

# A Stochastic Model of Latently Infected Cell Reactivation and Viral Blip Generation in Treated HIV Patients

Jessica M. Conway, Daniel Coombs\*

Department of Mathematics and Institute of Applied Mathematics, University of British Columbia, Vancouver, British Columbia, Canada

## Abstract

Motivated by viral persistence in HIV+ patients on long-term anti-retroviral treatment (ART), we present a stochastic model of HIV viral dynamics in the blood stream. We consider the hypothesis that the residual viremia in patients on ART can be explained principally by the activation of cells latently infected by HIV before the initiation of ART and that viral blips (clinically-observed short periods of detectable viral load) represent large deviations from the mean. We model the system as a continuous-time, multi-type branching process. Deriving equations for the probability generating function we use a novel numerical approach to extract the probability distributions for latent reservoir sizes and viral loads. We find that latent reservoir extinction-time distributions underscore the importance of considering reservoir dynamics beyond simply the half-life. We calculate blip amplitudes and frequencies by computing complete viral load probability distributions, and study the duration of viral blips via direct numerical simulation. We find that our model qualitatively reproduces short small-amplitude blips detected in clinical studies of treated HIV infection. Stochastic models of this type provide insight into treatment-outcome variability that cannot be found from deterministic models.

**Citation:** Conway JM, Coombs D (2011) A Stochastic Model of Latently Infected Cell Reactivation and Viral Blip Generation in Treated HIV Patients. *PLoS Comput Biol* 7(4): e1002033. doi:10.1371/journal.pcbi.1002033

**Editor:** Christophe Fraser, Imperial College London, United Kingdom

**Received:** November 4, 2010; **Accepted:** March 1, 2011; **Published:** April 28, 2011

**Copyright:** © 2011 Conway, Coombs. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by grants from NSERC (DC), MITACS (JMC and DC), and by the Canadian Institutes of Health Research [funding reference number HFE-105370 to JMC]. Also, this research has been enabled by the use of computing resources provided by WestGrid and Compute/Calcul Canada. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: coombs@math.ubc.ca

## Introduction

HIV infection can be effectively controlled by anti-retroviral drug therapy (ART) [1,2]. Different ART drugs inhibit different steps of HIV replication, and therefore truly effective therapy should halt viral production altogether. However, while plasma viral load is greatly decreased in patients on ART, it remains non-zero [3–5]. The sources of residual viremia remain under debate. One common argument is that the drugs may not be 100% effective, implying that the low-level viral load is associated with some residual viral replication. Older papers (pre-2004) present considerable evidence for this hypothesis [6–8]; for example, Havlir et al. [8] noted that, in patients on long-term suppressive therapy, the introduction of an improved drug into their regimen decreased the level of residual viremia.

However, the efficacy of ART drugs has improved substantially since their inception and the likelihood of substantial ongoing viral replication has correspondingly diminished. A recent phylogenetic study of virus before treatment and during structured treatment interruptions found that the viral samples were too closely related for there to have been significant ongoing replication [9]. Other studies measured residual viremia in patients on ART before and after treatment intensification, and found no change in residual viremia [10,11] (although the latter paper intriguingly discovered signs of replication in certain patients even though their plasma viral load was maintained at extremely low levels). Together, these works indicate that there are important sources of virus in treated patients and these sources are largely independent of ongoing viral replication.

## The latent reservoir during HIV infection

There are many locations in the body from which viruses could re-emerge during drug treatment; for a review, see [6]. Here, we will focus on the important possibility that viruses may emerge from a reservoir of latently infected cells. Usually when HIV infects target cells (such as CD4+ T cells and macrophages) the result is rapid virus production and cell death. However, a fraction of infected cells are known to enter a state of latent infection [12] where virus has integrated into the host cell DNA, but there is little, if any, viral gene expression. While cells are in this state, they are unaffected by ART and viral cytopathicity, and are effectively invisible to the host immune response [13]. However, upon reactivation, latent cells begin the normal processes of viral replication and production, and become immune targets [14]. A large fraction of the latent reservoir consists of resting memory CD4+ cells [15] and therefore, reactivation could occur as part of the normal immune response to a secondary pathogen [16]. However, we do not completely understand the reasons for activation of latently infected cells and it is likely that there is a pathogen-independent component as well. Indeed, the mechanisms for latency are generally poorly understood; there are differing opinions, but no consensus to date [12,17–19].

The population of latently infected cells is established as early as 10 days after symptoms of seroconversion, within a few weeks of initial infection [14]. Estimates of reservoir size differ but consistently show that latently infected cells constitute only a small fraction of the total number of T-cells [12,20]. Unfortunately, and in spite of its small relative size, the decline of this population is slow and it is estimated that it can persist for up to 70

## Author Summary

While on successful drug treatment, routine testing does not usually detect virus in the blood of an HIV patient. However, more sensitive techniques can detect extremely low levels of virus. Occasionally, routine blood tests show “viral blips”: short periods of elevated, detectable viral load. We explore the hypothesis that residual low-level viral load can be largely explained by re-activation of cells that were infected before the initiation of treatment, and that viral blips can be viewed as occasional statistical events. To do this, we propose a mathematical model of latently-infected cells, activated cells, and virus. The model captures random fluctuations of the system as well as the mean behaviour. We estimate the time it takes for all the latently-infected cells to be eradicated. Eradication of these cells is considered a major hurdle in eliminating infection. We predict a wide range of eradication times, highlighting the importance of studying latently-infected cells. We also estimate the frequency and duration of viral blips, and find qualitative agreement with clinical studies. By refining our models, we hope to find guidelines that can be used in practise to distinguish between clinically insignificant statistical blips, and instances of drug failure.

years [21]. This long lifetime probably arises from the intrinsic stability of resting memory CD4+ cells which is an important part of immune memory [22]. Recent evidence also indicates that latently infected cells can undergo cell division [15], potentially increasing the lifetime of the reservoir. These factors, in combination with long lifetime of the reservoir, indicate that latently infected cells are an important factor that must be addressed in the search for therapies to eradicate HIV infection [23].

### Viral blips during anti-retroviral treatment

While on successful anti-retroviral treatment (ART) for HIV, an infected individual’s viral load remains non-zero [3], though it is very low and usually undetectable using standard assays that have a detection limit of 50 copies/mL in plasma. Occasionally, however, regular blood tests show viral blips: periods of detectable viral load, preceded and followed by undetectable loads. At one time there was a concern that blips might signal imminent drug failure [24], including the emergence of new, drug-resistant variants of virus [25]. However, there is a substantial body of evidence dating from the early 2000s, indicating that most blips are not associated with virological failure [26–28]. With that said, a recent large-scale study of 3530 Canadian patients refined these results by showing a two-fold increase in the risk of drug failure following viral blips that exceeded 500 copies/mL, but importantly, smaller blips were not associated with drug failure [29]. In this study, blips were detected at a frequency of about 0.1/patient/year. This rate is compatible with data taken from the UK during 2006–2007 [30] and is lower than the rate estimated from earlier data [31]. The reduction in blip frequency over the last decade is likely a result of improved drug efficacy.

The underlying cause of viral blips remains controversial. There is some evidence that immune activation, through secondary infection or vaccination, may be correlated with viral blips, [32,33]. However, there have been observations of blips not associated with clinical or demographic variables. In an intensive 90-day study of 10 patients, Nettles et al. found that blips were fairly common, smaller in amplitude (mean 79 copies/mL) and short in duration (median less than 3 days), and that blip frequency

was unrelated to illness, vaccination, or drug concentrations [34]. Finally, we must acknowledge that accurate detection of small-amplitude blips is bedevilled by assay variability and sensitivity [30,34].

### Previous modeling work

There has been extensive modeling work done to characterize viral load and pathogen-immune system interaction in HIV-infected individuals. However, standard viral dynamics models do not capture residual viremia in treated patients and various modeling approaches have been applied.

Residual viremia can be captured by adding a latent cell reservoir to the standard model. Perelson et al. (1997) proposed the first model that included latent cell activation, in order to better understand decay characteristics of HIV-1-infected compartments during combination therapy [35]. This model was expanded to include a varying decay rate in the latent reservoir, and bystander proliferation in the latent reservoir, with the finding that a constant long-term activation rate for the latent reservoir, maintained through cell division, could explain residual viremia in treated patients [36]. We will use these elements in the development of our model of the latent reservoir.

Careful modelling of viral blips has been fruitful in analyzing different mechanisms of blip generation. The focus of previous models has been on blips associated with immune system activation due secondary infection or vaccination. One successful approach has been to consider short periods of sustained viral replication driven by stochastic activation of CD4+ and CD8+ T cells [37–39]. A further series of models including T cell expansion due to vaccination and secondary infection showed episodes of detectable viremia of long duration (2–3 months), with amplitudes in the range of several hundred copies/mL [40,41]. Viral blips can also result from production of virus following immune activation and clonal expansion of latently infected cells [33]; latent cell activation caused by sporadic immune activation has also been modeled as a source of viral blip generation. Indeed, antigen-induced latent cell activation has been modeled and shown to be a plausible source of viral blips [42]. Most recently, Rong and Perelson (2009) proposed a model with antigen-induced asymmetric activation and division of latently infected cells [43]. Blips produced by this model are of short duration, directly related to the length of stimulation, and of variable amplitude, consistent with observations. These models produce blips of larger (100 copies/mL) amplitude, with variable durations, and frequency depending directly on user-controlled periods of immune system activation in the model. The base mechanism in these models of the production of blips is immune system activation.

### A stochastic model of latent cell reactivation and viral load during ART

As noted above, there have been observations of small-amplitude blips not associated with clinical or demographic variables [34]. Such blips can be imagined as random biological or statistical variation around a mean undetectable viral load. In order to capture this kind of stochastic effect, continuous (differential equation based) models are inadequate.

Here, we propose a continuous-time branching process model of within-host viral dynamics for a patient undergoing successful treatment. We use this formulation to derive probability distribution functions for viral load as a function of time and examine the contribution of varying latent cell activation and proliferation to viral load. Using this methodology we first consider extinction times for the latent reservoir and examine the role of limited ongoing viral replication in replenishing the reservoir. We

then examine the hypothesis that stochastic activation of latently infected cells can maintain low-level plasma viremia and generate small intermittent viral blips. Finally, via Gillespie simulation of the branching process, we calculate viral blip durations.

### Methods

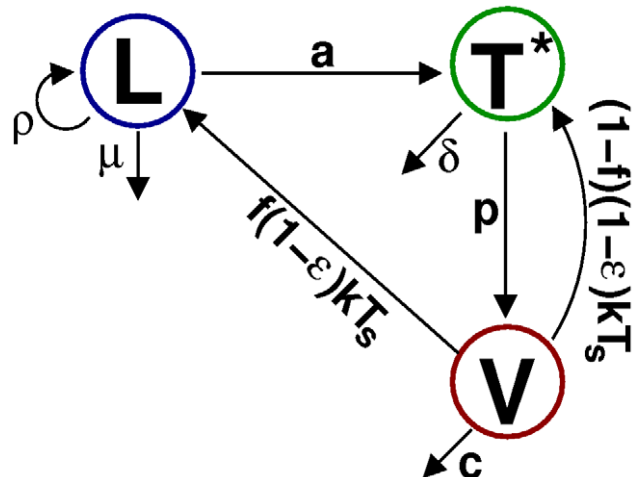
#### Viral dynamics in treated patients

We consider a simple model of latent cell reactivation, presented schematically in Figure 1. Our model has three compartments: the number of latently infected cells  $L$ , which can replicate at rate  $\rho$ , die at rate  $\mu$ , and activate at rate  $a$  to become productively infected cells; the number of productively infected cells  $T^*$ , which die at rate  $\delta$ ; and the number of virions  $V$ , produced by productively infected cells at rate  $p$ , which can die at a rate  $c$ . We allow for infection of new cells at rate  $k$ , of which a fraction  $f$  become latently infected. The efficacy of ART is given by  $\epsilon$ . We will assume that this efficacy is very high, and that therefore the number of uninfected T-cells remains approximately constant and equal to  $T_S$ . Clinical findings on viral blips show differing amplitudes [34,44,45]; though the small-amplitude blips were shown to be unassociated with clinical variables [34], it is possible that the larger-amplitude blips may be due to an immune response, increasing the activation rate  $a$  for a period of time. We will initially restrict ourselves to constant activation rate, but we consider variable  $a$  in a later section.

