Early HIV infection predictions: role of viral replication errors*

Jessica M. Conway† and Alan S. Perelson‡

Abstract. In order to prevent and/or control infections it is necessary to understand their early-time dynamics. However this is precisely the phase of HIV about which the least is known. To investigate the initial stages of HIV infection within a host we have developed a multi-type, continuous-time branching process model. This model is a stochastic extension of the standard viral dynamics model, under the assumption that the number of cell targets for viral infection is constant. We use our model to investigate three important clinical characteristics of early HIV infection following intravenous challenge: risk of infection, time to infection clearance (assuming failed infection), and time to infection detection. Our focus is on the impact of errors in viral replication that result in non-infectious virus production on these characteristics. Only a small fraction of circulating virus in any chronically infected individual is capable of infecting susceptible cells: estimates range from $1/10^4 - 1/10^3$. Characterization and quantification of the processes by which virus becomes defective remains incomplete. We consider two mechanisms that result in defective virus: (1) Copying errors, i.e., lethal errors in reverse transcription, which introduce mutations into the HIV-1 proviral genome, some of which may cripple the viral genome produced, and (2) Packaging errors, i.e., errors during viral packaging, at the end of the viral replication cycle, which cause defective virus by packaging new virions without, for example, viral RNA or key proteins required for infectivity. We show that assumptions on mechanisms of defective virus production can significantly impact early HIV infection model predictions. For example, the risk of infection is orders of magnitude higher if all defective virus is associated with packaging errors, but infection is predicted to be detectable sooner following HIV exposure if all defective virus is associated with copying errors. Thus, in order to make reliable predictions of risk, clearance time, and detection time, better characterization of viral replication is required.

1. Introduction.

HIV populations in chronically infected individuals are heterogeneous. HIV is constantly evolving, with different viral populations competing to become the dominant strain. But surprisingly, only small fraction of circulating virus in any infected individual is capable of infecting susceptible cells: estimates range from $1/10^4 - 1/10^3$ [9, 46, 60, 79]. Our aim is to investigate the effect of non-infectious viral production in the earliest stages of HIV infection.

Within-host events following exposure to HIV are critical in predicting whether infection will occur. We know from epidemiological studies that the probability or risk of infection is low, on the order of 0.1%–1% per sex act, percutaneous needlestick, or needle-sharing drug use [47, 74]. Further, from phylogenetic studies, we know that many infections arise due to expansion of a single viral strain [39], called the transmitted/founder virus. However, direct investigations of the early events in human or animal model infections are very difficult because viral and infected cell populations are very small. Mathematical modeling can be invaluable in investigating the earliest phase of infection, but it is important to understand how underlying

*Resubmitted to the editors X December 2017.

Funding: ASP acknowledges support by National Institutes of Health Grants R01-AI028433 and R01-OD011095.

†Department of Mathematics and Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, Pennsylvania, United States of America (jmconway@psu.edu).

‡Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, New Mexico, United States of America (asp@lanl.gov).

model assumptions such as viral infectiousness affect modeling predictions.

Infectious fractions ranging in $1/10^4 - 1/10^3$ [3, 46, 60, 79] are *in vivo* estimates based on viral samples taken from chronically infected individuals (a recent study suggests that that fraction may be larger than previously thought [73]). Though quantified through a simple fraction, non-infectious virus can arise through a number of mechanisms, such as host antibodies binding of viral surface proteins necessary for viral entry into a target cell. Here we focus specifically on viral replication errors that give rise to non-infectious virus. The reverse transcription step of viral replication, when the RNA of a virion that has penetrated its target cell, gets copied into DNA, is error prone - mutations, including insertions, deletions, and base substitutions [31], are introduced *in vivo* at a rate of $\sim O(10^{-5})$ per base per replication cycle [1,57,58] (the HIV genome is roughly 10^4 bases long). Frameshift mutations can also be introduced into the HIV-1 proviral genome [1]. Mutations are the primary driver of escape from cytotoxic T lymphocyte (CTL) responses and to the generation of drug-resistant viral strains. They can also fatally cripple the proviral genome so that any viral genomes produced by the provirus will not be infectious [22, 31]. Virus may also be rendered non-infectious by errors in virion assembly and packaging. For instance, to be infectious, virions must be packaged to include two RNA molecules encoding functional virus as well as the HIV enzymes reverse transcriptase and integrase and have sufficient surface proteins (gp120/gp41) necessary for viral binding and entry into target cells [51].

Our aim in this paper is to investigate how assumptions on viral replication errors leading to non-infectious virus affect predictions with regards to important variables in the earliest stages of HIV infection: (1) Risk of infection, i.e., the probability of becoming infected after exposure to a viral inoculum. Interventions such as prophylactic use of antiretroviral drugs (pre- or post-exposure, PrEP or PEP) [23,48], male circumcision [5], immunization [36], all aim to reduce risk. Therefore modeling predictions of risk can be of great value and clinical use. (2) Time to infection clearance, and (3) time to infection detection, in the case of unsuccessful or successful infection respectively. Obtaining these time distributions can help us characterize the course of early infection. They can also be of direct clinical use. For example, distributions on time to detection can offer some guidance for HIV testing windows [7, 67]. We investigate these measures in the context of intravenous exposure to HIV, i.e., via occupational needlestick exposure or intravenous drug use, where a well-mixed model with no spatial structure most likely applies.

Deterministic (differential equation) models have been very effective in characterizing HIV infection, for example in determining viral and infected cell clearance times [71]. Variants of the now-standard viral dynamics model, first developed to investigate HIV [65, 69, 71] have since been used to gain insight into a multitude of viral infections, including dengue [15], West Nile virus [6], cytomegalovirus [21], hepatitis B [64], hepatitis C [63], influenza [4,27,32], and Zika [2, 8, 66]. From a mathematical perspective, differential equation models represent average behavior of a system and are appropriate when numbers are large, as viral loads and infected cell concentrations certainly are in HIV chronically-infected individuals. Here we are interested in the earliest stages of HIV infection, when numbers of virions and in particular infected cells can be quite small. A stochastic approach is therefore more appropriate: stochastic models can give varying predictions in regimes where deterministic models, focusing on mean behavior only, cannot. For example, predictions on risk of infection - inaccessible

via differential equation modeling - is different if viral production is assumed to be continuous through an infected cell's life, or to occur in a burst at the end of the infected cell's life, for the same mean number of virus particles produced [68].

Our stochastic model will be built upon the standard viral dynamics model [70,71], involving cells susceptible to HIV infection, infected cells, and virus only. This simple model ignores spatial effects, the effects of long-lived and latently infected cells, which may affect long-term dynamics of HIV infection, as well as the possibility of cell-to-cell infection [24]. The model also ignores the delay from the time a virus enters a cell until it begins producing virus, i.e. the eclipse phase, although this can be included in the model [18,20,30,61]. Further, since we focus on the earliest stages of HIV infection during which time very few cells become infected, we will consider no immune response and ignore dynamics in the number of susceptible cells. By keeping the number of susceptible cells, i.e., target cells, constant the dynamical equations become linear and more amenable to analysis, although we recognize that in certain spatial locations target cells may be limiting and need to be accounted for in spatial models. We are interested in quantities such as risk of infection, time to viral clearance, and time to detection. Our model is best suited to gain qualitative insight into these quantities, for example, the relative impact of assumptions on mechanisms producing defective virus on these quantities. To get at the early infection quantities of interest, we will extend stochastic approaches used in our recent theoretical studies [17, 18, 68], which themselves build on previous stochastic modeling literature [29, 50, 82, 85].

We formulate our stochastic model as a multi-type branching process [28,37,41]. Branching process models have long been used to investigate and model biological processes as they are simply expressed, and yet include noise inherent in any biological system. Multi-type continuous time branching processes, have been used to gain insight into, for example, fluctuation theory [41], carcinogenesis [53, 62], cellular processes [55, 56], immunology and T-cell population dynamics [77], population dynamics and ecology [26, 44], and epidemiology [35]. Our own recent modeling of within-host HIV dynamics [17, 18, 68], which we extend in this present study, relied on much of this previously developed theory.

The structure of this paper is as follows. First we give a model overview. We then offer details on the calculations of risk of infection, time to clearance, and detection time, and give related results. Finally we discuss the results and their broader implications.

2. Viral dynamics model.

Our basic mathematical model of early HIV infection is presented schematically in Figure 1(a). There are four compartments: infectious and non-infectious virus, V and V_d, respectively, infected cells, I, and infected cells that produce only non-infectious virus, I_d. Since non-infectious virus is sometimes called defective virus, we use the subscript 'd.' There may be a continuum of infectivities across virus within a host, if we consider variation in envelope protein, but for simplicity we consider only the infectious and non-infectious extremes.

Infected cells I and I_d produce virus at rate p. Virions V and V_d are cleared at rate c. We assume mass-action kinetics for cell infection: Infectious virus V, infects target cells, T, at a rate proportional to their product, with proportionality constant k. We assume T to be constant, a necessary assumption since we wish to focus on specific mechanisms pertaining to viral replication rather than target cell limitation. We note that this assumption is probably

biologically reasonable for intravenous infection since very few cells are infected relative to the total number at risk during the earliest stages of HIV infection. Thus in our model, infectious virions make new infected cells at rate kT. Infected cells I and I_d die at rate δ. In principle, rates associated with infected cells I_d may differ from those associated with infected cells I. However, there are no estimates or qualitative studies to give us insights into how they may differ. We therefore make the simplifying assumption that rates associated with I and I_d are the same.

Mathematically speaking, the events we model - viral production and clearance, cell infection and death - are independent. In the standard model (cf. eq. (SM1)), cell infection is not an independent process, as it depends on the density of both virus and target cells. However since we make the assumption that the target cell population remains approximately constant at the earliest stage of infection, the target cell count is no longer a dependent variable, and each virion can infect a cell or clear independently of the remaining viral or infected cell populations. The constant target cell assumption is therefore required for independence to hold, and permits the use of branching processes.

Our focus is on errors in viral replication that result in non-infectious virus. There are two such mechanisms: lethal reverse transcription (copying) errors and packaging errors. Reverse transcription, or copying, errors introduce mutations into the HIV-1 proviral genome, some of which may cripple the viral genome produced in any of multiple ways, e.g., by the introduction of stop codons, large deletions or frameshifts. This can be exacerbated by the cellular antiviral enzyme APOBEG3G, that induces hypermutations in HIV DNA during reverse transcription [87,88]. These are our lethal copying errors. Because the host polymerase, which transcribes the HIV proviral genome into viral RNA, does so with high fidelity (error rate 10^{-9} per base pair per year [45]) we assume that cells with a crippled proviral genome only produce non-infectious virions. Packaging errors occur at the end of the viral replication cycle, during the packaging step: for example, virions may be packaged without RNA, without HIV enzymes such as reverse transcriptase, protease or integrase or with insufficient surface proteins (gp120/gp41) necessary for viral binding to target cells [51]. We'll assume these packaging errors occur at random in any produced virion.

We consider two classes of mutation.

1. Mutations that affect all virions produced by an infected cell equally. These are the reverse transcription or copying errors. In our model, the probability of a lethal copying error is $1-Q_c$, where Q_c is the probability of no, or non-lethal, copying errors. Therefore, new cell infection (at rate kT) yields
   - an infected cell I, containing an infectious proviral viral genome, with probability Q_c, or
   - an infected cell I_d, containing a *non*-infectious viral genome, with probability $1-Q_c$.
2. Mutations that affect individual virions produced by an infected cell. These are primarily routine packaging errors. In our model, the probability of a packaging error is $1-Q_p$, where Q_p is the probability of that a virion is correctly packaged. Therefore, for infected cells I, viral production (at rate p) yields
   - with probability Q_p, correctly packaged - and therefore infectious - virion V
   - with probability $1-Q_p$, incorrectly packaged - and therefore *non*-infectious - virion V_d

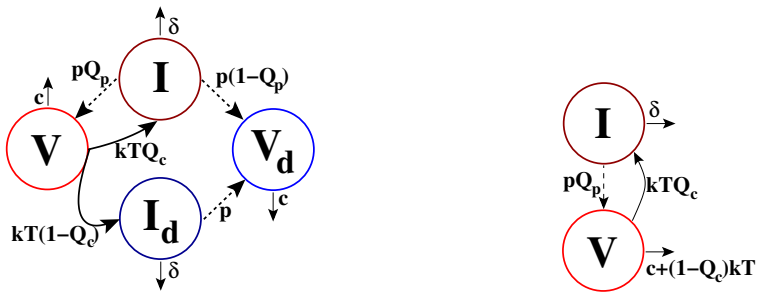

Figure 1. *Model schematics:* (a) *Basic model,* (b) *reduced model. I represent infected cells, V virions, with I_d and V_d representing defective infected cells and virions respectively. Each infected cell, I or I_d, dies at rate δ or produces a virion at rate p. Non-infectious infected cells I_d produce only non-infectious virus, V_d, while infectious infected cells I may also produce infectious virions, V, with probability Q_p. The dashed line indicates viral production without loss of the virus-producing infected cell. Each infectious virion, V, infects a target cell T with mass-action infectivity k; the newly infected cell is infectious, I, with probability Q_c and is otherwise non-infectious, I_d. Note that the number of target cells T is held constant. In the reduced model (b) we neglect the dynamics of non-infectious cells and virus, I_d and V_d.*

All virus produced by infected cells I_d are defective. Note that the mean number of infectious virus produced by infected cells I is $Q_p p/\delta$ where p/δ is the average number of virus particles produced during the infected cell's lifetime $(1/\delta)$, commonly called the "burst size" irrespective of whether the virus is produced continuously, as assumed here, or in a burst.

3. Mathematical approach. Traditionally mathematical modeling of viral dynamics takes a deterministic, ordinary differential equations approach. However since we seek to investigate events in early HIV infection, it is necessary to use a stochastic approach (see section SM2 for discussion).

The modeling framework we will use is continuous-time branching processes [28, 37, 41], extending previously-developed theory [17, 18, 68]. For our model shown in Figure 1(a) define

$$P_{n,v,m,w;n_0,v_0,m_0,w_0}(t) = P(I(t)=n, V(t)=v, I_d(t)=m, V_d(t)=w \\ |I(0)=n_0, V(0)=v_0, I_d(0)=m_0, V_d(0)=w_0)$$

as the probability that at time t there are n infected cells, v infectious virions, m defective infected cells, and w defective virions, given that at time 0 there were n_0 infected cells, v_0 infectious virions, m_0 defective infected cells, and w_0 defective virions.

For the purposes of computing extinction probabilities we need not explicitly include defective compartments (I_d, V_d) - we need only consider the reduced model shown in Figure 1(b). Infected cells produce (infectious-only) virus at rate pQ_p, and die at rate δ. Virions infect susceptible cells T at rate k, producing infected cells at rate $Q_c kT$.Virions are cleared at rate c. Define for this reduced model

$$P_{n,v;n_0,v_0}(t) = P(I(t)=n, V(t)=v | I(0)=n_0, V(0)=v_0),$$

as the probability that at time t there are n infected cells and v infectious virions, given that at time 0 there are n_0 infected cells and v_0 virions. In the calculations that follow we will

mainly use this reduced probability. The defective compartments will only be included in the of probability of detection calculation, since detection is of total viral load $(V + V_d)$, assuming the defective virus still contains HIV RNA.

We proceed by deriving the probability generating function differential equations for the probability $P_{n,v;n_0,v_0}(t)$. We will use these to derive expressions for risk of infection and time to infection clearance. To derive the time-dependent probability of infection detection, we use the extended equations and the associated probability generating function for $P_{n,m,v,w;n_0,m_0,v_0,w_0}(t)$.

3.1. Chapman-Kolmogorov differential equation and the probability generating function. We begin with the backwards Chapman-Kolmogorov differential equation (bCKde) for the probability $P_{n,v;n_0,v_0}(t)$,

$$\begin{aligned}\frac{d}{dt}P_{n,v;n_0,v_0} =&\delta n_0 P_{n,v;n_0-1,v_0} + pQ_p n_0 P_{n,v;n_0,v_0+1} + Q_c kT v_0 P_{n,v;n_0+1,v_0-1}\\ &+ (1-Q_c)kT v_0 P_{n,v;n_0,v_0-1} + cv_0 P_{n,v;n_0,v_0-1}\\ &- ((\delta + pQ_p)n_0 + (kT + c)v_0)\, P_{n,v;n_0,v_0}\end{aligned} \tag{1}$$

with initial condition $P_{n,v;n_0,v_0}(0) = \delta_{n,n_0}\delta_{v,v_0}$. The derivation of the bCKde is given subsection SM3.1.

