013) 1–13. E1003347.

[33] V. Zhdanov, Kinetic model of HIV infection including hematopoietic progenitor cells, Math. biosci. 236 (2012) 36–43.

[34] R.M.Z. dos Santos, S. Coutinho, Dynamics of HIV infection: A cellular automata approach, Phys. Rev. Lett. 87 (16) (2001) 168102-1–168102-4.