The mean behaviour of the system shown in Figure 1 is given by the linear system of ordinary differential equations

$$\begin{aligned} M'_L(t) &= (\rho - a - \mu)M_L(t) + f(1 - \epsilon)kT_S M_V(t) \\ M'_T(t) &= aM_L(t) - \delta M_T(t) + (1 - f)(1 - \epsilon)kT_S M_V(t) \quad (1) \\ M'_V(t) &= pM_T(t) - (c + (1 - \epsilon)kT_S)M_V(t), \end{aligned}$$

where  $M_L(t)$ ,  $M_T(t)$ , and  $M_V(t)$  represent the mean numbers of latently infected cells, productively infected cells, and virions respectively.



**Figure 1. Model schematic.** Latently infected cells (L) divide, die, and become activated with rates  $\rho$ ,  $\mu$  and  $a$  respectively. Productively infected cells ( $T^*$ ) die at rate  $\delta$  and produce virus (V) continuously, at rate  $p$ . Free virions are cleared at rate  $c$  and infect healthy cells at rate  $k \times T_S$ , reduced by drug treatment of efficacy  $\epsilon$ . A fraction  $f$  of newly infected cells become latently infected cells and the rest become productively infected cells.

doi:10.1371/journal.pcbi.1002033.g001

#### Probability distribution calculations

Our goal is to obtain probability distributions for the viral load and the size of the latent reservoir at time  $t$ . We assume that the events in the model can be described by a multi-type continuous time branching process with the rates given in Figure 1. Importantly, the model does not scale up and so any computations we perform must be over the total number of  $L, T^*$  and  $V$  in the patient.

The transition probabilities for each process in the model are stationary in time. We therefore know that our desired probability distributions depend only on the time since the initial state, and consider

$$P_{\tilde{\ell}, \tilde{n}, \tilde{v}; \ell, n, v}(t) = P(L(t) = \ell, T^*(t) = n, V(t) = v | L(0) = \tilde{\ell}, T^*(0) = \tilde{n}, V(0) = \tilde{v}),$$

the conditional probability that there are  $\ell$  latently infected cells  $L$  at time  $t$ ,  $n$  productively infected cells  $T^*$  at time  $t$ , and  $v$  virions  $V$  at time  $t$  given that there were initially  $\tilde{\ell}, \tilde{n}$ , and  $\tilde{v}$  of each species respectively. Then, given a joint initial distribution on these species  $P_{\tilde{\ell}, \tilde{n}, \tilde{v}}(0)$ , we can compute the joint probability distribution on each of these

$$p_{\ell, n, v}(t) = \sum_{\tilde{\ell}=0}^{\infty} \sum_{\tilde{n}=0}^{\infty} \sum_{\tilde{v}=0}^{\infty} P_{\tilde{\ell}, \tilde{n}, \tilde{v}; \ell, n, v}(t) p_{\tilde{\ell}, \tilde{n}, \tilde{v}}(0).$$

Note that, as the latent reservoir must be of finite size,  $P_{\tilde{\ell}, \tilde{n}, \tilde{v}; \ell, n, v}(t) \rightarrow 0$  and  $p_{\tilde{\ell}, \tilde{n}, \tilde{v}}(0) \rightarrow 0$  as  $\tilde{\ell}, \tilde{n}, \tilde{v} \rightarrow \infty$ .

By considering each process in Figure 1 in turn, we can derive the backwards Chapman-Kolmogorov differential equation for  $P_{\tilde{\ell}, \tilde{n}, \tilde{v}; \ell, n, v}(t)$  [46]:

$$\begin{aligned} \frac{dP_{\tilde{\ell}, \tilde{n}, \tilde{v}; \ell, n, v}(t)}{dt} &= a\tilde{\ell} \left( P_{\tilde{\ell}-1, \tilde{n}+1, \tilde{v}; \ell, n, v} - P_{\tilde{\ell}, \tilde{n}, \tilde{v}; \ell, n, v} \right) + \mu\tilde{\ell} \left( P_{\tilde{\ell}-1, \tilde{n}, \tilde{v}; \ell, n, v} - P_{\tilde{\ell}, \tilde{n}, \tilde{v}; \ell, n, v} \right) \\ &+ \rho\tilde{\ell} \left( P_{\tilde{\ell}+1, \tilde{n}, \tilde{v}; \ell, n, v} - P_{\tilde{\ell}, \tilde{n}, \tilde{v}; \ell, n, v} \right) + \delta\tilde{n} \left( P_{\tilde{\ell}, \tilde{n}-1, \tilde{v}; \ell, n, v} - P_{\tilde{\ell}, \tilde{n}, \tilde{v}; \ell, n, v} \right) \\ &+ p\tilde{n} \left( P_{\tilde{\ell}, \tilde{n}, \tilde{v}+1; \ell, n, v} - P_{\tilde{\ell}, \tilde{n}, \tilde{v}; \ell, n, v} \right) + c\tilde{v} \left( P_{\tilde{\ell}, \tilde{n}, \tilde{v}-1; \ell, n, v} - P_{\tilde{\ell}, \tilde{n}, \tilde{v}; \ell, n, v} \right) \quad (2) \\ &+ f(1 - \epsilon)kT_S\tilde{v} \left( P_{\tilde{\ell}+1, \tilde{n}, \tilde{v}-1; \ell, n, v} - P_{\tilde{\ell}, \tilde{n}, \tilde{v}; \ell, n, v} \right) \\ &+ (1 - f)(1 - \epsilon)kT_S\tilde{v} \left( P_{\tilde{\ell}, \tilde{n}+1, \tilde{v}-1; \ell, n, v} - P_{\tilde{\ell}, \tilde{n}, \tilde{v}; \ell, n, v} \right), \end{aligned}$$

with initial condition  $P_{\tilde{\ell}, \tilde{n}, \tilde{v}; \ell, n, v}(0) = \delta_{\tilde{\ell}\ell} \delta_{\tilde{n}n} \delta_{\tilde{v}v}$  ( $\delta_{jk}$  is the Kronecker delta function). Multiplying through by  $x^\ell y^n z^v$  and re-indexing yields an infinite-dimensional system of nonlinear ordinary differential equations for the conditional probability generating function

$$G_{\tilde{\ell}, \tilde{n}, \tilde{v}}(x, y, z; t) = E[x^{L(t)} y^{T^*(t)} z^{V(t)} | L(0) = \tilde{\ell},$$

$$T^*(0) = \tilde{n}, V(0) = \tilde{v}] = \sum_{\ell=0}^{\infty} \sum_{n=0}^{\infty} \sum_{v=0}^{\infty} P_{\tilde{\ell}, \tilde{n}, \tilde{v}; \ell, n, v}(t) x^\ell y^n z^v.$$

We reduce the infinite dimensional system to a system of three equations by exploiting the branching property  $G_{\tilde{\ell}, \tilde{n}, \tilde{v}} = (G_{100})^{\tilde{\ell}} (G_{010})^{\tilde{n}} (G_{001})^{\tilde{v}}$  [46],

$$\frac{dG_{100}(t)}{dt} = a(G_{010} - G_{100}) + \mu(1 - G_{100}) + \rho(G_{100}^2 - G_{100})$$

$$\frac{dG_{010}(t)}{dt} = \delta(1 - G_{010}) + p(G_{010}G_{001} - G_{010})$$

$$\frac{dG_{001}(t)}{dt} = c(1 - G_{001}) + f(1 - \epsilon)kT_S(G_{100} - G_{001}) + (1 - f)(1 - \epsilon)kT_S(G_{010} - G_{001})$$

with initial conditions  $G_{100}(0) = x, G_{010}(0) = y$  and  $G_{001}(0) = z$ . To our knowledge we cannot solve this nonlinear system analytically. Therefore to calculate  $G_{\ell\bar{n}\bar{v}}$ , we solve numerically using a standard differential equation integrator. Once  $G_{100}(t), G_{010}(t)$  and  $G_{001}(t)$  are calculated we can compute the full probability generating function, accounting for the initial distributions

$$G(x, y, z; t) = E[x^{L(t)}y^{T^*(t)}z^{V(t)}] = \sum_{\ell=0}^{\infty} \sum_{\bar{n}=0}^{\infty} \sum_{\bar{v}=0}^{\infty} P_{\ell, \bar{n}, \bar{v}}(0)(G_{100})^{\ell}(G_{010})^{\bar{n}}(G_{001})^{\bar{v}}.$$

Our goal is to recover the probability distributions of latently infected cells, productively infected cells, and virions at times  $t$ . These can be recovered from the probability generating function  $G_{\ell, \bar{n}, \bar{v}}$  by taking derivatives. For example, the probability that there are  $v$  virions at time  $t$  is given by

$$P(V = v, t) = \frac{1}{v!} \left. \frac{d^v G}{dz^v} \right|_{x=y=z=0}.$$

Because the distributions do not scale, we must perform computations over the total number of virions  $V$ . Given a mean viral load of 25 copies/ml (henceforth abbreviated as 5  $c$ /mL) within 5L of total blood volume, we must compute 125000 derivatives to get  $P(V = 25c/\text{mL}, t)$ ! Direct numerical differentiation would be difficult, so we exploit the Cauchy-Euler formula:

$$\left. \frac{d^n f}{dx^n} \right|_{x=a} = \frac{n!}{2\pi i} \oint_C \frac{f(w)}{(w-a)^{n+1}} dw,$$

where  $C$  is a closed curve in complex space which contains  $a$ , and  $f$  is analytic on a simply connected domain containing  $C$ . The probability generating function  $G(x, y, z; t)$  is a polynomial in  $x, y$ , and  $z$  and therefore satisfies the analyticity requirement. We want to evaluate integrals at  $z=0$ , so our contour  $C$  must contain the origin and it is simplest to use the unit circle,  $w = e^{i\theta}$ ,  $0 \leq \theta \leq 2\pi$ . Then

$$\begin{aligned} P(V = v, t) &= \frac{1}{v!} \left( \frac{v!}{2\pi i} \oint_C \frac{G(1, 1, w)}{w^{v+1}} dw \right) \\ &= \frac{1}{2\pi} \int_0^{2\pi} G(1, 1, e^{i\theta}) e^{-iv\theta} d\theta \\ &= \frac{1}{\pi} \text{Re} \left[ \int_0^{\pi} G(1, 1, e^{i\theta}) e^{-iv\theta} d\theta \right] \end{aligned}$$

where we have used the fact that  $G(1, 1, e^{i(2\pi-\theta)}) = G_{\ell, \bar{n}, \bar{v}}(1, 1, e^{i\theta})^*$ , where  $*$  indicates complex conjugate. By this method we can calculate our probabilities via straightforward and reliable numerical integration. The same approach can be used to compute joint probability distributions.

We can also use this formulation to directly calculate cumulative probabilities. As  $P(V \leq N, t) = \sum_{v=0}^N P(V = v, t)$ , we can write

$$P(V \leq N, t) = \frac{1}{\pi} \text{Re} \left[ \int_0^{\pi} G(1, 1, e^{i\theta}) \frac{1 - e^{-i(N+1)\theta}}{1 - e^{-i\theta}} d\theta \right]$$

by interchanging the order of integration and summation. This final formula will be useful in calculating blip probabilities at a time  $t$ ,  $P(V > 50; t) = 1 - P(V \leq 50, t)$ .

To our knowledge this is a novel method for computing probability distributions from single- or multi-type continuous time branching processes. We thoroughly tested our method and its implementation; see Figure S1 for comparisons with Gillespie simulations.