Define the probability generating function (PGF) $G_{n_0,v_0}(x, y; t)$:

$$G_{n_0,v_0}(x, y; t) = \sum_{n=0}^{\infty}\sum_{v=0}^{\infty} P_{n,v;n_0,v_0} x^n y^v. \tag{2}$$

We use the generating function because its derivatives give us individual probabilities and moments. For example, the marginal probability distribution on the viral load is given by derivatives of $G_{n_0,v_0}(1, y; t)$, and the mean viral load at time t is given by $\left.\frac{\partial G_{n_0,v_0}(1,y;t)}{\partial y}\right|_{y=1}$. Multiplying eq. (1) by $x^n y^v$ and summing over the exponents, we obtain an equation for G_{n_0,v_0}:

$$\begin{aligned}\frac{\partial}{\partial t}G_{n_0,v_0} =&\delta n_0 G_{n_0-1,v_0} + pQ_p n_0 G_{n_0,v_0+1} + Q_c kT v_0 G_{n_0+1,v_0-1}\\ &+ (1-Q_c)kT v_0 G_{n_0,v_0-1} + cv_0 G_{n_0,v_0-1}\\ &- ((\delta + pQ_p)n_0 + (kT + c)v_0)\, G_{n_0,v_0}\end{aligned} \tag{3}$$

with initial condition $G_{n_0,v_0}(0) = x^{n_0}y^{v_0}$. We can reduce this infinite-dimensional system to a two-dimensional system by exploiting the branching property [37]: $G_{n_0,v_0} = (G_{1,0})^{n_0}\,(G_{0,1})^{v_0}$. The branching property derives from our important assumption that cells and virions of each type behave identically and independently of all other cells and virions. We derive two ODEs from which we can recover the PGF,

$$\begin{aligned}\frac{\partial G_{1,0}}{\partial t} =&\delta + pQ_p G_{1,0}G_{0,1} - (\delta + pQ_p)G_{1,0}\\ \frac{\partial G_{0,1}}{\partial t} =&c + kT\left((1-Q_c) + Q_c G_{1,0}\right) - (c + kT)G_{0,1}\end{aligned} \tag{4}$$

with initial conditions $G_{1,0}(0) = x$, $G_{0,1}(0) = y$. We will use the PGF and the associated nonlinear ODEs to compute quantities of interest such as risk of infection and infection clearance times.

We note that the complete derivation of these equations for the probability generating function is similar to derivations for other models that rely on continuous time branching processes [28, 37, 41]. Different biological processes can be modeled in a similar manner. For example, in fluctuation theory [41] and stochastic models of carcinogenesis [53, 62], which focus on initially-homogeneous cell populations that accumulate mutants over time, the primary "branching" mechanism is mutation. Continuous viral production in our model, (1), is described in the same way as cell division with one identical, and one mutant, daughter cell [41]. Cell infection in our model is described in the same way as backwards mutation [41]. However to our knowledge, no biological processes have been modeled with the same combination of mechanisms (Figure 1) and resulting equations ((1) and (4)) as in this present study.

4. Parameters. Baseline parameters for simulation results are summarized in Table 1.

Table 1
Model parameters.

Parameter	Description	Estimate	Reference(s)
δ	Infected cell death rate	$1\,\text{day}^{-1}$	[59, 90]
p	Viral production rate	$2000\,\text{day}^{-1}$	see text
c	Viral clearance rate	$23\,\text{day}^{-1}$	[75]
Q_c	Probability of reverse transcription leading to infectious provirus	varied, $10^{-2} - 1$	see text
Q_p	Probability of correct viral packaging	varied, $10^{-2} - 1$	see text
Q	Infectious virion fraction in inoculum	$10^{-3} - 10^{-1}$	[54]

Rate parameters during the early and chronic stages of HIV infection may differ. However, because there are few reported parameter estimates from the earliest stages of HIV infection we mainly use estimates for their chronic infection counterparts. We use the mean lifetime of infected cells ($1/\delta$) estimate of 1 day [59]. Estimates on lifetime virion production (burst size) from a single infected cell vary significantly, from a few hundred virions to tens of thousands [13, 25]; we'll use an mid-range value, $B = 2000$ virions, which gives us a virus production rate $p = B\delta = 2000\,\text{day}^{-1}$. The viral clearance rate estimate we use is $c = 23\,\text{day}^{-1}$ [75]; while this is an estimate from the chronic stage of infection, there is evidence that suggests viral clearance is equally rapid during early HIV infection [90]. The infection rate kT we will compute from the expression for the basic reproduction number

$$R_0 = \frac{Q_c Q_p p k T}{\delta(c + kT)} = \tilde{Q} B \gamma,$$

where $B = p/\delta$ is the infected cell burst size, $\gamma = kT/(c + kT)$ is the probability that a virion will infect a cell [68], and we define $\tilde{Q} = Q_c Q_p$. Models that do not distinguish infectious from

defective virus have $R_0 = B\gamma$ (c.f. [68]) as they implicitly assume all virions are equivalent and infectious,i.e., $Q_c = Q_p = 1$. We will compute kT using within-host estimates of R_0 for early HIV infection as discussed below [78]. Finally, estimates on the replication-competent fraction $\tilde{Q} = Q_c Q_p$ range from $10^{-4} - 10^{-3}$ for chronic infection [42, 60, 79]. There is some evidence this fraction is higher during the early stages of HIV and SIV infection [42, 54, 86]. We therefore use the range $10^{-3} - 1$ for each of Q_c and Q_p.

We have not given estimates for the inoculum size N and the fraction of replication-competent virus in that inoculum Q. This inoculum related Q is distinct from $\tilde{Q}$, the fraction of replication competent virus produced by the newly-infected host. Our aim is to show that different mechanisms for production of defective virus can effect early-infection predictions only, so we will not explore sensitivity of our results to these parameters. An average inoculum size, or even a distribution on the inoculum size N, is difficult to determine, as it depends on exposure type, severity, and viral load in the HIV+ individual involved in the exposure or in a syringe if by needle stick injury. However epidemiological studies do estimate risks of infection averaged over all these exposures; the risk of infection from percutaneous needle stick, for example, has been estimated to be 0.3% [47]. As in a previous study [18], we will assume that exposures are uniformly distributed, and fit the maximum inoculum size to a desired risk of infection. Here we use an inoculum size of $N = 1000$ virions, within the range of inoculum sizes that give a risk of infection of 0.3%, corresponding to that of occupational exposure [47], if we assume that inoculum sizes are uniformly distributed across infecting donors (see subsection 5.3, below). And as briefly discussed above with regards to Q_c and Q_p, estimates on the replication-competent fraction Q during chronic infection range from $10^{-4} - 10^{-3}$ [42, 60, 79]. For this present study we will set the replication-competent fraction of virus in the inoculum $Q = 10^{-3}$, although as indicated in Table 1, Q could be as high as 10^{-1}.

The basic reproduction number R_0 is a key parameter in our model: it is the average number of new cell infections induced by an infected cell I during its lifetime $1/\delta$. R_0 impacts the probability of extinction (risk of infection), time to extinction, and times at which the viral load will be detectable. We use the individual R_0 estimates from [78], derived from viral load data obtained from 47 plasma donors, who were originally HIV^- and became HIV^+. We adjust these measurements to suit our model: the R_0 values in [78] account for the delay τ between infection of a cell and the beginning of viral production. They measure the viral growth r and from their viral dynamics model show $R_0 \approx (1 + r/\delta)e^{r\tau}$, where δ is the infected cell death rate [78]. Since our model does not include the delay τ, the corresponding R_0 for our model is $R_0 \approx (1 + r/\delta)$. We use the reported median R_0 value, not accounting for the delay τ, $R_0 = 2.77$, and the interquartile range $(R_0^{25}, R_0^{75}) = (2.28, 3.06)$.

We are interested in how assumptions on different mechanisms for production of defective virus, either reverse transcription errors (probability $1 - Q_c$) or packaging errors (probability $1 - Q_p$) affect modeling predictions on the clinical outcomes of risk of infection, time to infection clearance, and probability of detection at time t. For the results we show, we focus on parameter regimes within which the product $Q_c Q_p = \tilde{Q}$ is held constant. We also hold $R_0 = \tilde{Q}\gamma B$ constant, where B is the infected cell viral burst size, and $\gamma = kT/(c + kT)$ is the probability that a virion will infect a cell. We choose R_0 constant to focus on regimes where the deterministic model predictions would be constant (see section SM2 for details). As one

cannot simultaneously fix $R_0 = \tilde{Q}\gamma B$, B, $\tilde{Q}$, and γ, we allow gamma to vary. γ is a function of the infection rate kT ($\gamma = kT/(c+kT)$), for which we do not have a reliable estimate, even to within an order of magnitude. We choose the product $\tilde{Q} = Q_c Q_p$ constant to focus on regimes with constant total fractions of replication competent and defective virus.

We present results as a function of the fraction of errors attributable to reverse transcription, which we call copying errors, relative to the total fraction of defective virus. This fraction is

$$\text{Copying error fraction} = K = \frac{1-Q_c}{1-Q_cQ_p}.$$

The quantity K can be derived from the steady-state predictions of the standard viral dynamics ODE model (see eq. (SM1) in section SM2). The steady-state fraction of defective infected cells is $I_d/(I_d+I) = 1-Q_c$, the copying error probability, while the steady-state fraction of defective virus is $V_d/(V_d+V) = p(1-Q_cQ_p)/(p-\delta) \approx 1-Q_cQ_p$, since $p \gg \delta$. Thus $K = (1-Q_c)/(1-Q_cQ_p)$ can be interpreted as a potentially directly measurable quantity once the viral load set-point is reached.

5. On risk of infection. The probability that infected cells/virus go extinct can be interpreted in a clinically useful manner. Risk of infection, can be calculated as (1 - the overall probability of extinction (as time $\to \infty$)). Reducing risk is the main goal of HIV prevention strategies.

5.1. Calculation: Risk of infection from extinction as $t \to \infty$. The probability that the infection is extinct at or before time t is given by $P_{ext}(t) = P_{0,0;n_0,v_0}(t)$, since 0 is an absorbing boundary. Recall that $P_{n,v;n_0,v_0}(t) = P(I(t)=n, V(t)=v | I(0)=n_0, V(0)=v_0)$, the probability that there remain n infected cells and v virions at time t given n_0 infected cells and v_0 virions at time 0. Expressing the probability of extinction $P_{ext}(t)$ in terms of the generating function G_{n_0,v_0} in eq. (2),

$$P_{ext}(t) = G_{n_0,v_0}(0,0;t). \tag{5}$$

We compute the risk of infection from the limiting probability of extinction P_{ext}^{∞}. As $t \to \infty$,

$$\begin{aligned} P_{ext}^{\infty} &= \lim_{t\to\infty} G_{n_0,v_0}(0,0;t) \\ &= \left(\lim_{t\to\infty} G_{1,0}(0,0;t)\right)^{n_0} \left(\lim_{t\to\infty} G_{0,1}(0,0;t)\right)^{v_0} \end{aligned}$$

with $\lim_{t\to\infty} G_{1,0}(0,0;t)$, $\lim_{t\to\infty} G_{0,1}(0,0;t)$ the stable fixed point of eq. (4),

$$\lim_{t\to\infty} G_{1,0}(0,0;t) = G_{1,0}^* = \begin{cases} 1, & R_0 \le 1 \\ \frac{\delta(c+kT)}{pkTQ_cQ_p} = \frac{1}{R_0}, & R_0 > 1 \end{cases}$$

$$\lim_{t\to\infty} G_{0,1}(0,0;t) = G_{0,1}^* = \begin{cases} 1, & R_0 \le 1 \\ \frac{\delta}{pQ_p} + \frac{c+(1-Q_c)kT}{c+kT} & R_0 > 1 \end{cases}$$

where R_0 is the basic reproduction number, $R_0 = Q_cQ_pB\gamma = pQ_cQ_pkT/\delta(c+kT)$. Recall that $\gamma = kT/(c+kT)$ is the probability that a single replication-competent virion infects a

cell [68], and that $B = p/\delta$ is the burst size. If $Q_c = Q_p = 1$, $G^*_{0,1}$ for $R_0 > 1$ simplifies to $1 - (R_0 - 1)/B$, in agreement with Pearson et al. (2011) and Conway et al. (2013) [18, 68].

Starting from a state with n_0 infected cells capable of producing infectious virus, and v_0 infectious virions, the probability of extinction as $t \to \infty$ in terms of γ and B is

$$P^{\infty}_{ext} = \begin{cases} 1, & Q_c Q_p B\gamma \leq 1 \\ \left(\frac{1}{Q_c Q_p B\gamma}\right)^{n_0} \left(\frac{1}{BQ_p} + 1 - Q_c\gamma\right)^{v_0}, & Q_c Q_p B\gamma > 1. \end{cases}$$

Now consider a virus-only inoculum as is typically used in non-human primate infection experiments. The key piece of information is the probability of extinction starting with a single infectious virion, i.e., $v_0 = 1$ and $n_0 = 0$. For $R_0 > 1$,

$$P^{\infty}_{ext} = \frac{1}{Q_p B} + 1 - Q_c\gamma, \tag{6}$$

It is unlikely that any real viral inoculum will contain infectious virions only. We assume that given an inoculum of size N, each virion is infectious with probability Q, and use a binomial distribution. From the branching property - that is, the assumption that cells and virions of each type behave identically and independently of all other cells and virions - the probability that the infection will be extinct at time t, for an inoculum of size N, with each virion replication-competent with probability Q, is

$$\begin{aligned} P^{\infty}_{ext}|_N &= \sum_{\ell=0}^{N} \{(\text{Probability that an inoculum of size } N \text{ contains } \ell \text{ infectious virions}) \\ &\quad \times (\text{Probability that the infection goes extinct given an} \\ &\quad\quad \text{inoculum of } \ell \text{ infectious virions})\} \\ &= \sum_{\ell=0}^{N} \binom{N}{\ell} (1-Q)^{N-\ell} Q^{\ell} \left(\frac{1}{Q_p B} + 1 - Q_c\gamma\right)^{\ell} \\ &= \left(1 - Q\left(Q_c\gamma - \frac{1}{Q_p B}\right)\right)^{N}, \end{aligned} \tag{7}$$

for $R_0 > 1$. In the last step we noted that the sum corresponds to the binomial expansion of $(1 - Q\,(Q_c\gamma - 1/Q_p B))^N$. Then risk of infection for an inoculum of size N, when $R_0 > 1$, is

$$\text{Risk} = 1 - \left(1 - Q\left(Q_c\gamma - \frac{1}{Q_p B}\right)\right)^{N}. \tag{8}$$

It is immediately clear that Q_c and Q_p have different effects on this risk, which we will explore in the following.

5.2. Predicted risk of infection is lowered by copying errors . We begin by examining the risk of infection for a constant assumed total fraction of infectious virus $\tilde{Q} = Q_c Q_p$. Figure 2(a,b) shows the % risk of infection given a single virus inoculum as a function of K,

the fraction of errors attributable to copying errors,

$$\begin{aligned}\text{Risk}|_{N=1} =& 1-\left(1-Q\left(Q_c\gamma-\frac{1}{Q_pB}\right)\right)\\ =& \frac{Q\left(R_0-1\right)}{B}\left(1+(1-K)\left(\frac{1}{\tilde{Q}}-1\right)\right)\end{aligned} \tag{9}$$

for different total replication competent fractions $\tilde{Q}=Q_cQ_p=10^{-3}$, 10^{-2}, and 10^{-1}. Recall that Q is the probability that a virus in the inoculum N is infectious. We use inoculum size $N=1$ virion for this calculation to isolate the relative impact of K on risk. In Figure 2(a,b), solid lines indicate median R_0 and the shaded areas, risk within the 25th and 75th percentile in R_0 [78]. From eq. (9), single-virus risk linearly decreases with the copying error fraction K. In the limiting cases,

1. all errors attributable to packaging errors, $K=0$. The risk $|_{K=0}=Q(R_0-1)/B\tilde{Q}$ decreases as the total replication competent fraction $\tilde{Q}$ increases. This decrease is clear upon inspection of the left $y-$axis in Figure 2(a,b): curves for smallest $\tilde{Q}$ (10^{-3}; green) give higher risk than curves for larger $\tilde{Q}$ (10^{-1}; red). Further, as $\tilde{Q}$ increases, the range $(R_0^{25}/\tilde{Q},\, R_0^{75}/\tilde{Q})$ decreases, so the range in risk between the 25th and 75th percentile in R_0 is larger for $\tilde{Q}=10^{-3}$ (Figure 2(a), shaded area in green) than for $\tilde{Q}=10^{-1}$ (Figure 2(a), shaded area in red).
2. all errors attributable to copying errors, $K=1$: Risk $=Q(R_0-1)/B=5\times10^{-7}(R_0-1)$. Risk increases with R_0. However the coefficient preceding (R_0-1) is very small, $Q/B=5\times10^{-7}$ with our parameter choices of $Q=10^{-3}$ and $B=2\times10^3$, and since $R_0-1\sim O(1)$ the differences are slight on the scale of % risk shown in Figure 2(a,b).