### Extinction probabilities

We also wish to calculate the distribution of times to extinction for the latent reservoir. We choose parameters so that the probability of extinction of the latent reservoir is 1 as  $t \rightarrow \infty$ . However, as clearing the latent reservoir is considered a major hurdle in clearing HIV, the distribution of times to this inevitable extinction is of interest. We obtain the cumulative probability of latent reservoir extinction directly from the probability generating function. Since  $G(x, y, z; t) = \sum_{\ell=0}^{\infty} \sum_{\bar{n}=0}^{\infty} \sum_{\bar{v}=0}^{\infty} P_{\ell, \bar{n}, \bar{v}}(t) x^{\ell} y^{\bar{n}} z^{\bar{v}}$ ,

$$P_{ext}(t) = P(L = 0, t) = G(0, 1, 1; t).$$

Note that the marginal probability  $P(L = \ell, t) = G(x, 1, 1; t)$ . We then find the probability distribution of extinction times by differentiating,

$$p_{ext}(t) = \frac{d}{dt} P_{ext}(t).$$

If we assume that no newly infected cells become latently infected ( $f = 0$ ) or that treatment is completely effective ( $\epsilon = 1$ ), we can obtain the extinction probability analytically. In this case, the latent cell dynamics decouples from the rest of the model and can be represented as a pure birth-and-death process with master equation

$$L'_{\ell} = (a + \mu)((\ell + 1)L_{\ell+1} - \ell L_{\ell}) + \rho((\ell - 1)L_{\ell-1} - \ell L_{\ell}), \quad (3)$$

where  $L_{\ell}(t)$  is the probability that at time  $t$  there are  $\ell$  latently infected cells. This probability has the conditional probability generating function

$$G_{\tilde{\ell}}(x; t) = \sum_{\ell=0}^{\infty} L_{\ell}(t) x^{\ell} = \left[ \frac{(a + \mu)(1 - x) + e^{-(\rho - a - \mu)t}(\rho x - (a + \mu))}{\rho(1 - x) + e^{-(\rho - a - \mu)t}(\rho x - (a + \mu))} \right]^{\tilde{\ell}}$$

where  $\tilde{\ell}$  is the initial reservoir size. The cumulative distribution is  $P_{ext}(t)|_{L(0)=\tilde{\ell}} = G_{\tilde{\ell}}(0; t)$  and thus we obtain

$$\begin{aligned} P_{ext}(t)|_{L(0)=\tilde{\ell}} &= \frac{\tilde{\ell} e^{(\rho - a - \mu)t} (\rho - a - \mu)^2}{(a + \mu)(e^{(\rho - a - \mu)t} - 1)(\rho / (a + \mu) e^{(\rho - a - \mu)t} - 1)} \\ &\quad \left[ \frac{e^{(\rho - a - \mu)t} - 1}{\rho / (a + \mu) e^{(\rho - a - \mu)t} - 1} \right]^{\tilde{\ell}}. \end{aligned} \quad (4)$$

Then given the initial distribution on the latent reservoir  $p_{\tilde{\ell}}(0)$  we have can compute the extinction probability  $p_{ext}(t) = \sum_{\ell=0}^{\infty} p_{ext}(t)|_{L(0)=\tilde{\ell}} p_{\tilde{\ell}}(0)$ . We use the analytic expression to compute latent

reservoir extinction time distributions for  $f = 0$  or  $\epsilon = 1$ . Otherwise, we work numerically.

**Parameter estimation**

The parameters used for simulation results presented below are given in Table 1. In our simulations the parameters  $\mu$ ,  $\epsilon$  and  $T_S$  are chosen based on estimates from [36] and  $k$  based on estimates from [47]. The decay rate of the latent reservoir is chosen so that its half-life is  $t_{1/2} = 60$  months, as measured in patients exhibiting viral blips [21]. For  $\delta$ , the death rate of productively infected cells, and  $c$  the virion clearance rate we set to estimates from [48] ( $\delta = 1 \text{ day}^{-1}$ ) and [49] ( $c = 23 \text{ day}^{-1}$ ), respectively. The *in vivo* viral production rate  $p$  is not well established and therefore we will consider a range for this parameter. The fraction  $f$  of new viral infections that result in latency is also hard to estimate, but given the small size of the latent reservoir, it is likely to be rather small. For simplicity, we choose a baseline value of  $f = 0$ . We choose the initial mean latent reservoir size  $L_0$  consistent with the experimental estimates [21,50].

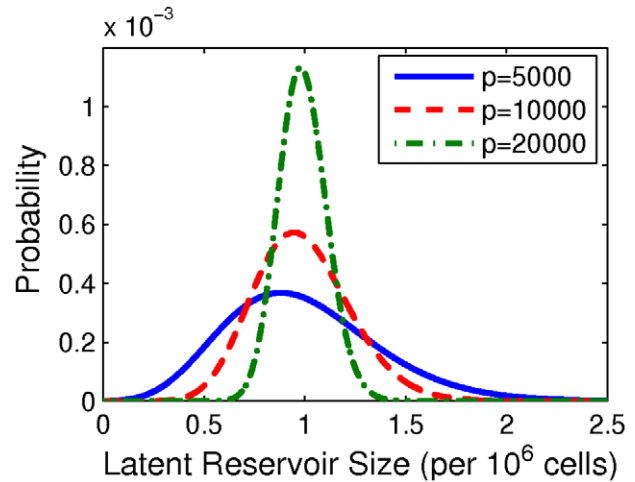
The activation rate  $a$  and replication rate  $\rho$  of latently infected cells remain unknown. We calculate values from the mean behaviour equations (1), taking  $f = 0$ . Since the dynamics of the productively infected cells and virus are more rapid than those of the latent reservoir, we can make a quasi-steady approximation to find  $T_0^* \simeq (c + (1 - \epsilon)kT_S)V_0/p$ . Then for an initial latent reservoir size of  $L_0$ ,  $a \simeq (\delta T_0^* - (1 - \epsilon)kT_S V_0)/L_0$ . To calculate  $\rho$ , we choose the mean decay rate so that the half-life of the latent reservoir is  $t_{1/2} = 60$  months. Thus we can set  $\rho = (-\log(2)/t_{1/2}) + a + \mu$ . The resulting  $(a, \rho)$  values for each production rate  $p$  and mean viral loads of 25c/mL or 35c/mL are given in Tables 2 and 3 respectively.

As noted above, we must perform calculations over the entire blood volume, which we take to be 5L. When presenting results below we report viral loads in copies per mL, as this is the standard measurement, but they are always obtained by re-scaling the axes for results over the entire blood volume.

**Table 1.** Baseline parameter values for latent cell activation model.

Parameter	Description	Estimate
$\mu$	Death rate of latently infected cells	$0.01 \text{ day}^{-1}$
$p$	Virion production rate for a productively infected cell	5000, 10000, 20000 $\text{day}^{-1}$
$\delta$	Death rate of productively infected cells	$1 \text{ day}^{-1}$
$c$	Clearance rate of free virions	$23 \text{ day}^{-1}$
$k$	Mass-action infectivity of free virions	$2.4 \times 10^{-5} \mu\text{L copy}^{-1} \text{ day}^{-1}$
$T_S$	Target cells	$595 \mu\text{L}^{-1}$
$\epsilon$	Drug efficacy	0.93
$f$	Fraction of new infections that result in latency	0
$t_{1/2}$	Half life of latent reservoir	60 months
$L_0$	Initial number of latently infected cells	1 per $10^6$ target cells

doi:10.1371/journal.pcbi.1002033.t001



**Figure 2.** Initial probability distribution on latent reservoir size.

We take an initial mean viral load of 25 c/mL and parameters given in Tables 1 and 2. Production rates  $p$  have units  $\text{day}^{-1}$ . Initial distributions assuming initial mean viral load of 35 c/mL (Table 3) are qualitatively similar (not shown).

doi:10.1371/journal.pcbi.1002033.g002

**Initial distributions**

In order to correctly simulate viral blips and latent reservoir extinction in patients with established treated infection, we should carefully choose the initial joint distribution  $p_{\vec{L}, \vec{n}, \vec{v}}$  so that it is close to the (moving) equilibrium of the ongoing dynamics. Otherwise, transient effects will pollute our results. In the mean, the dynamics of the latent reservoir are very slow compared to those of the productively infected cells or virions. We therefore focus on getting the initial latent reservoir distribution correct since errors in the other two compartments will resolve themselves quickly. Indeed, for a constant latent reservoir size, and our parameters, the distributions on  $V$  and  $T^*$  converge to stationary distributions in less than a month (results not shown).

In order to calculate a reasonable initial latent reservoir distribution we isolate its dynamics and consider the marginal probability distribution only, as in equation (3). We choose the marginal latent reservoir probability distribution at time  $\tau$  such that the variance is maximized. We reason that transient dynamics on the latent reservoir are dominated by the spreading of the distribution about the decaying mean, and that at maximum variance the probabilities are sufficiently spread for our purposes. For birth-and-death processes maximum variance occurs at the half life  $t_{1/2}$ . Therefore, in order to create the initial distribution on the latent reservoir  $p_{\vec{L}}^0$ , we solve (3) out to  $t_{1/2} = 60$  months, starting

**Table 2.** Calculated activation and replication rates for initial mean viral load of 25 c/mL.

Production rate $p$	Activation rate $a$	Replication rate $\rho$
5000 $\text{day}^{-1}$	$0.1513 \text{ day}^{-1}$	$0.1609 \text{ day}^{-1}$
10000 $\text{day}^{-1}$	$0.0546 \text{ day}^{-1}$	$0.0643 \text{ day}^{-1}$
20000 $\text{day}^{-1}$	$0.0063 \text{ day}^{-1}$	$0.0159 \text{ day}^{-1}$

We calculate the activation rates  $a$  for an initial mean viral load of 25 c/mL, given the virion production rate  $p$ . The replication rate  $\rho$  is then chosen so that the half-life of the latent reservoir is 60 months.

doi:10.1371/journal.pcbi.1002033.t002

with  $2L_0$  latently infected cells, where  $L_0$  is the desired mean latent reservoir size. The resulting distributions for different parameter sets are shown in Figure 2. Notice that results based on a virus production rate  $p = 5000\text{day}^{-1}$  have larger standard deviation. This is because lower production rates are associated with higher activation rates  $a$  (cf. Tables 2 and 3). The higher activation rate speeds the dynamics of the latent reservoir, increasing the spread of its probability distribution function. Finally, we combine the computed initial latent cell distribution with single initial numbers of productively infected cells and virus, to obtain the whole initial joint probability distribution:

$$p_{\ell,n,v}(t) = \sum_{\ell=0}^{\infty} = p_{\ell}^0 \delta_{\ell m} \delta_{v \bar{v}}$$

## Results

### Latent reservoir extinction

The reservoir of latently infected cells is considered a major obstacle to clearing HIV infection [19]. Within our model, when the reservoir goes extinct, viral load quickly goes to zero, since ongoing viral replication is too small to sustain the virus population. We are therefore interested in examining the reservoir lifetime after the onset of ART. To do this, we extend our approach to find the probability of reservoir extinction over time. We examine the reservoir lifetime using baseline parameters ( $f = 0$ ), and then allowing for the possibility of latent reservoir replenishment ( $f \neq 0$ ). Furthermore, since anti-retroviral treatments have improved substantially over the last 15 years, we also examine how the reservoir lifetime behaves as drug efficacy improves ( $\epsilon \rightarrow 1$ ).

**Latent reservoir extinction in the absence of re-seeding.** To begin, we use the baseline parameters (Tables 1 and 2). Since  $f = 0$ , we can use an analytic expression (4). Probability distribution functions for latent reservoir extinction are shown in Figure 3 for each of  $p = 5000, 10000$  and  $20000\text{day}^{-1}$ . Note that although the production rate  $p$  is not explicitly included in (4), the choice of  $p$  affects the activation and replication rates  $a$  and  $\rho$  (see Table 2 and the Methods). Figure 3 shows that the resulting distributions are asymmetric and leaning towards longer lifetimes, which is a characteristic of subcritical birth-and-death processes. We also observe that as the production rate  $p$  increases (so the activation rate  $a$  decreases), the distributions shift to the right, with increasing mean and variance. Exact means and variances can be calculated by integrating (4) over time: for  $p = 5000\text{day}^{-1}$  our model predicts a mean reservoir lifetime of 18.7 years with a standard deviation of 8.9 years; for

$p = 20000\text{day}^{-1}$ , it predicts a mean reservoir lifetime of 34.8 years with a standard deviation of 9.2 years.

At first these results seem surprising: since the reservoir half-life and size are identical in all three calculations, how can the time-to-extinction distributions differ so dramatically? Usually, in a deterministic framework, one would calculate the mean time to extinction assuming exponential decay with extinction when the mean falls below a small threshold. Taking this threshold to be a single cell, it is easy to calculate  $t_{\text{extinct}} = t_{1/2} \log(L_0)/\log(2) \approx 58\text{years}$ . However, the addition of activation and replication events complicates the picture and significantly changes the dynamics once the reservoir gets small. More specifically, the time to extinction decreases as the activation rate  $a$  increases, and  $a$  is inversely correlated with production rate  $p$  in our model.

However, beyond model- and parameter-specific results, we find an interesting insight: if we wish to predict the timescale for latent reservoir extinction, we must refine our understanding of latent reservoir dynamics beyond half-life estimates to include accurate estimates of birth, death and division, across the various populations that make up the latent reservoir.

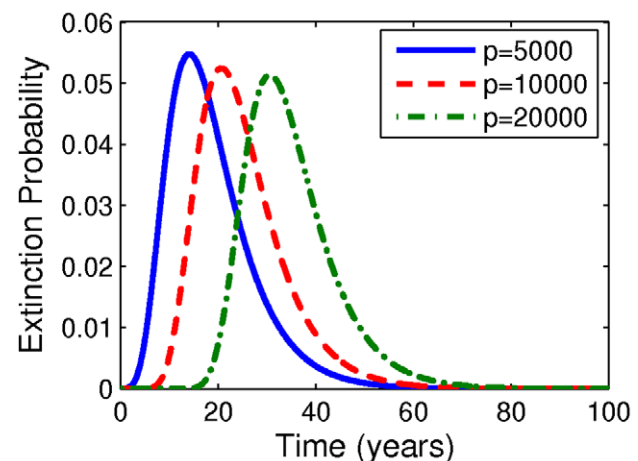
**Latent reservoir extinction with latent cell re-seeding.** So far we have assumed that the fraction  $f$  of new infections resulting in latency is zero. Under that assumption, the decay of latent cells is independent of the rest of the model, and in particular, of drug efficacy. However, intensification of drug treatment has been reported to speed latent reservoir decay [51]. To examine this effect, we now consider non-zero  $f$  so that latent reservoir replenishment is now in part through infection of new cells (Figure 1). As discussed in the Methods, the overall decay rate of the latent reservoir is kept the same via small corrections (less than two percent) in the replication rate  $\rho$ .

With improving drug efficacy (as  $\epsilon \rightarrow 1$ ) we predict that the mean latent reservoir lifetime decreases, as shown in Figure 4A for the parameters associated with  $p = 20000\text{day}^{-1}$ . Results for other values of  $p$  are very similar (not shown). The decrease is less than 1% for  $f = 5 \times 10^{-5}$  but more dramatic for  $f = 5 \times 10^{-3}$ , at almost 25%. This reduction is as expected from the model:  $f$  controls the contribution of newly infected cells back to the latent reservoir. But as the drug efficacy is improved, reservoir replenishment is decreased, and the latent reservoir goes extinct more quickly. When  $f$  is larger, the contribution of newly infected cells to latent

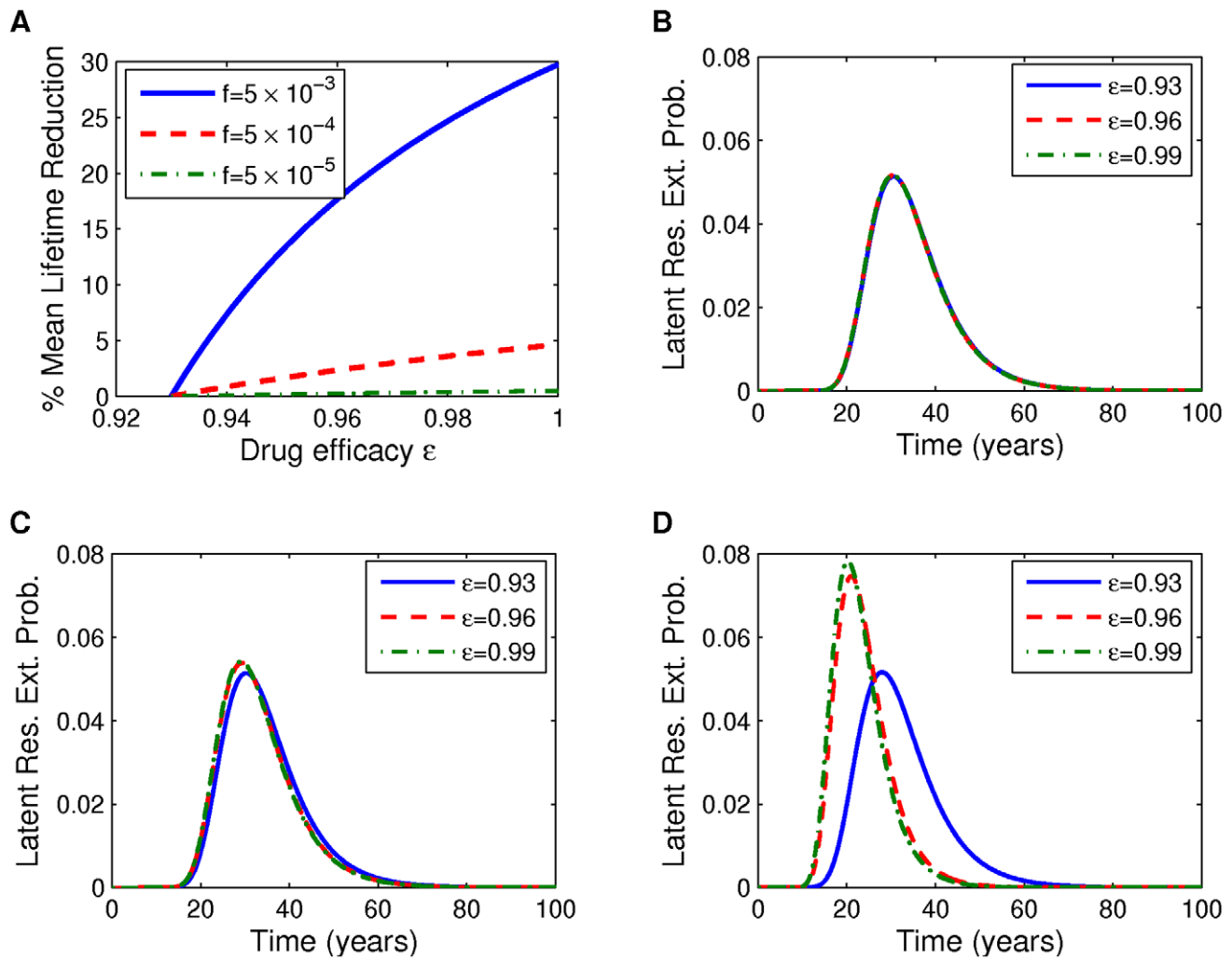
**Table 3.** Calculated activation and replication rates for initial mean viral load of 35 c/mL.

Production rate $p$	Activation rate $a$	Replication rate $\rho$
5000 $\text{day}^{-1}$	0.2118 $\text{day}^{-1}$	0.2214 $\text{day}^{-1}$
10000 $\text{day}^{-1}$	0.0765 $\text{day}^{-1}$	0.0861 $\text{day}^{-1}$
20000 $\text{day}^{-1}$	0.0088 $\text{day}^{-1}$	0.0185 $\text{day}^{-1}$

We calculate the activation rates  $a$  for an initial mean viral load of 35 c/mL, given the virion production rate  $p$ . The replication rate  $\rho$  is then chosen so that the half-life of the latent reservoir is 60 months.  
doi:10.1371/journal.pcbi.1002033.t003



**Figure 3.** Latent reservoir extinction probability over time. Parameters given in Tables 1 and 2. Production rates  $p$  have units  $\text{day}^{-1}$ .  
doi:10.1371/journal.pcbi.1002033.g003



**Figure 4. Reductions in latent reservoir lifetime with improving drug efficacy.** (A) Percent mean reduction in latent reservoir lifetime with improving drug efficacy  $\epsilon$ . (B–D) Corresponding latent reservoir extinction distributions with improving drug efficacy for fraction  $f$  of newly infected cells becoming latently infected (B)  $f = 5 \times 10^{-5}$ , (C)  $f = 5 \times 10^{-4}$ , (D)  $f = 5 \times 10^{-3}$ . Parameters: Tables 1 and 2 with  $p = 20000\text{day}^{-1}$ . doi:10.1371/journal.pcbi.1002033.g004

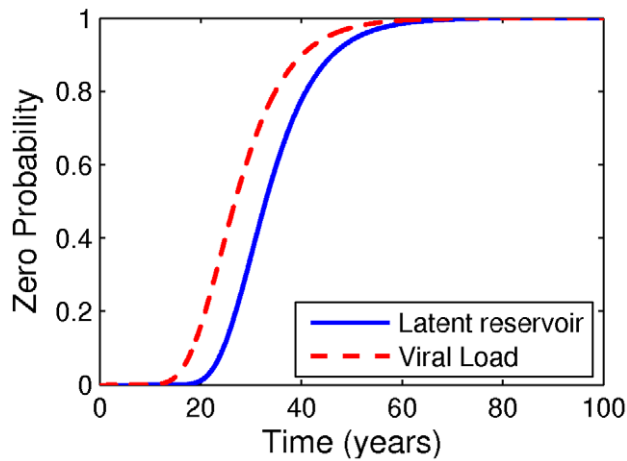
reservoir replenishment is larger. Therefore cutting off this supply results in a more significant reduction in reservoir lifetime.

Furthermore, as the drug efficacy increases ( $\epsilon \rightarrow 1$ ), the latent reservoir extinction time distributions narrow, so that the variance decreases and the asymmetric tail shrinks. This effect is illustrated for  $p = 20000\text{day}^{-1}$  in Figure 4B–D (results for other values of  $p$  are qualitatively similar, not shown). This is not unexpected: with improving drug efficacy, reservoir replenishment/birth is diminished, which is the source of the asymmetric tail of the distribution. Assuming our model is correct, this is a moderately encouraging result: with improving drug efficacy the range of possible lifetimes is reduced. Nonetheless, in our model we still find latent cell clearance only after decades of drug treatment.

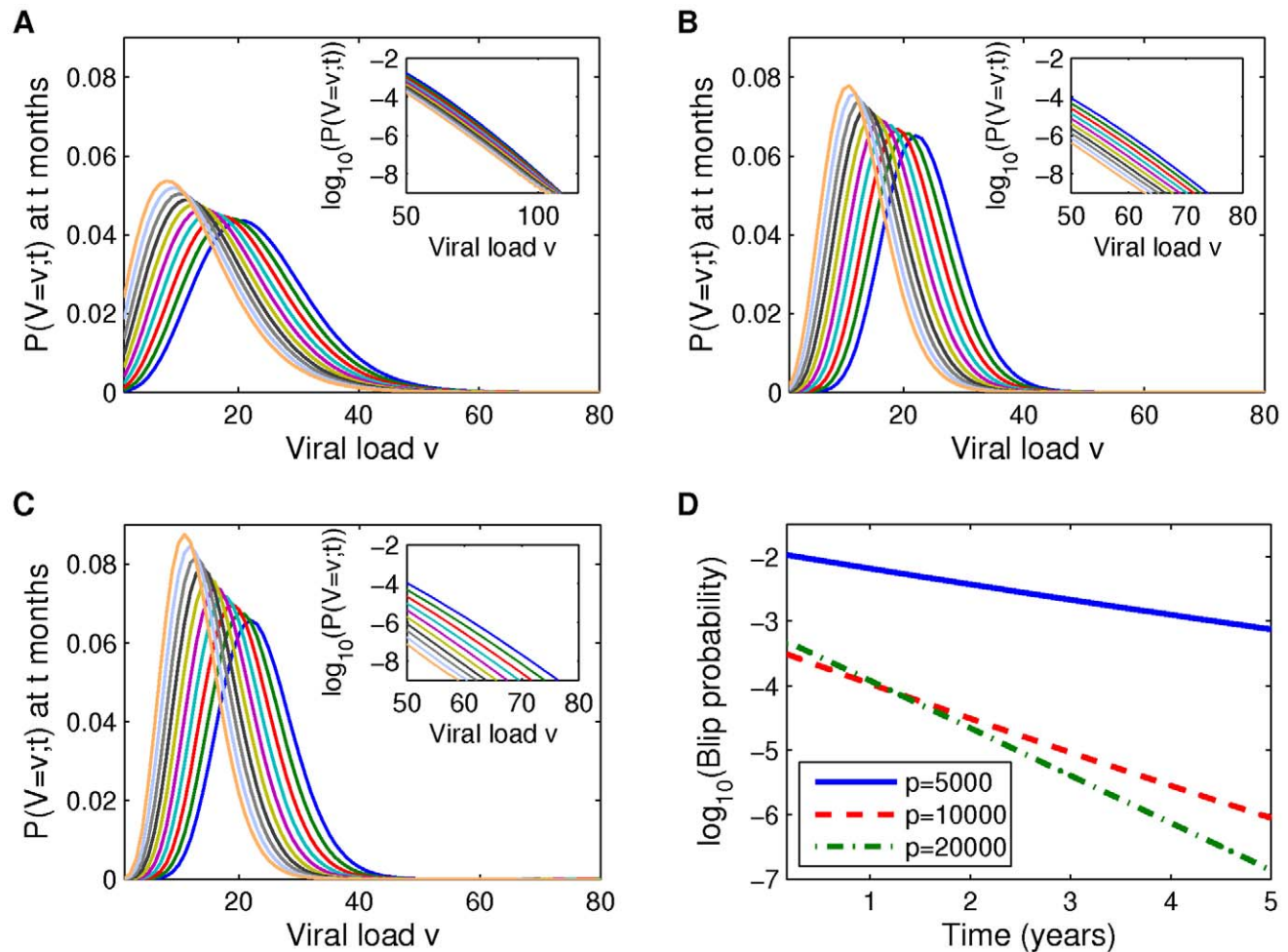
**Transient loss of free virus.** As described above, the timescale of true viral clearance in our model is set by the decline of the latent reservoir. However, it is possible to transiently achieve  $V = 0$  before the latent reservoir has disappeared altogether. In Figure 5 we plot the cumulative distribution function for the first occurrence of this event. Latent reservoir extinction lags transient viral clearance by approximately 6 years, with a mean time of 32 years, for these parameters.