In the intermediate cases where errors are a combination of copying errors and packaging errors, $0<K<1$, the risk decreases as the copying error fraction K increases. Further, as that fraction K increases to 1, the impact of replication errors decreases and, with it, the interquartile range in risk associated with the interquartile range in R_0 also decreases.

That single-virus risk of infection decreases as the total fraction of replication-competent fraction $\tilde{Q}$ *increases* seems counter-intuitive. This result is an artifact of keeping R_0 fixed. $R_0=Q_cQ_p\gamma B=\tilde{Q}\gamma B$; in order to keep R_0 fixed while varying $\tilde{Q}$ we adjust the probability of infection $\gamma=kT/(c+kT)$. As $\tilde{Q}$ increases, that probability of infection γ decreases accordingly, and thus the single-virus risk of infection decreases as well.

We note similar trends in the sensitivity of risk and inoculum size to the assumed viral production rate p (see Figure SM1). This is unsurprising since the viral production rate p is in the numerator of R_0. Higher p leads to higher per-virion risk of infection (Fig S2a,b) and lower required inoculum for a fixed risk of infection (Fig S2c). Note however that the sensitivity to p is more significant, given the uncertainty in p: as we move p through a biologically reasonable range of 200 to 20000 virions per day [13, 25], we recover commensurate order-of-magnitude changes in risk of infection.

Taken together, predictions on single-virus risk of infection increase - by orders of magnitude, see Figure 2(b) - with the assumed fraction of replication-incompetent virus attributable to packaging errors, and also with the basic reproduction number R_0. Practically speaking, the reason single-virus risk is lower assuming copying errors dominate relates to the two-step

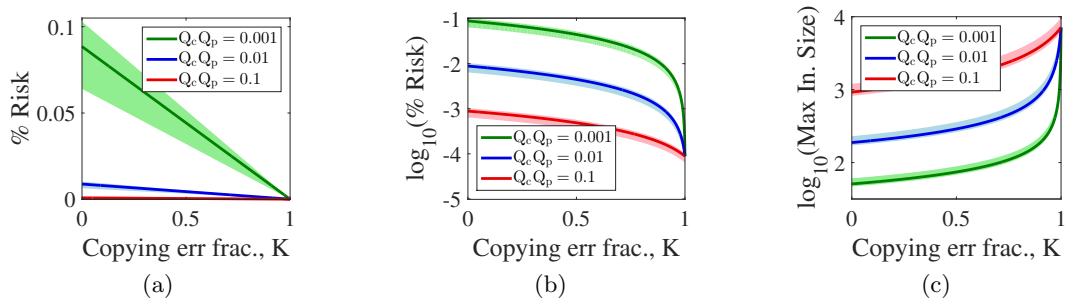

Figure 2. *Risk of infection and inoculum size for different total replication competent fractions* $\tilde{Q} = Q_cQ_p = 10^{-3}$, 10^{-2}, *and* 10^{-1}. (a-b) *% risk of infection given a single virus inoculum as a function of the fraction of errors attributable to copying errors, (9), on a* (a) *linear and* (b) *log scale.* (c) *Maximum inoculum size assuming a risk of infection of 0.3%, assuming inoculums to be uniformly distributed, computed from (10) using a nonlinear solver. Solid lines indicate median* $R_0 = 2.77$ *and the shaded areas, risk within the 25th and 75th percentile in* R_0, $(R_0^{25}, R_0^{75}) = (2.28, 3.06)$ *[78]. Remaining parameters can be found in Table 1.*

process of viral replication: in order to avoid extinction, inoculum virus must first infect cells (probability $Q_c kT/(c + kT)$), and only then those cells make viable virus. Hampering cell infection, as copying errors do, prevents the first step, successful cell infection, halting the potential of a propagating infection. Assuming packaging errors only, i.e., $Q_c = 1$, permits cell infection with much higher probability $(kT/(c + kT))$.

5.3. Inoculum size predictions increase with copying error fraction. Viral inoculum sizes associated with different high-HIV risk activities (e.g. needlestick exposure in hospital setting, unprotected vaginal intercourse) are unknown. However, epidemiological estimates of risk associated with these activities do exist; for example, occupational exposure in a hospital setting (e.g. needlestick) carries with it a risk of 0.3% [47]. These measures can be used to back-calculate the required inoculum size [18], in turn giving model-based insight into early infection dynamics. In the previous section we discussed risk of infection given an inoculum containing a single virus. To achieve a fixed risk of infection, the lower the single-virus risk, the higher the total viral inoculum required. The risk of infection initiated with a single virion decreases as the fraction of errors attributable to copying errors increases (as K increases from 0 to 1, cf. Figure 2(a,b)). We therefore anticipate that, to achieve a fixed risk of infection, as in occupational exposure, the required total inoculum increases with K.

To investigate total inoculum, assume that the risk of infection is 0.3%, corresponding to that of occupational exposure [47]. Occupational exposures vary in severity from needlestick exposures to blood splashes, and vary also according to the viral load of the donor, which can span orders of magnitude. In absence of information on a probability distribution on occupational-exposure inoculum size we assume a uniform distribution, $N \sim U(0, N_{max})$, as

in [18]. N_{max} is the maximum inoculum size. The total risk of infection is then

$$\begin{aligned}\text{Risk }|_{N\sim U(0,N_{max})} =& 1-\frac{1}{N_{max}}\sum_{j=0}^{N_{max}}\left(P_{ext}^{\infty}|_{N=1}\right)^{j}\\ =& 1-\frac{1}{N_{max}}\left(\frac{1-\left(P_{ext}^{\infty}|_{N=1}\right)^{N_{max}+1}}{1-P_{ext}^{\infty}|_{N=1}}\right)\end{aligned} \tag{10}$$

where $P_{ext}^{\infty}|_{N=1}$ is the probability of extinction $P_{ext}^{\infty}|_{N=1} = \left(1-Q\left(Q_c\gamma-\frac{1}{Q_pB}\right)\right)$, from eq. (7). We use a nonlinear solver to compute N_{max} such that the risk is fixed,

$$\text{Risk }|_{N\sim U(0,N_{max})} = 0.3\%,$$

and the total replication competent fraction $\tilde{Q} = Q_cQ_p$ is fixed, while varying the relative contributions of copying and packaging errors (increasing K from 0 to 1).

The result is shown in Figure 2(c) for different values of $\tilde{Q}$ and R_0 (as before, solid lines give R_0 median, with the shaded area giving the interquartile range (R_0^{25}, R_0^{75}) from [78]. The maximum inoculum size N_{max} required to achieve a risk of infection of 0.3% increases with copying error fraction K, as anticipated. N_{max} also increases with $\tilde{Q}$: since single-virus risk increases monotonically with $\tilde{Q}$, the required maximum inoculum size correspondingly decreases. If instead we assumed a Dirac-delta distribution on the inoculum size around a mean value N_m, i.e.,

$$\text{Risk} = 1-\left(P_{ext}^{\infty}|_{N=1}\right)^{N_m} \Rightarrow N_m = \frac{1-\text{Risk}}{P_{ext}^{\infty}|_{N=1}},$$

we recover qualitatively similar results. Quantitatively, $N_m < N_{max}$, which is not surprising since N_{max} must compensate for the equal-probability, very low inoculum sizes.

Inoculum size predictions are far more sensitive to assumptions on the viral production rate p within biologically reasonable ranges for p, $p = 200-20000$ virions per day, showing order of magnitude differences in prediction, Figure SM1c. But this should be anticipated, since our inoculum size calculation relies on the risk of infection, which also exhibits this sensitivity (see Figure SM1a,b), itself relying on R_0 which is linear in p, $R_0 = Q_1Q_2pkT/(\delta(c+kT))$.

6. On time to infection clearance. We define infection clearance as viral and infected cell clearance, $V = I = 0$. We therefore interpret the time to extinction as the time to infection clearance. Distributions of times to infection clearance may be useful in guiding experiments. For example, when dosing a rhesus macaque with an SIV inoculum (e.g. [38]), systemic infection may not develop. It would be useful to know when one can reasonably assume that a monkey who shows no detectable infection will not develop systemic infection and has cleared all infected cells and virus.

6.1. Calculating the time to infection clearance (extinction). In deriving from our model an expression for the risk of infection eq. (8), in subsection 5.1, we discussed the cumulative probability of infection extinction at time t, given by $P_{ext}(t) = P_{0,0;n_0,v_0}(t) = G_{n_0,v_0}(0,0;t)$ in eq. (5), where $G_{n_0,v_0}(x,y;t)$ is the probability generating function, eq. (2). We assume a

virus-only inoculum; since we assume that the virions behave identically and independently of all other cells and virus in the system, given v infectious virions, the probability of extinction at time t is $P_{ext}(t) = G_{0,1}(0,0;t)^v$ with $G_{0,1}(x,y;t)$ given by the solution of eq. (4).

As in our calculation of risk of infection in subsection 5.1, we assume an inoculum of size N, with each virion replication-competent with probability Q. Then

$$\begin{aligned} P_{ext}(t)|_N = &\sum_{\ell=0}^{N} \{(\text{Probability that the size-}N\text{ inoculum contains }\ell\text{ infectious virions}) \\ &\times (\text{Probability that the infection is extinct at time } t \text{ given} \\ &\quad \text{an inoculum of } \ell \text{ infectious virions})\} \\ = &\sum_{\ell=0}^{N} \left(\binom{N}{\ell} Q^\ell (1-Q)^{N-\ell} \right) \left(G_{0,1}(0,0;t)^\ell \right) \\ = &\left(QG_{0,1}(0,0;t) + (1-Q)\right)^N , \end{aligned} \tag{11}$$

noting again that the sum corresponds to the binomial expansion of $(QG_{0,1}(0,0;t)+(1-Q))^N$. Thus we can calculate the cumulative probability of extinction at time t from the solution of eq. (4) for $x = y = 0$.

Differentiating $P_{ext}(t)|_N$ with respect to t and normalizing gives us the probability density (under the condition that the infection goes extinct),

$$p_{ext}(t) = \frac{1}{\lim_{t\to\infty} P_{ext}(t)|_N} \frac{d\, P_{ext}(t)|_N}{dt}. \tag{12}$$

We can integrate this probability density to compute moments of the time to extinction.

6.2. Copying errors induce longer viral clearance times. We have seen that the single-virus risk of infection is orders of magnitude smaller if we assume that the only mechanism for defective virus production is copying errors. We now assume that infection will clear (so the normalized risk of infection is 0) and investigate the time to infection clearance. The inoculum size assumption does not affect qualitative results since inoculum affects risk and we condition on viral clearance. For convenience we take inoculum size $N = 1000$ virions, each replication competent with probability $Q = 10^{-3}$, in the range of inoculum sizes required to produce a risk of infection of 0.3%, see Figure 2(c).

If we assume that the source of defective virus is solely copying errors, predictions on time to infection clearance are longer in duration than if we assume that the only errors are packaging errors, given that the infection will clear. Figure 3(a) shows the normalized cumulative distribution of clearance times, computed from eq. (11), for median $R_0 = R_0^{\text{med}} = 2.77$. Note that for higher copying error fraction K the tail of the distribution increases. Recall that the probability of extinction for N virions is

$$P_{ext}^{\infty} = \left(1 - \frac{Q}{BQ_p}(R_0 - 1)\right)^N ,$$

from eq. (7) re-written in terms of $R_0 = B\gamma Q_c Q_p$. This probability of extinction increases as $Q_p \uparrow 1$; since we keep the total error $\tilde{Q} = Q_c Q_p$ fixed, $Q_p \uparrow 1$ corresponds to $Q_c \downarrow \tilde{Q}$ (copying

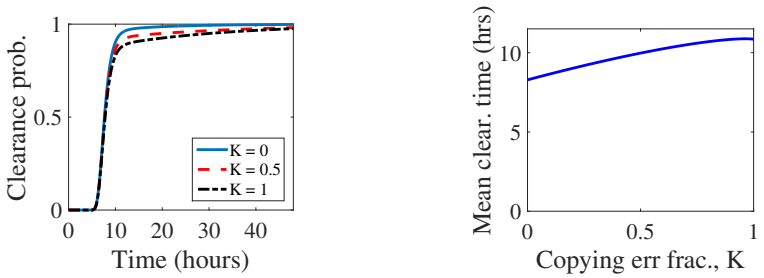

Figure 3. *Viral clearance times conditioned on viral extinction for basic reproduction number R_0 at its median value $R_0^{med} = 2.77$ [78]. Remaining parameters can be found in Table 1.* (a) *Cumulative distribution on time to viral clearance for different copying error fractions, computed from eq. (11) with generating function $G_{0,1}(0,0;t)$ calculated numerically from eq. (4) with zero initial conditions and normalized.* (b) *Mean clearance times computed from associated cumulative distributions. Note that the timescale is in hours.*

error fraction $K \uparrow 1$). The probability of extinction decreases more slowly with the number of virions N if we assume defective virus is mainly attributable to copying errors. Therefore infection extinction, given a few rounds of viral replication yielding circulating virus, is more likely if we assume defective virus arises from copying errors rather than packaging errors. This trend is not altered by the number of circulating infected cells since the probability of extinction arising from an infected cell is $1/R_0$, R_0 is kept fixed. As a result, the tail of the clearance probability distribution - for longer times, after a few rounds of viral replication - is wider. We confirm this result with the probability distribution function for the cumulative number of infected cells, conditioned on infection clearance, see Figure SM3. The probability of accumulating any infected cells is higher if we assume that defective virus are solely attributable to copying errors. We then anticipate that the mean clearance time would increase with the copying error fraction K, as shown in Figure 3(b). There also appears to be a maximum in the clearance time near, but not at, $K = 1$ (Figure 3(b)). This maximum will be addressed in the next section.

6.3. Error assumption changes clearance time dependence on R_0 . In the previous section we discussed clearance time for the basic reproduction number fixed to its median value R_0^{med} from [78]. Extending to R_0 values at the limits of the interquartile range we again see increasing mean clearance times with the copying error fraction K (Figure 4(a)), computed by integrating the associated cumulative distributions computed from eqs. (4) and (11). Intriguingly, however, we find that this increase is not monotonic in R_0: the mean clearance time curves for different values of R_0 in Figure 4(a) intersect.

Figure 4(b) shows contours in mean clearance time (in hours) to better illustrate the non-monoticity, with dashed and solid lines indicating $R_0^{\text{med}} = 2.77$, and the limits of the interquartile range $(R_0^{25}, R_0^{75}) = (2.28, 3.06)$, respectively. In the extreme case of copying error fraction $K = 0$, all defective virus associated with packaging errors, the mean clearance time decreases with basic reproduction number R_0 (Figure 4(c)). As R_0 increases, more new cell infections are engendered, on average, by each infected cell, and the probability of

extinction (eq. (7)) is reduced; after a few rounds of viral replication the viral load grows too large to expect the infection to go extinct. This concept is clear from the clearance time probability density functions, computed from eq. (12), shown in Figure 5(a). Observe that the mode height decreases only very slightly with increasing R_0: the most likely path to extinction is that the initial viral inoculum clears without infecting any target cells, each with rate c, thus the position of the peak doesn't change. But the probability that the initial viral inoculum clears without infecting any target cells, $c/(c+kT) = 1 - R_0/(B\tilde{Q})$ for each virion in the inoculum, decreases with increasing $R_0 < B\tilde{Q}$, and therefore the peak height drops very slightly. However, as R_0 increases, the size of the tail of the clearance time distribution shrinks, because as the viral load grows, the probability of infection extinction decreases. The mean of the distribution must also correspondingly decrease.

In the extreme case of copying error fraction $K = 1$, when all defective virus are attributable to copying errors, the mean clearance time increases with basic reproduction number R_0. This result counters the intuition discussed above. Figure 5(b) shows the clearance time PDFs in the case $K = 1$ and different values of R_0. As with the $K = 0$ case, the mode height decreases slightly: for early times, the clearance time cumulative density function for $K = 0$ and $K = 1$ (Figures 5(a) and (b)) are the same. The most likely scenario is inoculum clearance, which occurs at the same rate and with the same probability. In the $K = 1$ case however the tail of the distribution increases with R_0, see Figure 5(b), and so the mean clearance time increases with R_0. The clearance time probability density tail size increases because the density is computed with the condition that the infection goes extinct. A larger R_0 means that, on average, a larger number of secondary cell infections are induced by a single infected cell and - since $K = 1$ - all these infected cells will produce infectious virus. If the viral inoculum does not immediately clear, there will be more infected cells and virus to clear. Therefore, assuming that the infection does go extinct, clearance takes longer time.

The maximum mean clearance times in Figure 3(a) and Figure 4(a), near but away from $K = 1$, is the result of a transition between regimes where the mean clearance time decreases with R_0 (all packaging errors, $K = 0$) or increases with R_0 (all copying errors, $K = 1$).

7. On time to infection detection.

We have seen that in general, copying errors reduce predicted risk of infection and accelerate viral clearance. We now investigate how error assumptions affect predictions on viral detection, assuming that the infection is not cleared, i.e., does not go extinct. With improvements in technology, HIV tests are becoming more sensitive, and can detect increasingly small amounts of virus [14] and/or virus-associated proteins (for example, p24, an HIV viral capsid protein [89]). Further, there is an increased premium on early detection of HIV infection: early treatment has been shown to improve long-term patient outcomes in terms of quality and length of life [83], and very early treatment is also associated with post-treatment control of HIV [80]. Mathematical modeling predictions on infection detection times can offer insight and guidance into testing windows, that is, the time frame after exposure to HIV within which to get tested and be confident of the positive or negative result.

7.1. Calculating time to infection detection.

In a clinical setting, HIV is measured in a blood sample and is only detectable above a certain threshold, determined by the sensitivity of the assay. We define the probability of detection as the probability that, given some exposure

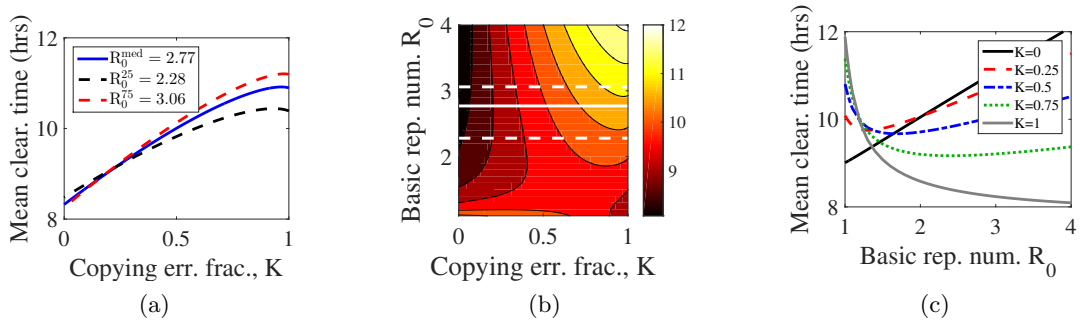

Figure 4. *Viral clearance times as a function of the copying error fraction K and basic reproduction number R_0.* (a-c) *Mean clearance time, in hours,* (a) *as a function of K for R_0 at its median value $R_0^{med} = 2.77$ and at the limits of the interquartile range $(R_0^{25}, R_0^{75}) = (2.28, 3.06)$ [78];* (b) *as contours in (R_0, K), with solid line indicating median R_0^{med} , dashed the interquartile limiting values (R_0^{25}, R_0^{75});* (c) *as a function of R_0 for $K = 0$ (all packaging errors) and $K = 1$ (all copying errors). Remaining parameters can be found in Table 1.*

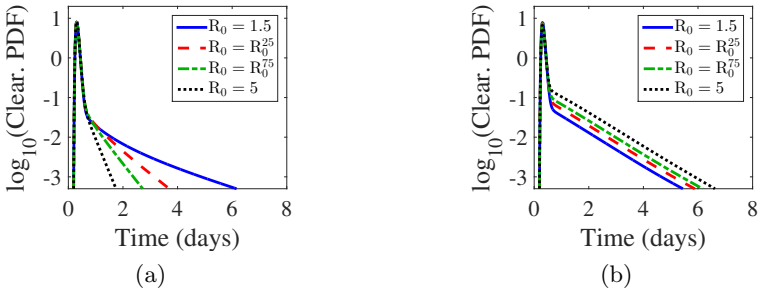

Figure 5. *Probability densities of viral clearance times conditioned on viral extinction for basic reproduction numbers R_0 at interquartile limit values $(R_0^{25}, R_0^{75}) = (2.28, 3.06)$ and smaller/larger values $R_0 = 1.5$ and 5, for copying error fraction* (a) *$K = 0$ and* (b) *$K = 1$. Curves computed using eq. (4), (11), and (12). Remaining parameters can be found in Table 1.*

at time 0, the viral load is above that detection threshold at time t.

For our model, the probability of detection is $P_{\text{det}}(t) = P(V(t) + V_d(t) \geq V_{\text{det}}) = 1 - P(V(t) + V_d(t) < V_{\text{det}})$ where V_{det} is the viral load detection threshold. Here we are assuming that the defective virus V_d does not have deletions that make the virus undetectable by standard viral nucleic acid blood assays [72]. Now,

$$P(V(t) + V_d(t) < V_{\text{det}}) = \sum_{k=0}^{V_{\text{det}}-1} P(V(t) + V_d(t) = k),$$

where $P(V(t) + V_d(t) = k)$ is the probability that the total viral load at time t is k. Note that we have dropped the initial condition for brevity; $P_{\text{det}}(t) = P(V(t) + V_d(t) \geq V_{\text{det}} | I(0) = n_0, V(0) = v_0, I_d(0) = m_0, V_d(0) = w_0)$. The probability of detection involves defective infected cells and virions, which we have not so far included in our calculations.

To calculate the probability of detection we return to the full model in Figure 1(a) and

use

$$P_{n,v,m,w;n_0,v_0,m_0,w_0}(t) = P(I(t)=n, V(t)=v, I_d(t)=m, V_d(t)=w \\ |I(0)=n_0, V(0)=v_0, I_d(0)=m_0, V_d(0)=w_0).$$

Define the corresponding probability generating function as

$$G_{n_0,v_0,m_0,w_0}(x,y,r,s;t) = \sum_{n=0}^{\infty}\sum_{v=0}^{\infty}\sum_{m=0}^{\infty}\sum_{w=0}^{\infty} P_{n,v,m,w;n_0,v_0,m_0,w_0}(t)x^n y^v r^m s^w.$$

But for the additional terms and indices, the derivation of the related bCKde and differential equations for $G_{n_0,v_0,m_0,w_0}(x,y,r,s;t)$ is identical to the derivations in subsection 3.1 and subsection SM3.1.

We can write the probability of detection in terms of the PGF,

$$\begin{aligned} P(V(t)+V_d(t)=k) =& \sum_{j=0}^{k} P(V=j, V_d=k-j) \\ =& \sum_{j=0}^{k} \frac{1}{j!}\frac{1}{(k-j)!}\frac{\partial^k}{\partial r^j \partial s^{k-j}} G_{n_0,v_0,m_0,w_0}\Big|_{x=r=1,y=s=0} \\ =& \sum_{j=0}^{k} \frac{1}{(2\pi)^2}\oint_{C_{z_1}}\oint_{C_{z_2}} \frac{G_{n_0,v_0,m_0,w_0}(1,z_1,1,z_2)}{z_1^{j+1}z_2^{k-j+1}}\,dz_2 dz_1. \\ =& \frac{1}{(2\pi)^2}\oint_{C_{z_1}}\oint_{C_{z_2}} G_{n_0,v_0,m_0,w_0}(1,z_1,1,z_2)\left(\frac{z_2^{-(k+1)}-z_1^{-(k+1)}}{z_1-z_2}\right) dz_2 dz_1. \end{aligned}$$

We used the Cauchy Gauss integral formula [10] to express derivatives as contour integrals for the third step, and summed the finite series in the fourth. Then the probability of not being detected at time t is

$$\begin{aligned} P(V(t)+V_d(t)<V_{\text{det}}) =& \sum_{k=0}^{V_{\text{det}}-1} \frac{1}{(2\pi)^2}\oint_{C_{z_1}}\oint_{C_{z_2}} G_{n_0,v_0,m_0,w_0}(1,z_1,1,z_2;t)\left(\frac{z_2^{-(k+1)}-z_1^{-(k+1)}}{z_1-z_2}\right) dz_2 dz_1 \\ =& \frac{1}{\pi}\mathbb{R}\left\{\int_0^{\pi} G_{n_0,v_0,m_0,w_0}(1,e^{i\theta},1,e^{i\theta};t)\left(\frac{1-e^{-iV_{\text{det}}\theta}}{1-e^{-i\theta}}\right) d\theta\right\} \end{aligned}$$

using the Residue Theorem [10] to reduce the double integral to a single integral, and using the unit circle $e^{i\theta}$ as our contour C_{z_1} (for details see the subsection SM3.2). Finally, if we want to take into account a virus-only inoculum of size N, each virion being infectious with probability Q, again assuming a binomial distribution,

$$P_{\text{det}}(t) = 1 - \sum_{j=0}^{N}\binom{N}{j} Q^j(1-Q)^{N-j}\left(\frac{1}{\pi}\mathbb{R}\left\{\int_0^{\pi} G_{0,j,0,N-j}(1,e^{i\theta},1,e^{i\theta};t)\left(\frac{1-e^{-iV_{\text{det}}\theta}}{1-e^{-i\theta}}\right) d\theta\right\}\right)$$

(13)

$$= 1 - \frac{1}{\pi}\mathbb{R}\left\{\int_0^{\pi}\left(QG_{0,1,0,0}(1,e^{i\theta},1,e^{i\theta};t) + (1-Q)G_{0,0,0,1}(1,e^{i\theta},1,e^{i\theta};t)\right)^N\left(\frac{1-e^{-iV_{\text{det}}\theta}}{1-e^{-i\theta}}\right) d\theta\right\},$$

where we have made use of the branching property

$$G_{n_0,v_0,m_0,w_0} = (G_{1,0,0,0})^{n_0}(G_{0,1,0,0})^{v_0}(G_{0,0,1,0})^{m_0}(G_{0,0,0,1})^{w_0}$$

and the binomial theorem to re-write the series

$$\sum_{j=0}^{N} \left(\begin{array}{c} N \\ j \end{array} \right) (QG_{0,1,0,0})^j \left((1-Q)G_{0,0,0,1}\right)^{N-j} = \left(QG_{0,1,0,0} + (1-Q)G_{0,0,0,1}\right)^N .$$

Note that for our purposes $P_{\text{det}}(t)$ will converge to 1: since we condition on no infection clearance, the viral load will eventually grow exponentially large, and the probability that the infection will go extinct $\to 0$. The detection threshold V_{det}, given by clinical constraints and explicitly quantified below, is very large and well into the exponential phase of infection. We can therefore consider the cumulative probability that the viral load will exceed the detection threshold at time t, $P_{\text{det}}(t)$, to be the probability of infection detection on or before time t.

Clinical investigations on early-time HIV and SIV infections seldom focus on cases where exposure does not result in infection. Data on viral load and CD4+ T-cell counts can only be collected when infection initiates. The probability of detection, above, includes the probability that the infection clears, and therefore may not be useful for some studies. If we only want to consider cases where infection does not clear we must condition the probability $P_{n_0,v_0,m_0,w_0}(t)$ on the infection not going extinct.

The probability of detection at time t conditioned on infection $\tilde{P}_{\text{det|inf}}(t)$ is

$$\begin{aligned} \tilde{P}_{\text{det|inf}}(t) =& P(\text{detection at time } t|\text{infection}) \\ =& \frac{P(\text{detection at time } t)}{P(\text{infection})} \\ =& \frac{P_{\text{det}}(t)}{\text{Risk}}, \end{aligned} \tag{14}$$

from the law of total probability, where $P_{\text{det}}(t)$ is given by eq. (13) and Risk is given by eq. (8). Note that as $t \to \infty$, $P_{\text{det}}(t) \to \text{Risk}$, and therefore $\tilde{P}_{\text{det|inf}}(t) \to 1$, as expected. The probability of not detecting infection at time t, given that infection occurs, is $\tilde{P}_{\text{no det|inf}}(t) = 1 - \tilde{P}_{\text{det|inf}}(t) = 1 - P_{\text{det}}(t)/\text{Risk}$.

Of greater clinical interest perhaps is the conditional risk of infection given an undetectable infection at time t. As tests for HIV viral load are becoming more sophisticated with a lower threshold of detection, this calculation can give insight into testing windows. That is, a blood bank, for example, might want to know times t beyond which this risk is sufficiently small, so that, in essence, an undetectable viral load means no infection. Using Bayes' rule we can write an expression for this conditional risk in terms of the risk of infection and the probability of detection:

$$P(\text{infection}|\text{no detection at time } t) = \frac{P(\text{no detection at time } t|\text{infection})P(\text{infection})}{P(\text{no detection at time } t)}$$

$$= \frac{(1 - P(\text{detection at time } t|\text{infection}))P(\text{infection})}{1 - P(\text{detection at time } t)}$$

$$P(\text{infection}|\text{no detection at time } t) = \left(\frac{1 - \tilde{P}_{\text{det}|\text{inf}}(t)}{1 - P_{\text{det}}(t)}\right) \text{Risk} \tag{15}$$

for $P_{\text{det}}(t)$ given by eq. (13) and $\tilde{P}_{\text{det}|\text{inf}}(t)$ given by eq. (14). Observe that as $t \to \infty$, since $\tilde{P}_{\text{det}|\text{inf}}(t) \to 1$, the probability of infection given no detection at time t goes to 0, again as expected.

7.2. Packaging errors delay infection detection. We assume that the detection threshold is 50 HIV RNA copies per mL, corresponding to the detection threshold for current commercial HIV testing assays. Unlike deterministic dynamics, in which the dynamics measured in 1 mL of intracellular fluid corresponds to the scaled version of full-body viral dynamics, stochastic dynamics do not scale, so we must compute the total viral load in an individual. Assuming an average person's total body extracellular fluid is 15 L, the total viral detection level is $V_{\text{det}} = 750\,000$ HIV RNA copies. We numerically integrate eq. (13) with, and calculate the conditional probability of detection as in eq. (14), with $V_{\text{det}} = 750\,000$, to obtain the probability of infection detection at time t.

Figure 6(a) shows the cumulative probability of the viral load exceeding the detection threshold at time t, conditioned on no extinction eq. (14). Assuming packaging errors only ($K = 0$) yields delayed model-based predictions on detection of infection relative to model predictions relying on copying error assumptions only ($K = 1$), see Figure 6(a). As shown in subsection 5.2, risk of infection decreases with copying error fraction, since for high copying error fractions, newly infected cells are more likely to produce only defective virus. However if we condition on no extinction, we eliminate the cases where the first few infected cells make defective virus (those cases lead to extinction). As a result the infectious viral load increases rapidly and we anticipate a more rapidly spreading infection, and therefore more rapid detection. In mathematical terms, from eq. (14) the probability of detection is normalized by the risk of infection. Smaller risk in the denominator translates to more rapid detection, i.e., higher probability of detection at time t.

Mean detection times decrease monotonically with the copying error fraction, as shown in Figure 6(b): a higher fraction of packaging errors delays the detection time. Figure 6(b) shows mean detection times assuming median R_0^{med} and its interquartile range (R_0^{25}, R_0^{75}). The trend of delayed detection-time predictions due to higher packaging error fraction (decreasing K) remains regardless of R_0. We also observe that as R_0 increases ($R_0^{25} < R_0^{\text{med}} < R_0^{75}$) mean detection times shorten. A larger value of R_0 indicates that each cell engenders, on average, more new cell infections, leading to increased viral load, which will therefore cross the detection threshold sooner, regardless of assumed error types.