### Viral load distributions in treated patients

We now focus on the time evolution of viral load and the likelihood of small-amplitude viral blips. We interpret viral load above the threshold of detection of 50 c/mL as a viral blip. Note that unless otherwise specified, the following calculations and computations assume the fraction  $f$  of newly infected cells that become latently infected is 0. In Figures 6 and 7, we plot full viral load distributions over time, assuming initial mean viral loads of 25 c/mL and 35 c/mL. As time advances, the mean viral load decreases as expected in all cases but the viral load distributions widen more significantly when  $p$  is smaller (e.g. distributions in Figure 6A are widest, those in Figure 6C are narrowest). This is because the lower values of  $p$  are associated with higher values of  $a$  and  $\rho$ , and the resulting dynamics on the latent reservoir cause the latent reservoir size probability distribution (not shown) to be wider. As a consequence the associated viral load distributions are wider, and this also causes higher blip amplitudes. This effect is more clearly understood by examining the insets in Figures 6 and 7, which represents a magnified view of the given probability distribution curves above the blip threshold (50 c/mL), using a log scale to more clearly distinguish the curves. We observe that viral



**Figure 5. Transient and permanent viral extinction.** We plot the probabilities that the viral load is zero (transient viral extinction) and that the latent reservoir is zero (permanent viral extinction) as a function of time. Parameters: Tables 1 and 2, with  $p=20000\text{day}^{-1}$ . doi:10.1371/journal.pcbi.1002033.g005



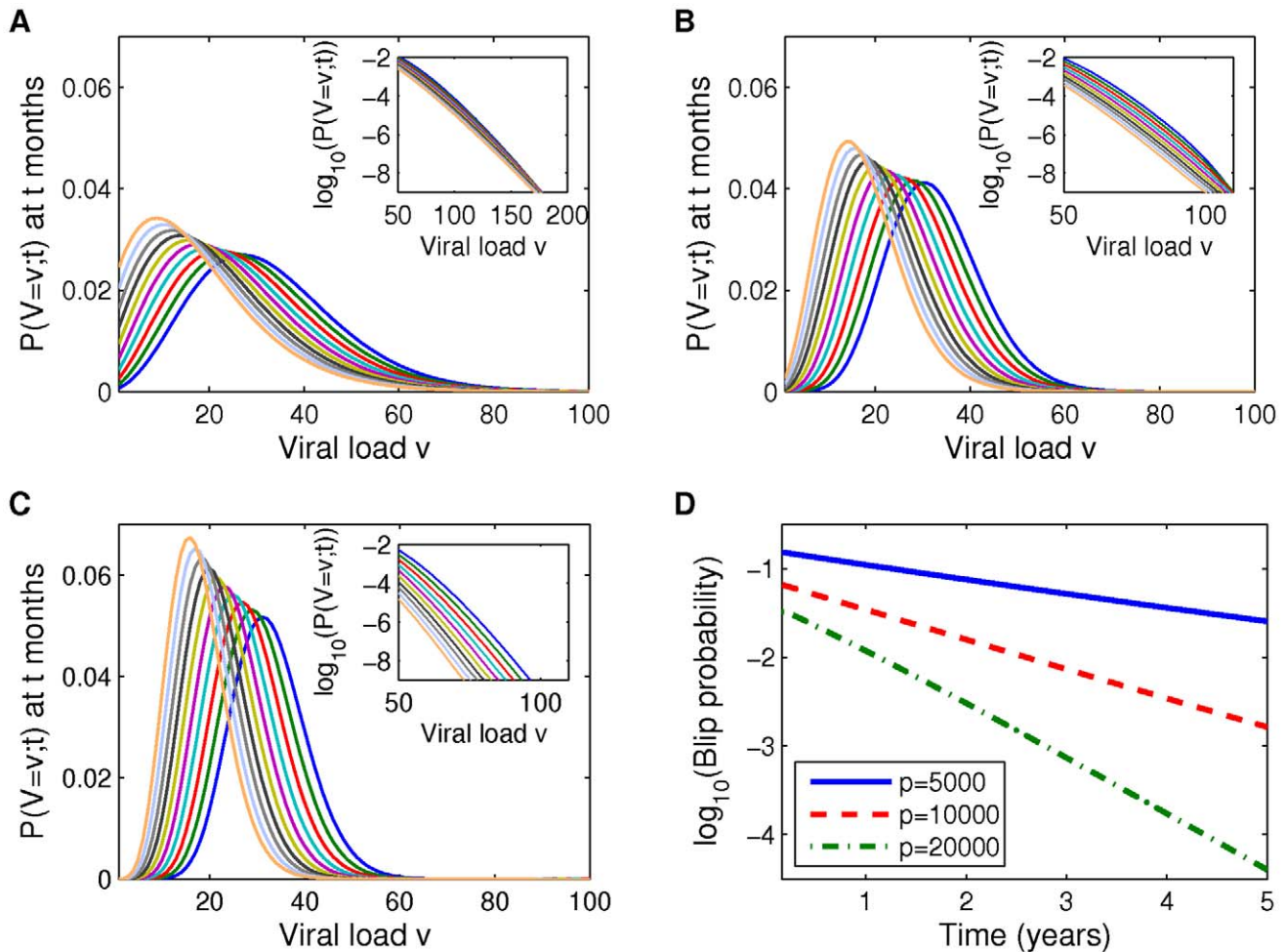
**Figure 6. Viral load probability distributions for initial mean viral load of 25 c/mL.** (A–C) Distribution functions are plotted at 6 month intervals for parameters given in Tables 1 and 2, and (A)  $p=5000\text{day}^{-1}$ , (B)  $p=10000\text{day}^{-1}$ , (C)  $p=20000\text{day}^{-1}$ . Insets: enlargement of probability distribution curves above the detection level,  $v=50\text{c/mL}$ ; a log scale is used to better distinguish the curves. As time advances the distributions move from right to left. (D) Blip probability plotted against time. The curves in (D) are computed by integrating the probability density functions from (A–C) over viral loads exceeding 50 c/mL. doi:10.1371/journal.pcbi.1002033.g006

blips occur with very small probability regardless of the production rate  $p$ . The blip amplitudes vary between parameter sets but remain approximately within the range of blips unassociated with clinical variables shown in [34], i.e. 50–100 c/mL. Over three years, the range of reasonably likely detectable viral loads decays slowly, but small blips remain possible throughout that time (Figure 6).

**Probability of detectable viremia declines exponentially as a function of time.** Figures 6D and 7D show the probability that an individual on ART has a detectable viral load at some given time  $t$ ,  $P(V > 50; t)$ . For all parameter sets, the probability of a blip declines exponentially over time. From the equations for mean viral load and latent reservoir size (1) we see that both are decaying exponentially with half life set by  $t_{1/2}=60\text{months}$ . However, blip probabilities decay much more quickly (with half-lives on the range of 6–18 months) than would be predicted from studying the mean behaviour of the system. This underlines the importance of taking a stochastic approach to predicting rare stochastic events.

As expected, in both cases the parameter set associated with the largest latent cell activation rate ( $p=5000\text{day}^{-1}$ ) yields the highest





**Figure 7. Viral load probability distributions for initial mean viral load of 35 c/mL.** (A–C) Distribution functions are plotted at 6 month intervals for parameters given in Tables 1 and 3, and (A)  $p=5000\text{day}^{-1}$ , (B)  $p=10000\text{day}^{-1}$ , (C)  $p=20000\text{day}^{-1}$ . Insets: enlargement of probability distribution curves above the detection level,  $v=50\text{c/mL}$ ; a log scale is used to better distinguish the curves. As time advances the distributions move from right to left. (D) Blip probability plotted against time. The curves in (D) are computed by integrating the probability density functions from (A–C) over viral loads exceeding 50 c/mL. doi:10.1371/journal.pcbi.1002033.g007

blip probability. Predictably, the likelihood of viral blips is substantially higher when the mean viral load is higher, and this is in agreement with a decline in blip detection as drug treatment has improved and reduced setpoint viral load. Our finding that blip probability declines over time disagrees with the report of Di Mascio et al. [31], where a constant rate of blips was observed. We attribute this difference to the fact that many of the blips considered by Di Mascio et al. were of large amplitude rather than the small blips we are examining here. Our basic modeling assumption is that small viral blips under consideration here represent large deviations from a small mean viral load, but that mechanisms for larger viral blips - not fully captured within our model - depend on external factors. Examples of these factors include target cell increase due to immune system activation in response to an unrelated infection, or treatment non-adherence; the frequency of such events is unlikely to change over time, which would result in a constant rate of viral blips over time.

**Impact of increasing drug efficacy.** With our baseline parameters, predicted viral loads are in the range of those measured in [3] but 3–5 times higher than viral loads measured recently [10,11]. Further, recent clinical observations reveal that

blips are now very rare. We can reproduce such observations in our model with improving drug efficacy  $\epsilon$ . We used a lower drug efficacy of  $\epsilon=0.93$  to reflect drug efficacy at a time when viral blips were observed. However as  $\epsilon \rightarrow 1$ , reflecting modern improvements in drug efficacy, we note that both the mean viral load and the variance decrease dramatically. For  $1-\epsilon=10^{-3}-10^{-5}$ , reasonable for current drug regimens [52], we compute (with baseline parameters, Tables 1 and 2) a mean initial viral load of 3–4 c/mL rather than 25 c/mL, and find blip probabilities to be smaller than we can calculate. This implies that, with excellent drug treatment and perfect adherence, any viral blips are almost certainly not due to stochastic reactivation of latent cells.

**Viral blips driven by secondary infection.** Nonetheless, blips continue to be observed in treated patients. One argument for their existence is poor compliance with the drug regimen, effectively reducing the drug efficacy averaged over time. A second possibility is that immune system activation due, for example, to transient secondary infection, can cause episodes of increased viremia. Specifically, during infection, the number of CD4 cells increases to fight infection; more target cells means more infected cells, producing more virus, increasing the viral load to a

detectable level [38,41]. Furthermore, memory T cells make up a part of the latent reservoir and infection may therefore induce an increased activation rate [19].

We experimented with adjusting our model to simulate a transient secondary infection. At initiation, we increase the activation rate  $a$  for 3 days and then return to the background value. We also increase the number of target cells  $T_h$  for 7 days, starting after 2 days. These parameter changes must be step-function-like because our method for solving the backwards equation only admits constant rates. Figure 8A shows the mean of maximum viral load for different activation rate and target cell number multipliers, using parameters  $p=20000\text{day}^{-1}$ ,  $f=5 \times 10^{-3}$  and  $\epsilon=0.999$ . We observe that the increase in target cells has the dominant effect on maximum viral load, although the increased activation rate does provide a boost. Figures 8B and C show the mean of maximum viral load  $\pm$  one standard deviation, as a function of activation rate multiplier (for target cell multiplier fixed at 100) and as a function of target cell multiplier (for activation rate multiplier fixed at 5), respectively. The variance is very small. These incomplete results show that for blips associated with immune system activation, the latent reservoir plays only a small role.