As copying errors come to dominate, mean detection time increases, but the variability in detection time decreases. Predictions on variability in detection time are essential when using modeling predictions to gain insight into testing windows; if variability is small, the

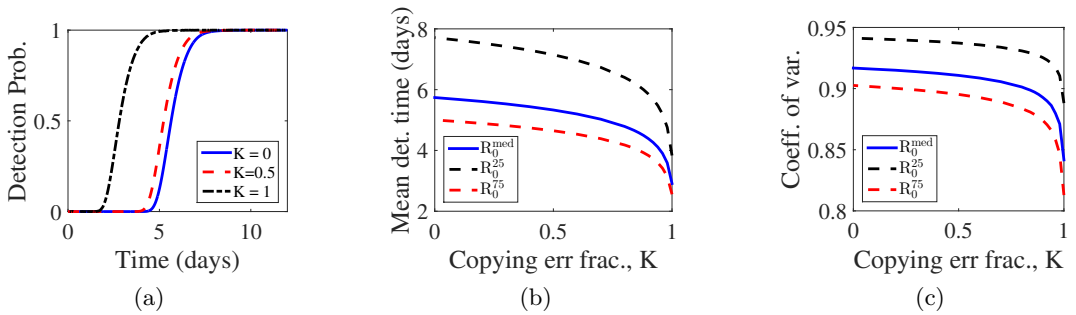

Figure 6. *Probability of detection at time t.* (a) *Cumulative probability of detection at time t for median* $R_0^{med} = 2.77$ *assuming that the copying error fraction is 0 (all packaging errors), 0.5, or 1 (all copying errors).* (b) *Mean detection time as a a function of the copying error fraction, for median basic reproduction number* R_0^{med} *and interquartile values* $(R_0^{25}, R_0^{75}) = (2.28, 3.06)$. (c) *Coefficient of variation in detection time as a a function of the copying error fraction, for median basic reproduction number* R_0^{med} *and interquartile values* (R_0^{25}, R_0^{75}). *Remaining parameters can be found in Table 1.*

mean may be a good guide, but if variability is large, it will not be. We use the coefficient of variation σ/μ (standard deviation over the mean), and plot against the copying error fraction, for median basic reproduction number R_0^{med} and interquartile range values (R_0^{25}, R_0^{75}), as shown in Figure 6(c). The coefficient of variation decreases with copying error, indicating less variability in detection times as copying errors come to dominate, across different R_0 values. Intuitively, decreased variability goes with shorter detection times. That is, there is a smaller time frame in which virus/infected cell "paths" can widen. We find that not only do packaging errors create delayed detection time predictions, they also increase the variability. However we should note that, for our parameters at least, the differences in the coefficient of variation are small - only very near the limiting case of copying error fraction $K = 1$ do the differences approach 10%.

Finally, using our simple model we can comment on the delay time between exposure and HIV testing. We use the basic reproduction number R_0^{25} since it produces the slowest detection times. Figure 7 shows the probability of infection given a negative HIV test, that is, no detection, at time t. The $y-$intercepts in Figure 7(a) gives the total risk of infection; to compare timing we condition on the process not going extinct, as seen in Figure 7(b). Assuming copying error fraction $K = 0$, since again it produces the slowest detection times, our model predicts that, assuming infection is successful, an HIV test should return a positive result at approximately 9.5 days, with 95% probability (Prob(infection — detectable at $t \leq 10$ days)$\approx$ 0.05), given our baseline inoculum size of 1000 virions which gives the high risk of infection of 6.2%. However, if we assume $R_0 = R_0^{75}$ and copying error fraction $K = 1$ the estimate gets as small as 3.6 days (not shown).

Rough empirical estimates for the window period between infection and first detection of HIV RNA lie in the 7-21 day range [16], within the same order of magnitude as our own predictions, albeit higher, understandable since our model neglects some of the complexity of early infection. We can only improve such estimates with better understanding of the different mechanisms underlying viral replication.

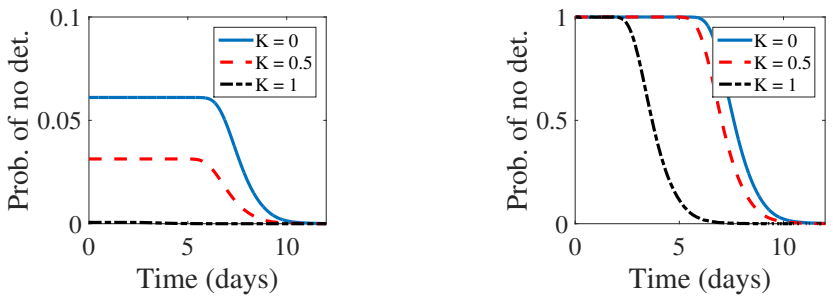

Figure 7. *Probability of infection given no detection at time t for basic reproduction number* $R_0^{25} = 2.28$*, computed numerically from eq.* (15). (a) *Cumulative probability.* (b) *Normalized version of* (a). *Remaining parameters can be found in Table* 1.

675 **8. Discussion.** Elucidating events that occur between exposure to HIV and detectable
676 infection is crucial in developing prevention strategies. Directly investigating these events in
677 humans is very difficult, as are indirect investigations using animal models, since infected cell
678 and viral populations are very small during this period. In this study we developed a simple
679 mathematical model that permits us to make predictions on important clinical characteristics
680 of early HIV infection: risk of infection, time to infection clearance (assuming failed infection),
681 and probability of detection (assuming successful infection).
682 Since cell and viral populations are small during the earliest phases of HIV infection,
683 we used a stochastic modeling approach. We employed continuous-time branching processes,
684 using and extending methods previously presented [17, 18, 68]. In particular we used tools
685 from complex analysis to derive an integral expression for the probability of detection, which
686 is, to our knowledge, a novel calculation.
687 We used our model to investigate the effect of viral replication errors, resulting in non-
688 infectious virus, on early-infection predictions. Estimates on the non-infectious viral fraction
689 are as high as 99.99% [46, 60, 79]. We focused on two mechanisms: (1) that a mutation
690 during reverse transcription fatally cripples the proviral genome so that any viral genomes
691 produced by the provirus will not be infectious [22] (here called copying errors), and (2) that
692 virus may also be rendered non-infectious by errors in the assembly and release phase of
693 viral replication (here called packaging errors), e.g. virions may be packaged with insufficient
694 surface proteins (gp120/gp41) necessary for viral infection or lack essential HIV enzymes such
695 as reverse transcriptase or integrase [51]. In summary, assuming a constant basic reproduction
696 number R_0 and non-infectious viral fraction, we found that:
697 **Risk:** The predicted risk of infection is much higher if we assume the source of non-infectious
698 virus is mainly packaging errors.
699 **Time to clearance:** For failed infections, exposures are predicted to clear more rapidly if
700 we assume the source of non-infectious virus is again mainly packaging errors.
701 **Probability of detection:** For successful infections, the infection is predicted to be de-
702 tectable earlier if we assume the source of non-infectious virus is mainly copying errors.
703 We also uncovered intriguing behavior in our investigation of clearance times. Mean clearance

times, conditioned on failed infection, are predicted to differ only by hours across a wide range of R_0 and copying error fraction K; no modern experiment could distinguish between 8- and 12-hour clearance time in tissues. Nonetheless, these mean clearance times are predicted to decrease with increasing R_0 if one assumes that all non-infectious virus is attributable to packaging errors. Intuitively, this is because we increased R_0 by increasing the infection rate kT. With increased kT, the infection spreads more rapidly, and therefore the associated probability of extinction (clearance) goes to 0 more rapidly. However, the mean clearance time increases with increasing R_0 assuming all non-infectious virus is attributable to copying errors. We traced this counter-intuitive result to the tail of the extinction-time distribution (Figure 5) and our conditioning the probability of clearance on infection extinction.

These results arise because we employed a stochastic model. In the deterministic version of the model (see eq. (SM1) in section SM2) the parameters dictating error type are not identifiable and results are not affected by error type, only total error. The deterministic model is appropriate for chronic infection, when viral and cell populations are very large. Early after exposure to HIV, these populations are small, and it is therefore inappropriate to use a deterministic model. When accounting for stochastic effects it is clear that the mechanism of defective virus production makes a significant and important difference.

Here we focused on errors in the viral replication cycle which lead to non-infectious virus. However there are other mechanisms. For example, neutralizing antibodies may bind free virus, rendering the virus non-infectious. But these events likely only occur in the deterministic limit, after viral populations are large, and our focus in this present study is events in early infection while viral populations are still small. APOBEC3G, an enzyme in cellular anti-viral immunity, is another important factor [33, 40, 43, 88]. APOBEC3G's primary mode of action is to interfere with reverse transcription, inducing hypermutation (copying errors) [43, 88]. Interestingly, infected cells generate APOBEC3G and the enzyme is packaged in the virions budding off the infected cell. Thus, APOBEC3G-induced hypermutation only occurs in cells infected by those budding virions. The HIV protein Vif acts against APOBEC3G, by both triggering its degradation and preventing its incorporation into HIV virions [88]. Thus the anti-viral action of APOBEC3G involves both viral packaging and reverse transcription, and its dynamics cannot be investigated with our simple model. To investigate the dynamics of APOBEC3G we must extend the model, which we leave for future work.

The existence of other mechanisms is one of the limitations of this present study. It is possible non-infectious or defective virus may help drive infection [22], potentially by stimulating an immune response [49] and creating more target cells. It is also possible that copying errors cause defects in the packaging signaling site of the genome, inducing packaging errors: an investigation of infected cells in treated, chronically-infected HIV$^+$ individuals showed that approximately 5% had deletions in the packaging signal portion of the genome [11]. These effects are also not included in our model. Further, our model is very simple, which facilitates the extensive analysis above. However, it does not account for some of the complexity involved in the earliest stages of infection. For example, our estimates for clearance focus on viral populations in the blood, the best indicator for humans. But even in needlestick injury, the virus will move into tissue, and clearance in the blood could therefore precede clearance in tissue. For vaginal sexual exposure, the viral inoculum has to cross vaginal tissues to reach target cells and the bloodstream. Dendritic cells are thought to take up virus and transport

748 them to lymph nodes. This is hypothesized to be one of the triggering events in HIV infec-
749 tion [34], which we previously modeled in [18]. In the very earliest stages, before there is much
750 viral dissemination, target cells may be limiting at particular spatial locations. Therefore, our
751 model is too simple to gain direct insight into animal model observations following vaginal
752 or rectal infection with SIV [52, 76], or to aid in design of such experiments. To address this
753 more complicated scenario, we would apply the mathematics developed in this current work to
754 compute, in particular, time to infection detection and clearance, to an extended model that
755 captures viral transport dynamics, such as [18]. Note that such an extended model would
756 retain considerable uncertainty, since viral dissemination dynamics following HIV exposure
757 remain poorly understood.

758 Even so, our simple model would be an appropriate starting point to investigate and design
759 experiments focusing on intravenous infection of HIV, when the viral inoculum is delivered
760 directly to the blood [19, 81]. Bruner et al. (2016) also showed that 40% of proviruses
761 generated after a single round of in vitro infection were defective [11]. In such studies and
762 experiments, focusing on intravenous infection of HIV, this 40% may inform the fraction of
763 errors associated with copying Q_c, suggesting $Q_c = 0.6$ as an upper bound. The intravenous
764 route of HIV infection is relevant to HIV epidemiology: in 2015, 6-9% of new infections in the
765 United States were associated with injection drug use [12].

766 In spite of limitations, stochastic modeling can be invaluable in investigations of early HIV
767 infection. In a previous study, it was shown that better understanding of viral production
768 would improve risk-of-infection predictions [68]. We have shown that in order to make reliable
769 predictions on risk, clearance time, and detection time, better characterization of viral repli-
770 cation is required. We can then use models such as ours to make practical predictions on HIV
771 testing windows, or to generate theoretical hypotheses on the potential impact of target cell
772 limitation, which we have neglected here but likely plays a role in sexual infection, as there are
773 normally very few $CD4^+$ T cells in the genital mucosa. However, improved stochastic models
774 of early infection also have other uses such as predicting the effect of vaccines on preventing
775 the establishment of infection. It remains to be determined if the mode of defective virus
776 production can impact vaccine efficacy, but this is a topic worthy of investigation.

777 Lastly, the viral dynamics model that we used (Figure 1) as a starting point or slight
778 variants have been used to model many viral infections, such as those due to hepatitis C [63],
779 hepatitis B [84], West Nile virus [6], Zika virus [8] and influenza [4]. As both infectious and
780 noninfectious viruses are produced in all of these infections, the techniques used here and the
781 results we derived should have applicability to these as well as other viral infections.

782 **Acknowledgments.** The authors would like to thank Frederik W. Wiegel, Grant Lythe,
783 and John Pearson for invaluable discussions in the development of this study.

REFERENCES

[1] M. Abram, A. Ferris, W. Shao, W. Alvord, and S. Hughes, *Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication*, J Virol, 84 (2010), pp. 9864–9878.

[2] M. Aid, P. Abbink, R. Larocca, M. Boyd, R. Nityanandam, O. Nanayakkara, A. Martinot, E. Moseley, E. Blass, E. Borducchi, A. Chandrashekar, A. Brinkman, K. Molloy, D. Jetton, L. Tartaglia, J. Liu, K. Best, A. Perelson, R. De La Barrera, M. Lewis, and

D. Barouch, *Zika virus persistence in the central nervous system and lymph nodes of rhesus monkeys*, Cell, 169 (2017), pp. 610–620.e14.

[3] A. Aldovini and R. Young, *Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus*, J Virol, 64 (1990), pp. 1920–1926.

[4] P. Baccam, C. Beauchemin, C. Macken, F. Hayden, and A. Perelson, *Kinetics of influenza A virus infection in humans*, J Virol, 80 (2006), pp. 7590–7599.

[5] R. Bailey, S. Moses, C. Parker, K. Agot, I. Maclean, J. Krieger, C. Williams, R. Campbell, and J. Ninya-Achola, *Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial*, Lancet, 369 (2007), pp. 643–656.

[6] S. Banerjee, J. Guedj, R. Ribeiro, M. Moses, and A. Perelson, *Estimating biologically relevant parameters under uncertainty for experimental within-host murine west nile virus infection*, J R Soc Interface, 13 (2016), p. 20160130.

[7] BC Centre for Disease Control, *HIV laboratory testing: a resource for health professionals*, June 2010.

[8] K. Best, J. Guedj, V. Madelain, X. de Lamballerie, S.-Y. Lim, C. Osuna, J. Whitney, and A. Perelson, *Zika plasma viral dynamics in nonhuman primates provides insights into early infection and antiviral strategies*, Proc Natl Acad Sci USA, 114 (2017), pp. 8847–8852.

[9] A. Bourinbaiar, *The ratio of defective HIV-1 particles to replication-competent infectious virions*, Acta Virol, 38 (1994), pp. 59–61.

[10] J. Brown and R. Churchill, *Complex Variables and Applications*, McGraw-Hill, 8 ed., 2008.

[11] K. M. Bruner, A. J. Murray, R. A. Pollack, M. G. Soliman, S. B. Laskey, A. A. Capoferri, J. Lai, M. C. Strain, S. M. Lada, R. Hoh, Y.-C. Ho, D. D. Richman, S. G. Deeks, J. D. Siliciano, and R. F. Siliciano, *Defective proviruses rapidly accumulate during acute HIV-1 infection*, Nat Med, 22 (2016), pp. 1043–1049.

[12] Centers for Disease Control and Prevention, *HIV Surveillance Report*, vol. 27, http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html. Published November 2016. Accessed 03/31/2017., 2016.

[13] H. Chen, M. Di Mascio, A. Perelson, D. Ho, and L. Zhang, *Determination of virus burst size in vivo using a single-cycle SIV in rhesus macaques*, P Natl Acad Sci USA, 104 (2007), pp. 19079–19084.

[14] A. Cillo, D. Vagratian, M. Bedison, E. Anderson, M. Kearney, E. Fyne, D. Koontz, J. Coffin, M. Piatak, Jr., and J. Mellors, *Improved single-copy assays for quantification of persistent HIV-1 viremia in patients on suppressive antiretroviral therapy*, J Clin Microbiol, 52 (2014), pp. 3944–3951.

[15] H. Clapham, V. Tricou, N. Van Vinh Chau, C. Simmons, and N. Ferguson, *Within-host viral dynamics of dengue serotype 1 infection*, J R Soc Interface, 11 (2014), p. 20140094.

[16] M. S. Cohen, C. L. Gay, M. P. Busch, and F. M. Hecht, *The detection of acute HIV infection*, J Infect Dis, 202 (2018), pp. S270–S277.

[17] J. Conway and D. Coombs, *A stochastic model of latently infected cell reactivation and viral blip generation in treated HIV patients*, PLoS Comput Biol, 7 (2011), p. e1002033.

[18] J. M. Conway, B. Konrad, and D. Coombs, *Stochastic analysis of pre- and post-exposure prophylaxis against HIV infection*, SIAM J Appl Math, 73 (2013), pp. 904–928.

[19] U. Dittmer, C. Stahl-Hennig, C. Coulibaly, T. Nisslein, W. Liike, D. Fuchs, W. Bodemer, H. Petry, and G. Hunsman, *Repeated exposure of rhesus macaques to low doses of simian immunodeficiency virus (SIV) did not protect them against the consequences of a high-dose SIV challenge*, J Gen Virol, 76 (1995), pp. 1307–1315.

[20] N. M. Dixit, M. Markowitz, D. D. Ho, and A. S. Perelson, *Estimates of intracellular delay and average drug efficacy from viral load data of HIV-infected individuals under antiretroviral therapy*, Antivir Ther, 9 (2004), pp. 237–246.

[21] V. Emery, A. Hassan-Walker, A. Burroughs, and P. Griffiths, *Human cytomegalovirus (HCMV) replication dynamics in HCMV-naive and -experienced immunocompromised hosts*, J Infect Dis, 185 (2002), pp. 1723–1728.