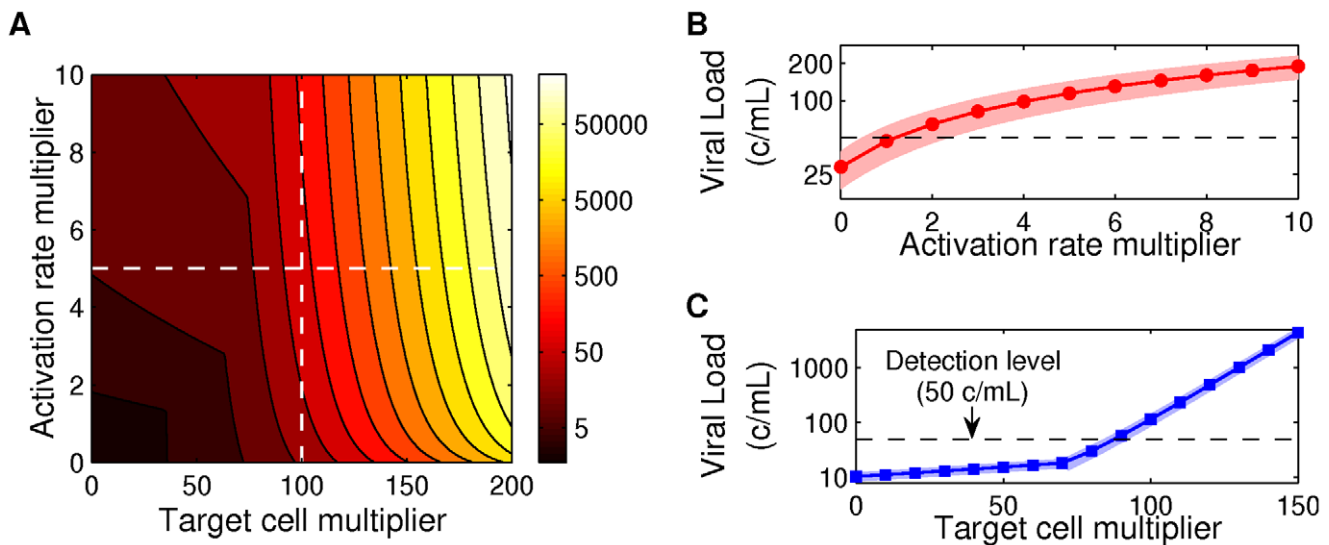
### Viral blip durations

In this section we consider the following question: given a blip, defined as a detectable viral load measurement, how long should we expect the viral load to remain above the threshold of detection? This question is of clinical interest, since a repeat measurement following a measurable viral load should be performed after enough time that a second positive result might have clinical significance, such as suggesting drug failure.

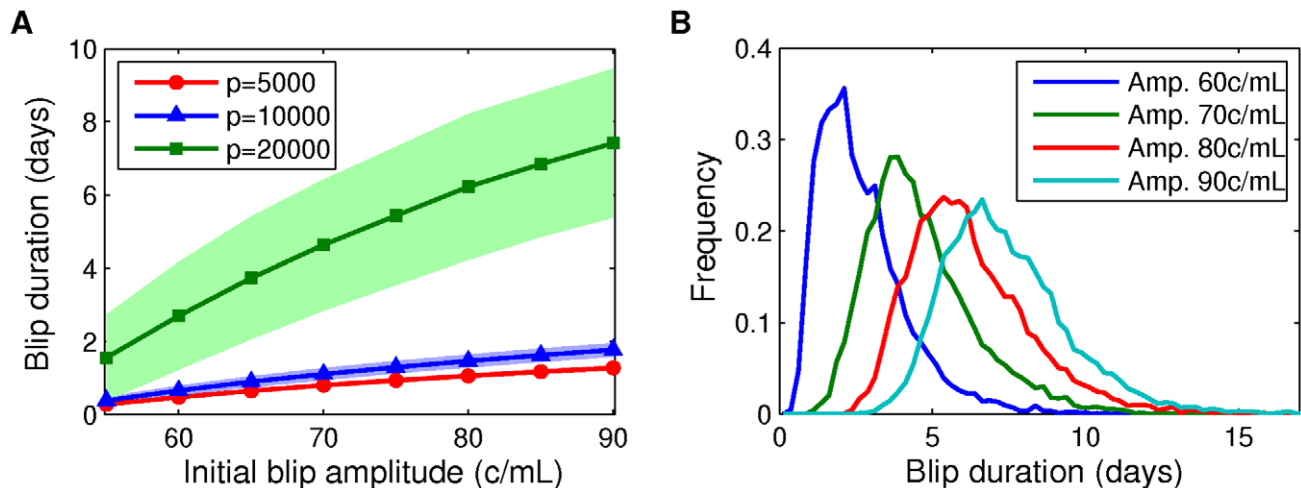
Different from the previous sections, all results in this section are computed via 10000 direct simulations of the branching process using the Gillespie algorithm, beginning with an initial “blip” condition. The initial conditions are chosen as follows: we set the latent reservoir size  $L(0)=L_0$  and viral load  $V(0)=V_0 > 50\text{c/mL}$ .

Since dynamics on the viral load  $V$  are so much faster than on the productively infected cells  $T^*$  (cf. Table 1), we then use a quasi-steady approximation to set the initial number of productively infected cells  $T^*(0)=(c+(1-\epsilon)kT_S)V_0/p$ .

**Blip duration dependence on initial viral load measurement.** We begin by calculating blip durations as a function of a detectable viral load measurement at time zero. Figure 9A shows the mean blip duration  $\pm$  1 standard deviation, over different values of the production rate  $p$ , with the latent reservoir size set to be 1 cell per  $10^6$ . As the initial measured blip amplitude increases, so do the mean and standard deviation of the blip duration. This is because the more productively infected cells there are, the longer it takes for enough of them to die and therefore reduce the viral load. Also, since mean duration is longer for larger blips we also expect a larger standard deviation, since there is more opportunity for variability. The mean duration increases with increasing production rate, since for higher  $p$  more virions are produced, and have the opportunity to infect healthy cells, before the productively infected cells die. Duration distributions for different initial detectable viral load measurements are shown in Figure 9B for  $p=20000\text{day}^{-1}$ . From this figure we see that the standard deviation alone does not fully determine blip duration variability. For smaller blip amplitudes the distribution is more asymmetric, with a relatively larger probability of longer blips (positive skew). As initial blip amplitude increases, the distribution becomes more symmetric. This observation is explained by noting that as the initial blip size is increased, we are moving the initial condition farther and farther from equilibrium. Therefore, the stochastic dynamics are increasingly driven by decay of  $T^*$  and  $V$  towards the equilibrium, and it becomes increasingly unlikely for the viral load to increase after the initial measurement. Direct observation of viral load evolution in different realizations of master equation simulations support this argument, as shown in Figures 10A and B. For example, after an initial measurement of amplitude 60 c/mL (Figure 10A), the viral load does not decay as rapidly as it does for a measurement of 80 c/mL (Figure 10B).



**Figure 8. Maximum viral load under immune system activation.** (A) Maximum mean viral load for different multiplicative increases in target cell populations  $T_S$  and activation rates  $a$ . Dashed lines indicate target cell multiplier 100 (vertical) and activation rate multiplier 5 (horizontal). (B) Maximum mean viral load (symbols)  $\pm$  one standard deviation (shaded area) depending on activation rate multiplier, for target cell multiplier 100 (along vertical line in (A)). (C) Maximum mean viral load (symbols)  $\pm$  one standard deviation (shaded area) depending on target cell multiplier, for activation rate multiplier 5 (along horizontal line in (A)). Parameters: Tables 1 and 2 with  $f=5 \times 10^{-3}$ ,  $p=20000\text{day}^{-1}$  and drug efficacy  $\epsilon=0.999$ . doi:10.1371/journal.pcbi.1002033.g008



**Figure 9. Blip durations depend on initial blip amplitude.** (A) Mean blip durations (symbols)  $\pm 1$  standard deviation (shaded area), computed over 10000 simulations, plotted as a function of the initial viral load measurement (initial blip amplitude). Production rates  $p$  have units  $\text{day}^{-1}$ . (B) Frequency plots of time distributions of detectable viral load given initial measurements of 60–90 c/mL, computed over 10000 simulations. Parameters: Tables 1 and 2; latent reservoir size 1 per  $10^6$  cells;  $p=20000\text{day}^{-1}$  in (B). doi:10.1371/journal.pcbi.1002033.g009

**Effect of latent reservoir size on blip duration.** We expect that larger reservoir sizes should be associated with longer blips, since a higher reservoir size is associated with a higher quasi-steady mean viral load. Our results confirm this expectation. Figure 11A shows the mean and standard deviation of an amplitude-60 c/mL blip increasing with the reservoir size, for latent reservoir sizes  $L_0$  between 0 and 1.5 cells per  $10^6$ . As the reservoir size nears 2 cells per  $10^6$  we anticipate that the duration gets very large: at this level, the associated quasi-steady viral load is above the detection threshold of 50 c/mL, and we must wait until the viral load decays naturally to a mean below that threshold. This is extremely unlikely in our model: the probability that the latent reservoir size reaches 2 cells per  $10^6$  is initially very small  $O(10^{-13})$  for  $p=20000\text{day}^{-1}$  (see Figure 2) and only decreases over time. Notice also in Figure 11A that, as before, larger production rates  $p$  result in blips with longer durations.

We plot duration distributions across different latent reservoir sizes in Figure 11B. Interestingly, although the mean duration increases with reservoir size as shown in Figure 10A, the peak of the distribution stays in the range of 2–3 days. Therefore our modeling suggests that regardless of reservoir size, given a viral load measurement of 60 c/mL, the viral load in most patients should drop below detection level after 2–3 days. In contrast to the results with increasing initial blip measurement, we observe that the asymmetry in blip duration distributions is increasing - the tail is getting heavier with increasing reservoir size. This is because larger reservoir sizes are associated with larger associated mean viral loads (see Figure 11A). For the smaller reservoir sizes, the initial viral load of 60 c/mL is further away from the associated mean and viral decline is therefore quicker. Again this is supported by direct observation of viral load evolution in different simulations (Figures 10A and C).

## Discussion

We have presented a simple but fully stochastic model of HIV viral dynamics in individuals on antiretroviral treatment, focusing in particular on the role of the latent reservoir. In our model we included dynamics of only three compartments: the numbers of latently infected cells  $L$ , productively infected cells  $T^*$ , and virions

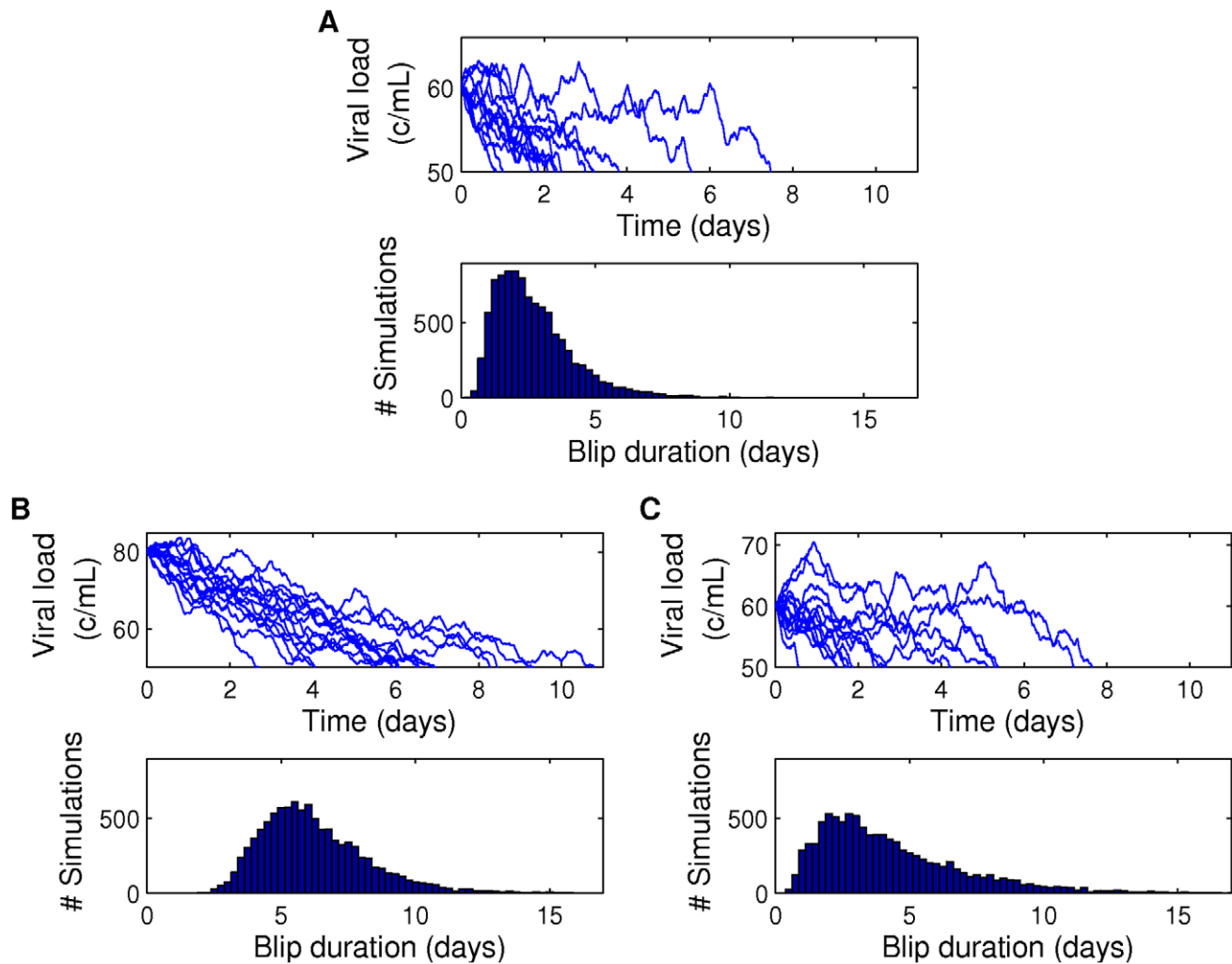
$V$ . We assumed that all rates correspond to exponentially-distributed transition probabilities and that therefore dynamics could be described by a continuous-time, multi-type branching process. We then derived equations for the probability generating function and using novel numerical techniques we computed the probability distributions on viral load over time, recovering features that are hard to study with approaches based on differential equations or direct simulation.

Our model reproduces interesting features of successfully treated infection, namely a usually low, undetectable viral load [3] and brief periods of low-amplitude detectable viral load, unassociated with clinical or demographic parameters, as discussed in [34]. This shows that the hypothesis that random activation of latent cells plays a major role in residual viremia on treatment, as has been suggested by clinical evidence (e.g. [9]), is reasonable and is compatible with reasonable parameter estimates. We were also able to use our model to look at the slow decline of the latent reservoir itself.

## Latent reservoir extinction

Clinical results on latent reservoir decay (e.g. [21]) make predictions on latent reservoir lifetimes that are based on purely exponential decay. Our model results showed that, for the same mean decay rate, the time distribution - and the mean time to extinction - is sensitive to dynamics on the latent reservoir. Further, assuming some reservoir replenishment due to latency in newly infected cells, our model predicted only limited lifetime reduction associated with improving drug efficacy. Eradication of the latent reservoir is considered a major hurdle in eradicating HIV infection [19,23], and these results demonstrate the importance of understanding the underlying dynamics on the latent reservoir: the reservoir half-life is only a small part of the equation.

From a clinical point of view our results on the latent reservoir lifetime are quite depressing - in our model we essentially study perfectly drug-adherent patients, and even with perfect drugs, decades of drug treatment are needed to clear the latent cell reservoir. Our model predicts that drug treatments that increase the activation rate of latently infected cells should reduce the lifetime of the reservoir. This approach has been tried several



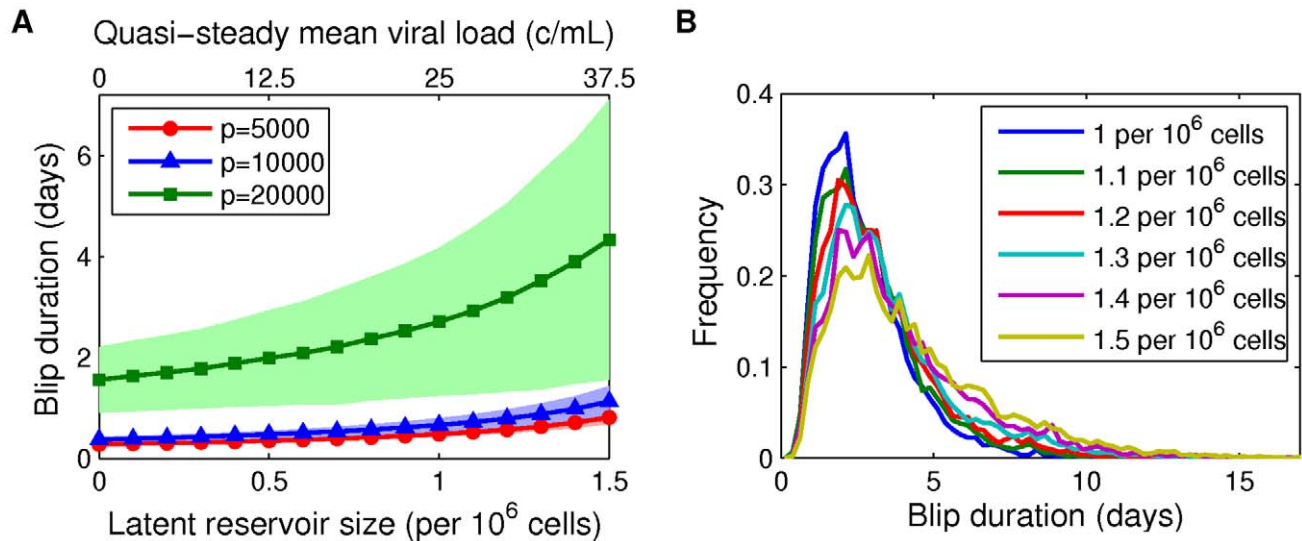
**Figure 10. Realizations of Gillespie simulations showing viral load evolution.** (A–C) show sample viral load evolutions and the associated histogram of durations until the viral load is below 50, over 10000 simulations, given an initial viral load measurement and latent reservoir size. (A) Initial viral load measurement of 60 c/mL with latent reservoir size 1 per  $10^6$  cells; (B) initial viral load measurement of 80 c/mL with latent reservoir size 1 per  $10^6$  cells; (C) initial viral load measurement of 60 c/mL with latent reservoir size 1.5 per  $10^6$  cells. Parameters: Tables 1 and 2, for  $p=20000\text{day}^{-1}$ .  
doi:10.1371/journal.pcbi.1002033.g010

times but so far without real success (reviews in [1,53]). A candidate drug would need to work on the whole heterogeneous population of cells seeded during initial infection, and in particular on the longest-lived subpopulation. We would predict that early treatment with such a drug, along with aggressive ART, would be most likely to reduce the size of the latent reservoir. This is in line with current research in treating HIV infection earlier, to enhance survival on an individual level [54] and limit transmission on the population level [55]. Of concern with earlier treatment is the possibility of emergent drug resistance (DR). In future work we plan to expand our stochastic model to examine the likelihood of different mechanisms of acquired DR in patients on treatment, such as mutation during ongoing viral replication and activation of a cell latently infected with a DR strain [56].

#### Viral blip frequency over time

We also examined the evolution of viral load over time, finding that as time progresses, the viral load distributions become more asymmetric, with a long tail towards higher viral loads. This can

be explained by viewing our model as an extended subcritical birth-and-death process. Such processes produce asymmetric distributions (see Figures 6 and 7). The asymmetry is more pronounced for smaller production rates ( $p=5000\text{virions/day}$  vs  $p=20000\text{virions/day}$ ), associated with larger activation rates, and for larger initial viral loads (35c/mL vs 25c/mL). When examining blip probability we found that our model predicts that these probabilities decay exponentially over time. This decay is more dramatic for larger production rates, associated with smaller activation ('birth') rates, and smaller initial mean viral loads. In only one case is the decay so slow ( $p=5000\text{virions/day}$  with mean initial viral load 35c/mL) that our model predictions are broadly consistent with previous observations that blip probabilities don't decay over time [45]. We also observed that blip probabilities show great sensitivity to model parameters, varying by orders of magnitude. Given sufficient high-quality data on blips (which does not currently exist), these would be the ideal results to compare with data for the purpose of parameter fitting, in order to gain some insight into latent reservoir dynamics - into the activation rate  $a$ , for example.



**Figure 11. Blip durations depend on initial latent reservoir size.** Mean blip durations (symbols)  $\pm 1$  standard deviation (shaded area), computed over 10000 simulations, plotted as a function of the initial latent reservoir size. (B) Frequency plots of time distributions of detectable viral load given initial latent reservoir sizes of 1–1.5 cells per  $10^6$ . Parameters: Tables 1 and 2; initial viral load measurement of 60 c/mL;  $p = 20000\text{day}^{-1}$  in (B). doi:10.1371/journal.pcbi.1002033.g011

With improving drug efficacy ( $\epsilon \rightarrow 1$ ) our model predicts a significant decrease in baseline viral load, down to 3–4 c/mL, in accordance with recent viral load observations [10,11]. From this low baseline, we found blip probabilities smaller than we can calculate. However, we found that by increasing the activation rate (roughly simulating immune system activation due to secondary infection) viral loads exceeding the threshold of detection were attainable. Therefore, our model supports the hypothesis that, for patients adhering to modern ART, viral blips signal an underlying secondary condition.

#### Duration of viral blips

We examined the duration of viral blips through direct (Gillespie) simulation of the model. We sought to answer the question “Given a patient measurement of  $X$  c/mL, how long can we expect the viral load to remain detectable?” We found, unsurprisingly, that blips of larger initial amplitude have longer mean duration and larger standard deviation in duration. Perhaps more interestingly, we found very strong dependence on the production rate  $p$ . Given an initial blip of amplitude 90 c/mL, doubling the production rate from  $p = 10000\text{virions/day}$  to  $p = 20000\text{virions/day}$ , and changing other parameters accordingly, more than triples the predicted mean blip duration (for parameters as in Figure 9). We also considered blip duration as a function of the latent reservoir size, anticipating longer durations for larger reservoir sizes, since associated with these is a higher quasi-steady mean viral load. Our expectations were confirmed by simulation results (see Figure 11). The sensitivity to production rate  $p$  and associated parameters was also recovered.

Repeat-blip measurements in patients are, predictably, rather rare, since blips are already quite unusual events. Across all the parameter sets we examined, we found that detectable viremia should be expected to vanish within 8–10 days at most. This result is in general agreement with previous reports [34,45] and indicates that repeat low-level detectable viremia within 8–10 days could be due to a statistical fluctuation rather than drug resistance or other pathology.

#### Stochastic modeling of viral infection

Over the last 15 years, enormous numbers of differential-equation models have been generated to study different aspects of

various viral infections. We believe that stochastic models of the kind described here have an important role to play in certain situations where viral or cell populations are small enough that random effects still play a role. The obvious settings are during the first few days of any new infection (see also [57] and, very recently, [58]), during drug treatment of a chronic infection, and during the extinction phase of an acute infection. One issue with stochastic modeling of rare events (such as viral blips in our model) is that simulation-driven studies can require enormous numbers of simulations to reliably sample the rare events. The method we describe here is an alternative to simulation (or methods to capture rare-events) and provides a direct and relatively straightforward way to calculate probability distribution functions. We hope to adapt this method to other situations in viral dynamics in future work.

#### Supporting Information

**Figure S1** Comparison between our probability distribution function calculations and direct numerical simulations using the Gillespie algorithm. Distributions over the number of productively infected cells are plotted at 1 year, starting with 1 per  $10^6$  latently infected cells only, for parameters given in Tables 1 and 2. (A–C) Frequencies over  $10^5$  stochastic simulations are compared to probability distributions derived using our method, for (A)  $p = 5000\text{day}^{-1}$ , (B)  $p = 10000\text{day}^{-1}$  (C)  $p = 20000\text{day}^{-1}$ . (D) Enlargement of tail in (C), using a log scale for clarity, with frequencies over  $10^6$  Gillespie simulations. Notice that direct calculation of the probability distribution is clearly preferable to simulation when rare events are studied. (TIFF)

#### Acknowledgments

We acknowledge Rafael Meza, Michael Gilchrist, Michele Di Mascio and Alan Perelson for helpful discussions.

#### Author Contributions

Conceived and designed the experiments: JMC DC. Performed the experiments: JMC. Analyzed the data: JMC DC. Contributed reagents/materials/analysis tools: JMC DC. Wrote the paper: JMC DC. Designed and implemented algorithms: JMC.