[22] D. Finzi, S. Plaeger, and C. Dieffenbach, *Defective virus drives human immunodeficiency virus infection, persistence, and pathogenesis*, Clin Vaccine Immunol, 13 (2006), pp. 715–721.

[23] R. Grant, J. Lama, P. Anderson, V. McMahan, A. Liu, L. Vargas, P. Goicochea, M. Casapía,

J. Guanira-Carranza, M. Ramirez-Cardich, O. Montoya-Herrera, T. Fernández, V. Veloso, S. Buchbinder, S. Chariyalertsak, M. Schechter, L.-G. Bekker, K. Mayer, E. Kallás, K. Amico, K. Mulligan, L. Bushman, R. Hance, C. Ganoza, P. Defechereux, B. Postle, F. Wang, J. McConnell, J.-H. Zheng, J. Lee, J. Rooney, H. Jaffe, A. Martinez, D. Burns, D. Glidden, and for the iPrEx Study Team, *Preexposure chemoprophylaxis for HIV prevention in men who have sex with men*, New Engl J Med, 363 (2010), pp. 2587–2599.

[24] F. Graw and A. Perelson, *Modeling viral spread*, Annu Rev Virol, 3 (2016), pp. 555–572.

[25] A. T. Haase, K. Henry, M. Zupancic, G. Sedgewick, R. A. Faust, H. Melroe, W. Cavert, K. Gebhard, K. Staskus, Z.-Q. Zhang, P. J. Dailey, J. Balfour, H. H., A. Erice, and A. S. Perelson, *Quantitative image analysis of HIV-1 infection in lymphoid tissue*, Science, 274 (1996), pp. 985–989.

[26] P. Haccou, P. Jagers, and V. A. Vatutin, *Branching Processes: Variation, Growth, and Extinction of Populations*, Cam, 2005.

[27] A. Handel, I. Longini Jr., and R. Antia, *Towards a quantitative understanding of the within-host dynamics of influenza A infections*, J R Soc Interface, 7 (2010), pp. 35–47.

[28] T. Harris, *The Theory of Branching Processes*, Dover Publications, New York, USA, 1989.

[29] J. Heffernan and L. Wahl, *Monte Carlo estimates of natural variation in HIV infection*, J Theor Biol, 236 (2005), pp. 137–153.

[30] A. Herz, S. Bonhoeffer, R. Anderson, R. May, and M. Nowak, *Viral dynamics in vivo: limitations on estimates of intracellular delay and virus decay*, Proc Natl Acad Sci USA, 93 (1996), pp. 7247–7251.

[31] Y.-C. Ho, L. Shan, N. Hosmane, J. Wang, S. Laskey, D. Rosenbloom, J. Lai, J. Blankson, J. Siliciano, and R. Siliciano, *Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure*, Cell, 155 (2013), pp. 540–551.

[32] B. Holder and C. Beauchemin, *Exploring the effect of biological delays in kinetic models of influenza within a host or cell culture*, BMC Public Health, 11 (2011), p. S10.

[33] I. Hosseini and F. Mac Gabhann, *Multi-scale modeling of HIV infection in vitro and APOBEC3G-based anti-retroviral therapy*, PLoS Comput Biol, 8 (2012), p. e1002371.

[34] N. Izquierdo-Useros, M. Naranjo-Gmez, I. Erkizia, M. Puertas, F. Borràs, J. Blanco, and J. Martinez-Picado, *HIV and mature dendritic cells: Trojan exosomes riding the Trojan Horse?*, PLoS Pathog, 6 (2010), p. e1000740.

[35] C. Jacob, *Branching processes: Their role in epidemiology*, Int J Environ Res Public Health, 7 (2010), pp. 1186–1204.

[36] M. Johnston and A. Fauci, *An HIV vaccine — evolving concepts*, New Engl J Med, 356 (2007), pp. 2073–2081.

[37] S. Karlin and H. Taylor, *A First Course in Stochastic Processes*, Academic Press, New York, USA, 1975.

[38] B. Keele, H. Li, G. Learn, P. Hraber, E. Giorgi, T. Grayson, C. Sun, Y. Chen, W. Yeh, N. Letvin, J. Mascola, G. Nabel, B. Haynes, T. Bhattacharya, A. Perelson, B. Korber, B. Hahn, and G. Shaw, *Low-dose rectal inoculation of rhesus macaques by SIVsmE660 or SIVmac251 recapitulates human mucosal infection by HIV-1*, J Exp Med, 206 (2009), pp. 1117 – 1134.

[39] B. F. Keele, E. E. Giorgi, J. F. Salazar-Gonzalez, J. M. Decker, K. T. Pham, M. G. Salazar, C. X. Sun, T. Grayson, S. Y. Wang, H. Li, X. P. Wei, C. L. Jiang, J. L. Kirchherr, F. Gao, J. A. Anderson, L. H. Ping, R. Swanstrom, G. D. Tomaras, W. A. Blattner, P. A. Goepfert, J. M. Kilby, M. S. Saag, E. L. Delwart, M. P. Busch, M. S. Cohen, D. C. Montefiori, B. F. Haynes, B. Gaschen, G. S. Athreya, H. Y. Lee, N. Wood, C. Seoighe, A. S. Perelson, T. Bhattacharya, B. T. Korber, B. H. Hahn, and G. M. Shaw, *Identification and characterisation of transmitted and early founder virus envelopes in primary HIV-1 infection*, Proc Natl Acad Sci USA, 105 (2008), pp. 7552–7557.

[40] T. Kikuchi, Y. Iwabu, T. Tada, A. Kawana-Tachikawa, M. Koga, N. Hosoya, S. Nomura, Z. Brumme, H. Jessen, F. Pereyra, A. Trocha, B. Walker, A. Iwamoto, K. Tokunaga, and T. Miura, *Anti-APOBEC3G activity of HIV-1 Vif protein is attenuated in elite controllers*, J Virol, 89 (2015), pp. 4992–5001.

[41] M. Kimmel and D. E. Axelrod, *Branching Processes in Biology*, Springer-Verlag, New York, 2002.

[42] S. Kleinman, N. Lelie, and M. Busch, *Infectivity of human immunodeficiency virus-1, hepatitis C*

virus, and hepatitis B virus and risk of transmission by transfusion, Transfusion, 49 (2009), pp. 2454–2489.

[43] T. Kobayashi, Y. Koizumi, J. Takeuchi, N. Misawa, Y. Kimura, S. Morita, K. Aihara, Y. Koyanagi, S. Iwami, and K. Sato, *Quantification of deaminase activity-dependent and -independent restriction of HIV-1 replication mediated by APOBEC3F and APOBEC3G through experimental-mathematical investigation*, J Virol, 88 (2014), pp. 5881–5887.

[44] M. Kot, *Elements of Mathematical Ecology*, Cambridge University Press, 2001.

[45] S. Kumar and S. Subramanian, *Mutation rates in mammalian genomes*, Proc Natl Acad Sci USA, 99 (2001), pp. 803 – 808.

[46] Y. Kwon, G. Hung, W. Anderson, C.-A. Peng, and H. Yu, *Determination of infectious retrovirus concentration from colony-forming assay with quantitative analysis*, J Virol, 77 (2003), pp. 5712–5720.

[47] R. Landovitz, *Occupational and nonoccupational postexposure prophylaxis for HIV in 2009*, Top HIV Med, 17 (2009), pp. 104–108.

[48] R. Landovitz and J. Currier, *Postexposure prophylaxis for HIV infection*, New Engl J Med, 361 (2009), pp. 1768 – 1775.

[49] M. Larsson, J.-F. Fonteneau, M. Lirvall, P. Haslett, J. Lifson, and N. Bhardwaj, *Activation of HIV-1 specific CD4 and CD8 T cells by human dendritic cells: roles for cross-presentation and non-infectious HIV-1 virus*, AIDS, 16 (2002), pp. 1319–1329.

[50] E. Le Corfec and H. Tuckwell, *Variability in early HIV-1 population dynamics*, AIDS, 12 (1998), pp. 960–962.

[51] J. Levy, *HIV and the Pathogenesis of AIDS*, ASM Press, 3rd ed., 2007.

[52] J. Liu, B. Keele, H. Li, S. Keating, P. Norris, A. Caville, K. Mansfield, G. Tomaras, B. Hynes, D. Kolodkin-Gal, N. Letvin, B. Hahn, G. Shaw, and D. Barouch, *Low-dose mucosal simian immunodeficiency virus infection restricts early replication kinetics and transmitted virus variants in rhesus monkeys*, J Virol, 84 (2010), pp. 10406–10412.

[53] E. G. Luebeck and S. H. Moolgavkar, *Multistage carcinogenesis and the incidence of colorectal cancer*, Proc Natl Acad Sci USA, 99 (2002), pp. 15095–15100.

[54] Z.-M. Ma, M. Stone, M. Piatak, Jr., B. Schweighardt, N. Haigwood, D. Montefiori, J. Lifson, M. Busch, and C. Miller, *High specific infectivity of plasma virus from the pre-ramp-up and ramp-up stages of acute simian immunodeficiency virus infection*, J Virol, 83 (2009), pp. 3288–3297.

[55] C. A. Macken and A. S. Perelson, *Branching Processes Applied to Cell Surface Aggregation Phenomena*, vol. 58 of Lecture Notes in Biomathematics, Springer-Verlag, New York, 1985.

[56] C. A. Macken and A. S. Perelson, *Stem Cell Proliferation and Differentiation: A Multitype Branching Process Model*, vol. 76 of Lecture Notes in Biomathematics, Springer-Verlag, New York, 1988.

[57] L. Mansky, *Forward mutation rate of human immunodeficiency virus type 1 in a T lymphoid cell line*, AIDS Res Hum Retrov, 12 (1996), pp. 307–314.

[58] L. Mansky and H. Temin, *Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase*, J Virol, 69 (1995), pp. 5087 – 5094.

[59] M. Markowitz, M. Louie, A. Hurley, E. Sun, and M. Di Mascio, *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 (2003), pp. 5037–5038.

[60] A. Marozsan, E. Fraundorf, A. Abraha, H. Baird, D. Moore, R. Troyer, I. Nankja, and E. Arts, *Relationships between infectious titer, capsid protein levels, and reverse transcriptase activities of diverse human immunodeficiency virus type 1 isolates*, J Virol, 78 (2004), pp. 11130–11141.

[61] J. Mittler, B. Sulzer, A. Neumann, and A. Perelson, *Influence of delayed viral production on viral dynamics in HIV-1 infected patients*, Math Biosci, 152 (1998), pp. 143–163.

[62] S. H. Moolgavkar and A. G. Knudson Jr., *Mutation and cancer: A model for human carcinogenesis*, J Natl Cancer Inst, 66 (1981), pp. 1037–1052.

[63] A. Neumann, N. Lam, H. Dahari, D. Gretch, T. Wiley, T. Layden, and A. Perelson, *Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy*, Science, 282 (1998), pp. 103–107.

[64] M. Nowak, S. Bonhoeffer, A. Hill, R. Boehme, H. Thomas, and H. McDade, *Viral dynamics in hepatitis B virus infection*, Proc Natl Acad Sci USA, 93 (1996), pp. 4398–4402.

[65] M. Nowak and R. May, *AIDS pathogenesis: mathematical models of HIV and SIV infections*, AIDS, 7

(1993), pp. S3–S18.

[66] C. Osuna, S.-Y. Lim, C. Deleage, B. Griffin, D. Stein, L. Schroeder, R. Omange, K. Best, M. Luo, P. Hraber, H. Andersen-Elyard, E. Ojeda, S. Huang, D. Vanlandingham, S. Higgs, A. Perelson, J. Estes, D. Safronetz, M. Lewis, and J. Whitney, *Zika viral dynamics and shedding in rhesus and cynomolgus macaques*, Nat Med, 22 (2016), pp. 1448–1455.

[67] P. Patel, B. Bennett, T. Sullivan, M. Parker, J. Heffelfinger, P. Sullivan, and for the CDC AHI Study Group, *Rapid HIV screening: Missed opportunities for HIV diagnosis and prevention*, J Clin Virol, 54 (2012), pp. 42–47.

[68] J. Pearson, P. Krapivsky, and A. Perelson, *Stochastic theory of early viral infection: Continuous versus burst production of virions*, PLoS Comput Biol, 7 (2011), p. e1001058.

[69] A. Perelson, P. Essunger, Y. Cao, M. Vesanen, A. Huerley, K. Saksela, M. Markowitz, and D. Ho, *Decay characteristics of HIV-1-infected compartments during combination therapy*, Nature, 387 (1997), pp. 188–191.

[70] A. Perelson and P. Nelson, *Mathematical analysis of HIV-1 in vivo*, SIAM J Appl Math, 41 (1999), pp. 3–44.

[71] A. Perelson, A. Neumann, M. Markowitz, J. Leonard, and D. Ho, *HIV-1 dynamics in vivo: Virion clearance rate, infected cell life-span, and viral generation time*, Science, 271 (1996), pp. 1582 – 1586.

[72] C. Pilcher, S. Fiscus, T. Nguyen, E. Foust, L. Wolf, D. Williams, R. Ashby, J. O'Dowd, J. McPherson, B. Stalzer, L. Hightow, W. Miller, J. Eron, Jr., M. Cohen, and P. Leone, *Detection of acute infections during HIV testing in North Carolina*, New Engl J Med, 352 (2005), pp. 1873–1883.

[73] E. Platt, S. L. Kozak, J. Durnin, T. Hope, and D. Kabat, *Rapid dissociation of HIV-1 from cultured cells severely limits infectivity assays, causes the inactivation ascribed to entry inhibitors, and masks the inherently high level of infectivity of virions*, J Virol, 84 (2010), pp. 3106–3110.

[74] K. Powers, C. Poole, A. Pettifor, and M. Cohen, *Rethinking the heterosexual infectivity of HIV-1: a systematic review and meta-analysis*, Lancet Infect Dis, 8 (2008), pp. 553–563.

[75] B. Ramratnam, S. Bonhoeffer, J. Binley, A. Hurley, and L. Zhang, *Rapid production and clearance of HIV-1 and hepatitis C virus assessed by large volume plasma apheresis*, Lancet Infect Dis, 354 (1999), pp. 1782–1785.

[76] R. Regoes, *The role of exposure history on HIV acquisition: Insights from repeated low-dose challenge studies*, PLoS Comput Biol, 8 (2012), p. e1002767.

[77] J. Reynolds, M. Coles, G. Lythe, and C. Molina-París, *Deterministic and stochastic naive T cell population dynamics: symmetric and asymmetric cell division*, Dynamical Systems, 27 (2012), pp. 75–103.

[78] R. Ribeiro, L. Qin, L. Chavez, D. Li, S. Self, and A. Perelson, *Estimation of the initial viral growth rate and basic reproductive number during acute HIV-1 infection*, J Virol, 84 (2010), pp. 6096–6102.

[79] P. Rusert, M. Fischer, B. Joos, C. Leemann, H. Kuster, M. Flepp, S. Bonhoeffer, H. F. Günthard, and A. Trkola, *Quantification of infectious HIV-1 plasma viral load using a boosted in vitro infection protocol*, Virology, 326 (2004), pp. 113–129.

[80] A. Sáez-Cirión, C. Bacchus, L. Hocqueloux, V. Avettand-Fenoel, I. Girault, C. Lecuroux, V. Potard, P. Versmisse, A. Melard, T. Prazuck, B. Descours, J. Guergnon, J.-P. Viard, F. Boufassa, O. Lambotte, C. Goujard, L. Meyer, D. Costagliola, A. Venet, G. Pancino, B. Autran, C. Rouzioux, and the ANRS VISCONTI Study Group, *Post-treatment HIV-1 controllers with a long-term virological remission after the interruption of early initiated antiretroviral therapy ANRS VISCONTI study*, PLoS Pathog, 9 (2013), p. e1003211.

[81] M. Sutton, C. Burns, A. Weiler, A. Balgeman, A. Braasch, G. Lehrer-Brey, T. Friedrich, and S. O'Connor, *Vaccination with live attenuated simian immunodeficiency virus (SIV) protects from mucosal, but not necessarily intravenous, challenge with a minimally heterologous SIV*, J Virol, 90 (2016), p. 55415548.

[82] W.-Y. Tan and H. Wu, *Stochastic modeling of the dynamics of CD4+ T-cell infection by HIV and some Monte Carlo studies*, Math Biosci, 147 (1998), pp. 173–205.

[83] The INSIGHT START Study Group,, *Initiation of antiretroviral therapy in early asymptomatic HIV infection*, New Engl J Med, 373 (2015), pp. 795–807.