## References

- Bowman MC, Archin NM, Margolis DM (2009) Pharmaceutical approaches to eradication of persistent HIV infection. *Expert Rev Mol Med* 11: e6.
- Violari A, Cotton MF, Gibb DM, Babiker AG, Steyn J, et al. (2008) Early Antiretroviral Therapy and Mortality among HIV-Infected Infants. *New Engl J Med* 359: 2233–2244.
- Dornadula G, Zhang H, VanUitert B, Stern J, Livornese L, et al. (1999) Residual HIV-1 RNA in blood plasma of patients taking suppressive highly active antiretroviral therapy. *JAMA - J Am Med Assoc* 282: 1627–1632.
- Pamer EG (2004) Immune responses to *Listeria monocytogenes*. *Nat Rev Immunol* 4: 812–823.
- Maldarelli F, Palmer S, King MS, Wiegand A, Polis MA, et al. (2007) ART suppresses plasma HIV-1 RNA to a stable set point predicted by pretherapy viremia. *PLoS Pathog* 3: e46.
- Blankson JN, Persaud D, Siliciano RF (2002) The challenge of viral reservoirs in HIV-1 infection. *Annu Rev Med* 53: 557–593.
- Schrager LK, D'Souza MP (1998) Cellular and anatomical reservoirs of HIV-1 in patients receiving potent antiretroviral combination therapy. *JAMA - J Am Med Assoc* 280: 67–71.
- Havlic DV, Strain MC, Clerici M, Ignacio C, Trabattini D, et al. (2003) Productive infection maintains a dynamic steady state of residual viremia in human immunodeficiency virus type 1-infected persons treated with suppressive antiretroviral therapy for five years. *J Virol* 77: 11212–11219.
- Joos B, Fischer M, Kuster H, Pillai SK, Wong JK, et al. (2008) HIV rebounds from latently infected cells, rather than from continuing low-level replication. *Proc Natl Acad Sci USA* 105: 16725–16730.
- Dinosa JB, Kim SY, Wiegand AM, Palmer SE, Gange SJ, et al. (2009) Treatment intensification does not reduce residual HIV-1 viremia in patients on highly active antiretroviral therapy. *Proc Natl Acad Sci USA* 106: 9403–9408.
- Buzón MJ, Massanella M, Llibre JM, Esteve A, Dahl V, et al. (2010) HIV-1 replication and immune dynamics are affected by raltegravir intensification of HAART-suppressed subjects. *Nat Med* 16: 460–466.
- Chun TW, Carruth L, Finzi D, Shen XF, DiGiuseppe JA, et al. (1997) Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* 387: 183–188.
- Lassen K, Han YF, Zhou Y, Siliciano J, Siliciano RF (2004) The multifactorial nature of HIV-1 latency. *Trends Mol Med* 10: 525–531.
- Chun TW, Engel D, Mizell SB, Ehler LA, Fauci AS (1998) Induction of HIV-1 replication in latently infected CD4(+) T cells using a combination of cytokines. *J Exp Med* 188: 83–91.
- Chomont N, El-Far M, Ancuta P, Trautmann L, Procopio FA, et al. (2009) HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat Med* 15: 893–899.
- Muller V, Viguera-Gomez JF, Bonhoeffer S (2002) Decelerating decay of latently infected cells during prolonged therapy for human immunodeficiency virus type 1 infection. *J Virol* 76: 8963–8965.
- Brooks DG, Kitchen SG, Kitchen CMR, Scripture-Adams DD, Zack JA (2001) Generation of HIV latency during thymopoiesis. *Nat Med* 7: 459–464.
- Pierson T, McArthur T, Siliciano RF (2000) Reservoirs for HIV-1: Mechanisms for viral persistence in the presence of antiviral immune responses and antiretroviral therapy. *Annu Rev Immunol* 18: 665–708.
- Coiras M, López-Huertas MR, Pérez-Olmeda M, Alcami J (2009) Understanding HIV-1 latency provides clues for the eradication of long-term reservoirs. *Nat Rev Microbiol* 7: 798–812.
- Fondere JM, Planas JF, Huguet MF, Baillat V, Bolos F, et al. (2004) Enumeration of latently infected CD4(+) T cells from HIV-1-infected patients using an HIV-1 antigen ELISPOT assay. *J Clin Virol* 29: 33–38.
- Siliciano JD, Kajdas J, Finzi D, Quinn TC, Chadwick K, et al. (2003) Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4(+) T cells. *Nat Med* 9: 727–728.
- Monie D, Simmons RP, Nettles RE, Kieffer TL, Zhou Y, et al. (2005) A novel assay allows genotyping of the latent reservoir for human immunodeficiency virus type 1 in the resting CD4(+) T cells of viremic patients. *J Virol* 79: 5185–5202.
- Richman DD, Margolis DM, Delaney M, Greene WC, Hazuda D, et al. (2009) The challenge of finding a cure for HIV infection. *Science* 323: 1304–1307.
- Gunthard HF, Havlic DV, Fiscus S, Zhang ZQ, Eron J, et al. (2001) Residual human immunodeficiency virus (HIV) type 1 RNA and DNA in lymph nodes and HIV RNA in genital secretions and in cerebrospinal fluid after suppression of viremia for 2 years. *J Infect Dis* 183: 1318–1327.
- Macias J, Palomares JC, Mira JA, Torres MJ, Garcia-Garcia JA, et al. (2005) Transient rebounds of HIV plasma viremia are associated with the emergence of drug resistance mutations in patients on highly active antiretroviral therapy. *J Infection* 51: 195–200.
- Havlic DV, Bassett R, Levitan D, Gilbert P, Tebas P, et al. (2001) Prevalence and predictive value of intermittent viremia with combination HIV therapy. *JAMA - J Am Med Assoc* 286: 171–179.
- Mira JA, Macias J, Nogales C, Fernandez-Rivera J, Garcia JA, et al. (2002) Transient rebounds of low-level viraemia among HIV-infected patients under HAART are not associated with virological or immunological failure. *Antivir Ther* 7: 251–256.
- Sklar PA, Ward DJ, Baker RK, Wood KC, Gafoor Z, et al. (2002) Prevalence and clinical correlates of HIV viremia ("blips") in patients with previous suppression below the limits of quantification. *AIDS* 16: 2035–2041.
- Grennan T, Raboud J, Su D, Cooper C, Klein M, et al. (2010) Impact of different HIV viral load assays on viral load "blip" rates and their association with virologic failure [Abstract O008]. In: 19th Annual Canadian Conference on HIV/AIDS Research; May 13–16, 2010; Saskatoon, Saskatchewan, Canada. CAHR 2010. Available: [http://www.cahr-acrv.ca/english/resources/abstracts\\_2010/abs.php?abs=O008](http://www.cahr-acrv.ca/english/resources/abstracts_2010/abs.php?abs=O008). Accessed 8 March 2011.
- Smit E, Bhattacharya S, Osman H, Taylor S (2009) Increased frequency of HIV-1 viral load blip rate observed after switching from Roche Cobas Amplicor to Cobas Taqman Assay. *JAIDS* 51: 364–365.
- Di Mascio M, Markowitz M, Louie M, Hogan C, Hurley A, et al. (2003) Viral blip dynamics during highly active antiretroviral therapy. *J Virol* 77: 12165–12172.
- Gunthard HF, Wong JK, Spina CA, Ignacio G, Kwok S, et al. (2000) Effect of influenza vaccination on viral replication and immune response in persons infected with human immunodeficiency virus receiving potent antiretroviral therapy. *J Infect Dis* 181: 522–531.
- Tobin NH, Learn GH, Holte SE, Wang Y, Melvin AJ, et al. (2005) Evidence that low-level viremias during effective highly active antiretroviral therapy result from two processes: Expression of archival virus and replication of virus. *J Virol* 79: 9625–9634.
- Nettles RE, Kieffer TL, Kwon P, Monie D, Han YF, et al. (2005) Intermittent HIV-1 viremia (blips) and drug resistance in patients receiving HAART. *JAMA - J Am Med Assoc* 293: 817–829.
- Perelson AS, Essunger P, Cao YZ, Vesanan M, Hurley A, et al. (1997) Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* 387: 188–191.
- Kim HJ, Perelson AS (2006) Viral and latent reservoir persistence in HIV-1-infected patients on therapy. *PLoS Comput Biol* 2: e135.
- Ferguson NM, deWolf F, Ghani AC, Fraser C, Donnelly CA, et al. (1999) Antigen-driven CD4+T cell and HIV-1 dynamics: Residual viral replication under highly active antiretroviral therapy. *Proc Natl Acad Sci USA* 96: 15167–15172.
- Fraser C, Ferguson NM, Anderson RM (2001) Quantification of intrinsic residual viral replication in treated HIV-1 patients. *Proc Natl Acad Sci USA* 98: 15167–15172.
- Fraser C, Ferguson NM, de Wolf D, Anderson RM (2001) The role of antigenic stimulation and cytotoxic T cell activity in regulating the long-term immunopathogenesis of HIV: mechanisms and clinical implications. *Proc R Soc B* 268: 2085–2095.
- Jones LE, Perelson AS (2002) Modeling the effects of vaccination on chronically infected HIV-positive patients. *JAIDS* 31: 369–377.
- Jones LE, Perelson AS (2005) Opportunistic infection as a cause of transient viremia in chronically infected HIV patients under treatment with HAART. *Bull Math Biol* 67: 1227–1251.
- Jones LE, Perelson AS (2007) Transient viremia, plasma viral load, and reservoir replenishment in HIV-Infected patients on antiretroviral therapy. *J AIDS* 45: 483–493.
- Rong L, Perelson AS (2009) Modeling HIV persistence, the latent reservoir, and viral blips. *J Theor Biol* 260: 308–31.
- Di Mascio M, Dornadula G, Zhang H, Sullivan J, Xu Y, et al. (2003) In a subset of subjects on highly active antiretroviral therapy, human immunodeficiency virus type 1 RNA in plasma decays from 50 to <5 copies per milliliter, with a half-life of 6 months. *J Virol* 77: 2271–2275.
- Di Mascio M, Percus JK, Percus OE, Markowitz M, Ho DD, et al. (2005) Duration of an intermittent episode of viremia. *Bull Math Biol* 67: 885–900.
- Karlin S, Taylor HM (1975) *A First Course in Stochastic Processes* Academic Press, Inc.
- Perelson AS, Kirschner DE, de Boer R (1993) Dynamics of HIV infection of CD4+ T cells. *Math Biosci* 114: 81.
- Markowitz M, Louie M, Hurley A, Sun E, Di Mascio M (2003) A novel antiviral intervention results in more accurate assessment of human immunodeficiency virus type 1 replication dynamics and T-Cell decay in vivo. *J Virol* 77: 5037–5038.
- Ramratnam B, Bonhoeffer S, Binley J, Hurley A, Zhang LQ, et al. (1999) Rapid production and clearance of HIV-1 and hepatitis C virus assessed by large volume plasma apheresis. *Lancet* 354: 1782–1785.
- Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, et al. (1997) Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science* 278: 1295–1300.
- Ramratnam B, Ribeiro R, He T, Chung C, Simon V, et al. (2004) Intensification of antiretroviral therapy accelerates the decay of the HIV-1 latent reservoir and decreases, but does not eliminate, ongoing virus replication. *J AIDS* 35: 33–37.
- Krakovska O, Wahl L (2007) Optimal drug treatment regimens for HIV depend on adherence. *J Theor Biol* 246: 499–509.
- Margolis DM (2010) Mechanisms of HIV Latency: an Emerging Picture of Complexity. *Curr HIV/AIDS Rep* 7: 37–43.
- Kitahata MM, Gange SJ, Abraham AG, Merriman B, Saag MS, et al. (2009) Effect of early versus deferred antiretroviral therapy for HIV on survival. *New Engl J Med* 360: 1815–1826.

55. Granich R, Gilks C, Dye C, De Cock K, BG W (2009) Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model. *Lancet* 373: 48–57.
56. Ribeiro RM, Bonhoeffer S (2000) Production of resistant HIV mutants during antiretroviral therapy. *Proc Natl Acad Sci USA* 97: 7681–7686.
57. Merrill SJ (1989) Modeling the interaction of HIV with cells of the immune system. In: Castillo-Chavez C, ed. *Mathematical and statistical approaches to AIDS epidemiology*. New York: Springer-Verlag, volume 83 of *Lecture Notes in Bioinformatics*. pp 371–385.
58. Pearson JE, Krapivsky P, Perelson AS (2011) Stochastic theory of early viral infection: Continuous versus burst production of virions. *PLoS Comput Biol* 7: e1001058.