[84] M. Tsiang, J. F. Rooney, J. J. Toole, and C. S. Gibbs, *Biphasic clearance kinetics of hepatitis b virus from patients during adefovir dipivoxil therapy*, Hepatology, 29 (1999), pp. 1863–1869.

[85] H. Tuckwell and E. Le Corfec, *A stochastic model for early HIV-1 population dynamics*, J Theor Biol, 195 (1998), pp. 451–463.

[86] N. Vaidya, R. Ribeiro, C. Miller, and A. Perelson, *Viral dynamics during primary simian immunodeficiency virus infection: Effect of time-dependent virus infectivity*, J Virol, 84 (2010), pp. 4302–4310.

[87] X. Wang, Z. Ao, L. Chen, G. Kobinger, J. Peng, and X. Yaoa, *The cellular antiviral protein APOBEC3G interacts with HIV-1 reverse transcriptase and inhibits its function during viral replication*, J Virol, 86 (2012), pp. 3777–3786.

[88] Y. Wang, B. Kinlock, Q. Shao, T. Turner, and B. Liu, *HIV-1 Vif inhibits G to A hypermutations catalyzed by virus-encapsidated APOBEC3G to maintain HIV-1 infectivity*, Retrovirology, 11 (2014), p. 89.

[89] M. Wu, G.and Swanson, A. Talla, D. Graham, J. Strizki, D. Gorman, R. Barnard, W. Blair, O. Søgaard, M. Tolstrup, L. Østergaard, T. Rasmussen, R.-P. Sekaly, N. Archin, D. Margolis, D. Hazuda, and B. Howell, *HDAC inhibition induces HIV-1 protein and enables immune-based clearance following latency reversal*, JCI Insight, 2 (2017), p. e92901.

[90] L. Zhang, R. Ribeiro, J. Mascola, M. Lewis, G. Stiegler, H. Katinger, A. Perelson, and M. Davenport, *Effects of antibody on viral kinetics in simian/human immunodeficiency virus infection: Implications for vaccination*, J Virol, 78 (2004), pp. 5520–5522.

SUPPLEMENTARY MATERIALS: Early HIV infection predictions: role of viral replication errors*

Jessica M. Conway† and Alan S. Perelson‡

SM1. Supporting figures.

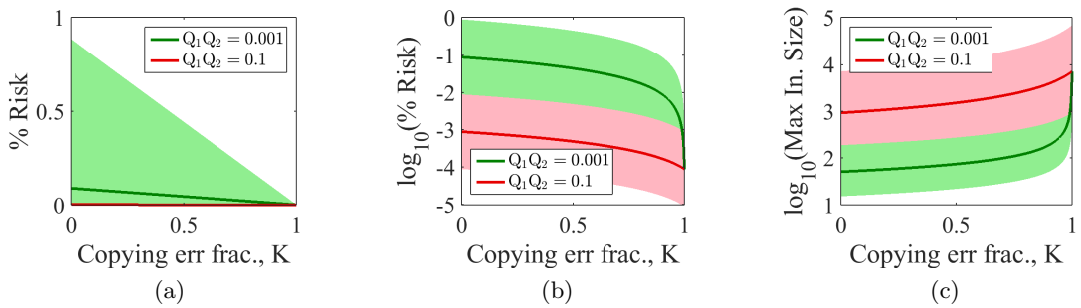

Figure SM1. *Sensitivity of risk of infection and inoculum size to viral production rate p for different total replication competent fractions,* $\tilde{Q} = Q_c Q_p = 10^{-3}$ *and* 10^{-1}. (a-b) *% risk of infection given a single virus inoculum as a function of the fraction of errors attributable to copying errors, on a* (a) *linear and* (b) *log scale.* (c) *Maximum inoculum size assuming a risk of infection of 0.3%, assuming inoculums to be uniformly distributed. Solid lines indicate baseline viral production rate* $p = 2000$ *virions per cell per day, and the shaded areas, within viral production rates* $p = 200$ *and* $p = 20000$ *virions per day. Remaining parameters can be found in Table 1.*

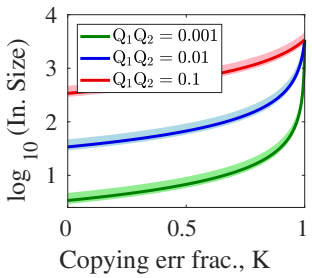

Figure SM2. *Inoculum size prediction assuming a Delta-Dirac distribution. Mean inoculum size assuming a risk of infection of 0.3%, assuming a peaked, Delta-Dirac distribution on inoculum sizes. Solid lines indicate median* $R_0 = 2.77$ *and the shaded areas, risk within the 25th and 75th percentile in* R_0, $(R_0^{25}, R_0^{75}) = (2.28, 3.06)$ *[SM12]. Remaining parameters can be found in Table 1.*

*Resubmitted to the editors X December 2017.

Funding: ASP acknowledges support by National Institutes of Health Grants R01-AI028433 and R01-OD011095.

†Department of Mathematics and Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, Pennsylvania, United States of America (jmconway@psu.edu).

‡Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, New Mexico, United States of America (asp@lanl.gov).

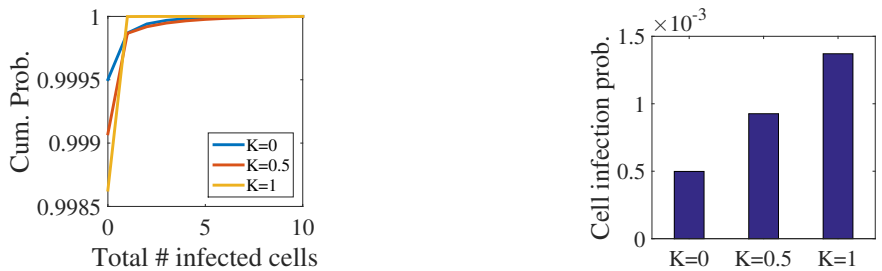

Figure SM3. *Probabilities on number of cells infected in a cleared infection assuming* $R_0 = 2.77$ *for copying error fractions* $K = 0$ *(all packaging errors), 0.5, 1 (all copying errors). (a) Cumulative probability distribution on the number of cells infected in a cleared infection. (b) Probability of any cell infection in a cleared infection. Calculation described in subsection* SM3.3*. Remaining parameters can be found in Table* 1*.*

SM2. Motivation for stochastic approach.

SM2.1. Early infection models: stochastic approach required. The standard approach in viral dynamics modeling is to use ordinary differential equations; the equations one would use to for our model, shown in the main text, Figure 1(a), are

$$\begin{aligned}\frac{dI}{dt} &= Q_c kT\,V - \delta I\\ \frac{dI_d}{dt} &= (1-Q_c)kT\,V - \delta I_d\\ \frac{dV}{dt} &= Q_p pI - (c+kT)V\\ \frac{dV_d}{dt} &= (1-Q_p)pI + pI_d - cV_d.\end{aligned} \tag{SM1}$$

These ODEs are recognizable as the standard viral dynamics model [SM7, SM9, SM11], extended to include defective infected cells and virus, assuming the number of target cells T remain constant. Such deterministic models are suitable when populations of virus and cells are large, since ODEs give average system behavior.

We seek to investigate the following questions, which can only be addressed using a stochastic approach:

1. *Risk of infection:* The basic reproduction number R_0 gives the average number of secondary cell infections caused by the introduction of a single virus into the system; it is used as a measure of severity of infection (or epidemic). Using the next generation method [SM5, SM6] we compute $R_0 = pQ_cQ_pkT/(\delta(c+kT))$. For $R_0 > 1$ the deterministic model eq. (SM1) gives exponential growth only, suggesting a risk of infection of 1. It is well-established (see for example [SM1, SM4, SM8]), and clear from our calculations below, that for stochastic models the probability of extinction for $R_0 > 1$ is non-zero, giving a risk of infection (1-probability of extinction) less than 1.
2. *Time to infection clearance (extinction):* For $R_0 > 1$, the model eq. (SM1) gives no extinction. For $R_0 < 1$, the quantities $(I(t), I_d(t), V(t), V_d(t)) \to 0$ as $t \to \infty$. Traditionally to compute the eradication time using ODEs, one would set some small threshold and claim the extinction occurs when $(I(t), I_d(t), V(t), V_d(t))$ crosses that threshold. The linear system has eigenvalues

$$-c,\ -\delta,\ -\frac{1}{2}(c+\delta) \pm \frac{1}{2}\sqrt{(c-\delta)^2 + 4Q_cQ_pkpT}$$

with decay $\sim \exp\left(-\frac{1}{2}(c+\delta) \pm \frac{1}{2}\sqrt{(c-\delta)^2 + 4Q_cQ_pkpT}\right)$ for $R_0 < 1$. Therefore crossing the threshold in the ODE model depends only on the exponential decay rate. Hence changing dynamic parameters but keeping the decay rate constant will yield identical threshold-crossing times in the ODE model. In stochastic models, the times to clearance differ [SM3]. Stochastic modeling will also yield a distribution of clearance times. Deterministic models are not the ideal approach to investigate clearance.
3. *Probability of detection:* As with the time to extinction, the approach to calculating the time to detection of infection with the ODE model (SM1) would be to set some

threshold, say in the viral load $V + V_d$, and call the time τ that the mean viral load crosses the detection threshold, the mean time to detection. Again, we rely on the system eigenvalues, and find that the viral load expansion will be dominated by the exponential growth rate $\sim \exp\left(-\frac{1}{2}(c+\delta) \pm \frac{1}{2}\sqrt{(c-\delta)^2 + 4Q_cQ_ppkT}\right)$ for $R_0 > 1$. For small amounts of virus, the individual dynamic parameters will play a role and influence the distribution of detection probabilities [SM3,SM8], an effect not captured using exponential growth only. Deterministic models are not the ideal approach to investigate detection.

Therefore if we seek to investigate events in early HIV infection it is necessary to use a stochastic approach.

SM2.2. Identifiability of defective virus fractions. Past models that account for non-infectious virus have focused on chronic HIV infection and employed a deterministic (ODE) approach [SM10]. The focus has mainly been to evaluate efficacy of protease inhibitors, the class of anti-retroviral drugs which interferes with maturation of viral particles. Therefore the theoretical emphases have been on errors in individual virions (which occur in our model with probability 1-Q_p). But practically speaking, the emphasis doesn't matter: the different mechanisms generating defective virus are non-identifiable in deterministic models. In the previous section we discussed R_0 and eigenvalues giving decay or growth - in each of these Q_c and Q_p appear as a product Q_cQ_p only. This non-identifiability is arguably a strength. That is, while the different mechanisms are poorly characterized, we only need know the total non-infectious fraction. With improved understanding individual estimates may change but modeling predictions will remain.

For small viral populations, as is the case in the earliest stages of HIV infection, a stochastic approach is the appropriate choice. When employing a stochastic approach we must be more careful about the underlying mechanisms we model. For example, a previous investigation showed that the assumption on viral production - continuous production by an infected cell or burst production at infected cell death only - will change predictions on risk [SM8]. We will see below that when using a stochastic approach, the individual viral replication error probabilities (lethal copying errors with probability $1 - Q_c$ and packaging errors with probability $1 - Q_p$) will alter predictions on risk, clearance, and probability of detection.

SM3. Details of mathematical derivation.

SM3.1. Backwards Chapman Kolmogorov Differential Equation. In the main text we use a multi-type branching process model of viral dynamics, main text Figure 1. Underpinning the calculations that followed was the backwards Chapman Kolmogorov differential equation (bCKde) for the probability $P_{n,v;n_0,v_0}(t) = P(I(t) = n, V(t) = v | I(0) = n_0, V(0) = v_0)$. We provide here the derivation of the bCKde for the interested reader.

The probability $P(I(t+\tau) = n, V(t+\tau) = v | I(\tau) = n_0, V(\tau) = v_0)$ is a transition probability between the state at time τ, $(I(\tau), V(\tau)) = (n_0, v_0)$, and the state at time $t+\tau$, $(I(t+\tau), V(t+\tau)) = (n, v)$. We assume the probabilities are homogeneous in time, i.e., that they they depend on the duration of the transition time t but not the individual times $t+\tau$ or τ. Therefore, when computing the transition probability, we can directly consider $P_{n,v;n_0,v_0}(t) = P(I(t) = n, V(t) = v | I(0) = n_0, V(0) = v_0)$, setting $\tau = 0$, only.

To model the system in main text Figure 1, we assume the transition probabilities obey the following postulates as $h \downarrow 0$:

1. $P_{n-1,v;n,v}(h) = \delta n h + o(h)$
2. $P_{n,v+1;n,v}(h) = pQ_p n h + o(h)$
3. $P_{n+1,v-1;n,v}(h) = Q_c k T v h + o(h)$
4. $P_{n,v-1;n,v}(h) = (c + (1-Q_c)kT)vh + o(h)$
5. $P_{n,v;n,v}(h) = 1 - ((\delta + pQ_p)n + (c+kT)v)\, h + o(h)$
6. $P_{n,v;n_0,v_0}(h) = \delta_{nn_0}\delta_{vv_0}$

for $n, v \geq 0$, where δ_{jk} is the Kronecker-Delta function. That is, we assume a Poisson process.

Because we assume $P_{n,v;n_0,v_0}(t)$ is homogeneous in time, the Chapman-Kolmogorov equation holds:

$$P_{n,v;n_0,v_0}(t+h) = \sum_{j,k=0}^{\infty} P_{n,v;j,k}(t) P_{j,k;n_0,v_0}(h). \tag{SM2}$$

The $(0, t+h)$ time interval is split into $(0,h)$ and $(h, t+h)$; this equation says that the probability of transitioning from (n_0, v_0) to (n, v) in time $t+h$ is equal to the probability of starting at (n_0, v_0) at 0 and stopping at some midpoint (j,k) at time h, then going from (j,k) to (n,v) in time $t+h$, summed over all possible midpoints (j,k). We can re-write eq. (SM2) as

$$\begin{aligned} P_{n,v;n_0,v_0}(t+h) =& P_{n,v;n_0-1,v_0}(t) P_{n_0-1,v_0;n_0,v_0}(h) + P_{n,v;n_0,v_0+1}(t) P_{n_0,v_0+1;n_0,v_0}(h) \\ &+ P_{n,v;n_0+1,v_0-1}(t) P_{n_0+1,v_0-1;n_0,v_0}(h) + P_{n,v;n_0,v_0-1}(t) P_{n_0,v_0-1;n_0,v_0}(h) \\ &+ P_{n,v;n_0,v_0}(t) P_{n_0,v_0;n_0,v_0}(h) + \left(\sum_{j,k=0}^{\infty} P_{n,v;j,k}(t) P_{j,k;n_0,v_0}(h) \right)'. \end{aligned} \tag{SM3}$$

$\left(\sum_{j,k=0}^{\infty} P_{n,v;j,k}(t) P_{j,k;n_0,v_0}(h)\right)'$ denotes the remaining terms, where j and k do not take on values used in the previous terms. Using the postulates we can show that these remaining

terms are $o(h)$,

$$\begin{aligned}\left(\sum_{j,k=0}^{\infty} P_{n,v;j,k}(t)P_{j,k;n_0,v_0}(h)\right)' \leq& \left(\sum_{j,k=0}^{\infty} P_{j,k;n_0,v_0}(h)\right)'\\ =&1-(P_{n_0-1,v_0;n_0,v_0}(h)+P_{n_0,v_0+1;n_0,v_0}(h)+P_{n_0+1,v_0-1;n_0,v_0}(h)\\ &+P_{n_0,v_0-1;n_0,v_0}(h)+P_{n_0,v_0;n_0,v_0}(h))\\ =&1-(o(h)+\delta n_0 h+pQ_p n_0 h+Q_c kTv_0 h+(c+(1-Q_c)kT)v_0 h\\ &+1-((\delta+pQ_p)n_0+(c+kT)v_0)\,h)\\ =&o(h).\end{aligned}$$

Thus, $\left(\sum_{j,k=0}^{\infty} P_{n,v;j,k}(t)P_{j,k;n_0,v_0}(h)\right)' \leq o(h)$. With this and the postulates we can re-write eq. (SM3) as

$$\begin{aligned}P_{n,v;n_0,v_0}(t+h) =&\delta n_0 h P_{n,v;n_0-1,v_0}(t)+pQ_p n_0 h P_{n,v;n_0,v_0+1}(t)+Q_c kTv_0 h P_{n,v;n_0+1,v_0-1}(t)\\ &+(1-((\delta+pQ_p)n_0+(c+kT)v_0)\,h)\,P_{n,v;n_0,v_0}(t)\\ &+(c+(1-Q_c)kT)v_0 h P_{n,v;n_0,v_0-1}(t)+o(h)\\ =&P_{n,v;n_0,v_0}(t)+h\,(\delta n_0 P_{n,v;n_0-1,v_0}(t)+pQ_p n_0 P_{n,v;n_0,v_0+1}(t)\\ &+Q_c kTv_0 P_{n,v;n_0+1,v_0-1}(t)+(c+(1-Q_c)kT)v_0 P_{n,v;n_0,v_0-1}(t)\\ &-((\delta+pQ_p)n_0+(c+kT)v_0)\,P_{n,v;n_0,v_0}(t))+o(h).\end{aligned}$$

Moving $P_{n,v;n_0,v_0}(t)$ to the left side, dividing by h, and taking the limit $h \to 0$, we obtain the differential equation

$$\begin{aligned}\frac{d}{dt}P_{n,v;n_0,v_0} =&\delta n_0 P_{n,v;n_0-1,v_0}+pQ_p n_0 P_{n,v;n_0,v_0+1}+Q_c kTv_0 P_{n,v;n_0+1,v_0-1}\\ &+(1-Q_c)kTv_0 P_{n,v;n_0,v_0-1}+cv_0 P_{n,v;n_0,v_0-1}\\ &-((\delta+pQ_p)n_0+(kT+c)v_0)\,P_{n,v;n_0,v_0}\end{aligned} \tag{SM4}$$

From postulate 3 we have the initial condition is $P_{n_0,v_0}(0)=\delta_{n,n_0}\delta_{v,v_0}$. This is the backward Chapman Kolmogorov differential equation (bCKde) for the probability $P_{n_0,v_0}(t)$. We cannot solve eq. (SM4) directly analytically, and numerically we can only solve it by simulating exact solution paths, using the SSA (Gillespie) algorithm and plotting the resulting histogram. From eq. (SM4), however, we can derive differential equations for the probability generating function, discussed in the main text and given by eqs. (3,refeq:PGFodes). These will allow us to compute probabilities.

Had we instead incremented time as $(0,t)$ and $(t,t+h)$, using instead the Chapman-Kolmogorov equation

$$P_{n,v;n_0,v_0}(t+h)=\sum_{j,k=0}^{\infty} P_{n,v;j,k}(t+h)P_{j,k;n_0,v_0}(t). \tag{SM5}$$

and followed a similar derivation, we would obtain the corresponding forward Chapman Kolmogorov differential equation for our probability $P_{n_0,v_0}(t)$, also commonly referred to as the

Master Equation. We use the bCKde instead of the more often-used Master equation because the resulting equations for the generating function, are more tractable.

SM3.2. Probability of detection. The probability of detection is $P_{\text{det}}(t) = P(V(t) + V_d(t) \geq V_{\text{det}}) = 1 - P(V(t) + V_d(t) < V_{\text{det}})$, where V_{det} is the viral load detection threshold. Now,

$$P(V(t) + V_d(t) < V_{\text{det}}) = \sum_{k=0}^{V_{\text{det}}-1} P(V(t) + V_d(t) = k),$$

where $P(V(t)+V_d(t) = k)$ is the probability that the viral load at time t is k. Note that we have dropped the initial condition for brevity; $P_{\text{det}}(t) = P(V(t) + V_d(t) \geq V_{\text{det}} | I(0) = n_0, V(0) = v_0, I_d(0) = m_0, V_d(0) = w_0)$. The probability of detection involves defective infected cells and virions, which we have not so far included in our calculations. For the probability of detection we return to the full model in main text Figure 1(a) and use

$$\begin{aligned} P_{n,v,m,w;n_0,v_0,m_0,w_0}(t) = P(&I(t) = n, V(t) = v, I_d(t) = m, V_d(t) = w \\ &| I(0) = n_0, V(0) = v_0, I_d(0) = m_0, V_d(0) = w_0). \end{aligned}$$

Define the corresponding probability generating function as

$$G_{n_0,v_0,m_0,w_0}(x, y, r, s; t) = \sum_{n=0}^{\infty}\sum_{v=0}^{\infty}\sum_{m=0}^{\infty}\sum_{w=0}^{\infty} P_{n,v,m,w;n_0,v_0,m_0,w_0}(t) x^n y^v r^m s^w.$$

But for the additional terms and indices, the derivation of the related bCKde and differential equations for $G_{n_0,v_0,m_0,w_0}(x, y, r, s; t)$ is identical to the derivations in Sec. SM3.1.

We can write the probability of detection in terms of the PGF,

$$\begin{aligned} P(V(t) + V_d(t) = k) &= \sum_{j=0}^{k} P(V = j, V_d = k - j) \\ &= \sum_{j=0}^{k} \frac{1}{j!}\frac{1}{(k-j)!}\frac{\partial^k}{\partial r^j \partial s^{k-j}} G_{n_0,v_0,m_0,w_0}\Big|_{x=r=1,y=s=0} \\ &= \sum_{j=0}^{k} \frac{1}{(2\pi i)^2}\oint_{C_{z_1}}\oint_{C_{z_2}} \frac{G_{n_0,v_0,m_0,w_0}(1, z_1, 1, z_2)}{z_1^{j+1} z_2^{k-j+1}}\, dz_2 dz_1 \\ &= \frac{1}{(2\pi)^2}\oint_{C_{z_1}}\oint_{C_{z_2}} G_{n_0,v_0,m_0,w_0}(1, z_1, 1, z_2)\left(\frac{z_2^{-(k+1)} - z_1^{-(k+1)}}{z_1 - z_2}\right) dz_2 dz_1. \end{aligned}$$

We used the Cauchy Gauss integral formula [SM2] in the third step, and summed the finite

series in the fourth. Then our probability of detection is

$$
\begin{aligned}
P(V(t)+V_d(t)<V_{\text{det}}) &= \sum_{k=0}^{V_{\text{det}}-1} \frac{1}{(2\pi)^2} \oint_{C_{z_1}} \oint_{C_{z_2}} G_{n_0,v_0,m_0,w_0}(1,z_1,1,z_2;t) \left(\frac{z_2^{-(k+1)} - z_1^{-(k+1)}}{z_1 - z_2} \right) dz_2 dz_1 \\
&= \frac{1}{(2\pi)^2} \oint_{C_{z_1}} \oint_{C_{z_2}} G_{n_0,v_0,m_0,w_0}(1,z_1,1,z_2;t) \sum_{k=0}^{V_{\text{det}}-1} \left(\frac{z_2^{-(k+1)} - z_1^{-(k+1)}}{z_1 - z_2} \right) dz_2 dz_1 \\
&= \frac{1}{(2\pi)^2} \oint_{C_{z_1}} \oint_{C_{z_2}} \frac{G_{n_0,v_0,m_0,w_0}(1,z_1,1,z_2;t)}{z_1 - z_2} \left(\frac{1 - z_2^{-V_{\text{det}}}}{z_2 - 1} - \frac{1 - z_1^{-V_{\text{det}}}}{z_1 - 1} \right) dz_2 dz_1 \\
&= -\frac{1}{(2\pi)^2} \oint_{C_{z_1}} \oint_{C_{z_2}} \frac{G_{n_0,v_0,m_0,w_0}(1,z_1,1,z_2;t)}{z_2 - z_1} \left(\frac{1 - z_2^{-V_{\text{det}}}}{z_2 - 1} \right) dz_2 dz_1 \\
&\quad - \frac{1}{(2\pi)^2} \oint_{C_{z_1}} \oint_{C_{z_2}} \frac{G_{n_0,v_0,m_0,w_0}(1,z_1,1,z_2;t)}{z_1 - z_2} \left(\frac{1 - z_1^{-V_{\text{det}}}}{z_1 - 1} \right) dz_2 dz_1.
\end{aligned}
$$

In the last step we split the integrals so that we can evaluate them. We start with the second integral, exchanging the order of integration.

$$
\frac{1}{(2\pi)^2} \oint_{C_{z_2}} \oint_{C_{z_1}} \frac{G_{n_0,v_0,m_0,w_0}(1,z_1,1,z_2;t)}{z_1 - z_2} \left(\frac{1 - z_1^{-V_{\text{det}}}}{z_1 - 1} \right) dz_2 dz_1.
$$

Taking the inner integral only, holding z constant, we note that the contour integral

$$
\oint_{C_{z_1}} \frac{G_{n_0,v_0,m_0,w_0}(1,z_1,1,z_2;t)}{z_1 - z_2} \left(\frac{1 - z_1^{-V_{\text{det}}}}{z_1 - 1} \right) dz_2
$$

is over a function that is analytic everywhere except at $z_1 = z_2$ (singularity at $z_1 = 1$ is removable). Using the residue theorem,

$$
\oint_{C_{z_1}} \frac{G_{n_0,v_0,m_0,w_0}(1,z_1,1,z_2;t)}{z_1 - z_2} \left(\frac{1 - z_1^{-V_{\text{det}}}}{z_1 - 1} \right) dz_2 = \pi i \, G_{n_0,v_0,m_0,w_0}(1,z_2,1,z_2;t) \left(\frac{1 - z_2^{-V_{\text{det}}}}{z_2 - 1} \right)
$$

(note that w is on the contour boundary so the residue is multiplied by πi only) and the second integral becomes

$$
\begin{aligned}
&\frac{1}{(2\pi)^2} \oint_{C_{z_2}} \oint_{C_{z_1}} \frac{G_{n_0,v_0,m_0,w_0}(1,z_1,1,z_2;t)}{z_1 - z_2} \left(\frac{1 - z_1^{-V_{\text{det}}}}{z_1 - 1} \right) dz_2 dz_1 = \\
&\qquad \frac{i}{4\pi} \oint_{C_{z_2}} G_{n_0,v_0,m_0,w_0}(1,z_2,1,z_2;t) \left(\frac{1 - z_2^{-V_{\text{det}}}}{z_2 - 1} \right) dz_2.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\frac{1}{(2\pi)^2} \oint_{C_{z_1}} \oint_{C_{z_2}} \frac{G_{n_0,v_0,m_0,w_0}(1,z_1,1,z_2;t)}{z_2 - z_1} \left(\frac{1 - z_2^{-V_{\text{det}}}}{z_2 - 1} \right) dz_2 dz_1 = \\
&\qquad \frac{i}{4\pi} \oint_{C_{z_1}} G_{n_0,v_0,m_0,w_0}(1,z_1,1,z_1;t) \left(\frac{1 - z_1^{-V_{\text{det}}}}{z_1 - 1} \right) dz_1.
\end{aligned}
$$

Then

$$P(V(t)+V_d(t)<V_{\rm det}) = -\frac{i}{2\pi}\oint_{C_z} G_{n_0,v_0,m_0,w_0}(1,z,1,z;t)\left(\frac{1-z^{-V_{\rm det}}}{z-1}\right)dz$$

The contour C_z is the unit circle, $z=e^{i\theta}$, $dz=ie^{i\theta}d\theta$, and

$$\begin{aligned}P(V(t)+V_d(t)<V_{\rm det}) =& -\frac{i}{2\pi}\int_0^{2\pi} G_{n_0,v_0,m_0,w_0}(1,e^{i\theta},1,e^{i\theta};t)\left(\frac{1-e^{-iV_{\rm det}\theta}}{e^{i\theta}-1}\right)ie^{i\theta}\,d\theta\\ =&\frac{1}{2\pi}\int_0^{2\pi} G_{n_0,v_0,m_0,w_0}(1,e^{i\theta},1,e^{i\theta};t)\left(\frac{1-e^{-iV_{\rm det}\theta}}{1-e^{-i\theta}}\right)d\theta\\ =&\frac{1}{\pi}\mathbb{R}\left\{\int_0^{\pi} G_{n_0,v_0,m_0,w_0}(1,e^{i\theta},1,e^{i\theta};t)\left(\frac{1-e^{-iV_{\rm det}\theta}}{1-e^{-i\theta}}\right)d\theta\right\}.\end{aligned}$$

This result of this integration subtracted from 1 gives us the probability of detection at time t. Finally, if we want to take into account a virus-only inoculum of size N, each virion being infectious with probability Q, again assuming a binomial distribution,

$$\begin{aligned}P_{\rm det}(t) =&1-\sum_{j=0}^{N}\binom{N}{j}Q^j(1-Q)^{N-j}\left(\frac{1}{\pi}\mathbb{R}\left\{\int_0^{\pi} G_{0,j,0,N-j}(1,e^{i\theta},1,e^{i\theta};t)\left(\frac{1-e^{-iV_{\rm det}\theta}}{1-e^{-i\theta}}\right)d\theta\right\}\right)\\ =&1-\sum_{j=0}^{N}\binom{N}{j}Q^j(1-Q)^{N-j}\left(\frac{1}{\pi}\mathbb{R}\left\{\int_0^{\pi}\left(G_{0,1,0,0}(1,e^{i\theta},1,e^{i\theta};t)\right)^j\left(G_4(1,e^{i\theta},1,e^{i\theta};t)\right)^{N-j}\left(\frac{1-e^{-iV_{\rm det}\theta}}{1-e^{-i\theta}}\right)d\theta\right\}\right)\\ =&1-\frac{1}{\pi}\mathbb{R}\left\{\int_0^{\pi}\left(QG_{0,1,0,0}(1,e^{i\theta},1,e^{i\theta};t)+(1-Q)G_4(1,e^{i\theta},1,e^{i\theta};t)\right)^N\left(\frac{1-e^{-iV_{\rm det}\theta}}{1-e^{-i\theta}}\right)d\theta\right\},\end{aligned}\tag{SM6}$$

where we have made use of the branching property,

$$G_{n_0,v_0,m_0,w_0} = (G_{1,0,0,0})^{n_0}(G_{0,1,0,0})^{v_0}(G_{0,0,1,0})^{m_0}(G_{0,0,0,1})^{w_0}.$$

SM3.3. Probability distribution on the number of cumulative number of cell infections.

In Supporting Figure S3 we show the probability distribution on the cumulative number of cell infections conditioned on infection clearance. Here we provide details of that calculation.

In order to count the number of cell infections, we extend our model to include the state variable C which increases with every cell infection and does not otherwise alter infection dynamics. Define $P_{n,v,c;n_0,v_0,c_0}(t) = P(I(t)=n, V(t)=v, C(t)=c | I(0)=n_0, V(0)=v_0, C(0)=c_0)$. Note that $c_0=0$ for the purposes of calculation below.

We then assume that the transition probabilities obey the following postulates as $h\downarrow 0$:

1. $P_{n-1,v,c;n,v,c}(h) = \delta nh + o(h)$
2. $P_{n,v+1,c;n,v,c}(h) = pQ_pnh + o(h)$
3. $P_{n+1,v-1,c+1;n,v,c}(h) = Q_ckTvh + o(h)$
4. $P_{n,v-1,c+1;n,v,c}(h) = (1-Q_c)kTvh + o(h)$
5. $P_{n,v-1,c;n,v,c}(h) = cvh + o(h)$
6. $P_{n,v,c;n,v,c}(h) = 1-((\delta+pQ_p)n + (c+kT)v)\,h + o(h)$
7. $P_{n,v,c;n_0,v_0,c_0}(h) = \delta_{nn_0}\delta_{vv_0}\delta_{cc_0}$

for $n,v,c\geq 0$, where δ_{jk} is the Kronecker-Delta function.

Following the derivation in subsection SM3.1, the bCKde for the probability $P_{n,v,c;n_0,v_0,c_0}$, associated with this process, is

$$\begin{aligned}\frac{d}{dt}P_{n,v,c;n_0,v_0,c_0} =&\delta n_0 P_{n,v,c;n_0-1,v_0,c_0} + pQ_p n_0 P_{n,v,c;n_0,v_0+1,c_0}\\ &+ Q_c kT v_0 P_{n,v,c;n_0+1,v_0-1,c_0+1} + (1-Q_c)kT v_0 P_{n,v,c;n_0,v_0-1,c_0+1}\\ &+ cv_0 P_{n,v,c;n_0,v_0-1,c_0} - ((\delta + pQ_p)n_0 + (kT+c)v_0)\, P_{n,v,c;n_0,v_0,c_0}.\end{aligned} \tag{SM7}$$

We then define the generating function

$$H_{n_0,v_0,c_0}(x,y,r;t) = \sum_{n=0}^{\infty}\sum_{v=0}^{\infty}\sum_{c=0}^{\infty} P_{n,v,c;n_0,v_0,c_0}(t)x^n y^v r^c$$

and following subsection 3.1 in the main text, we can derive from (SM7) equations for the probability generating function,

$$\begin{aligned}\frac{\partial H_{1,0,0}}{\partial t} =&\delta + pQ_p H_{1,0,0}H_{0,1,0} - (\delta + pQ_p)H_{1,0,0}\\ \frac{\partial H_{0,1,0}}{\partial t} =&c + kT\left((1-Q_c) + Q_c H_{1,0,0}\right)H_{0,0,1} - (c+kT)H_{0,1,0}\\ \frac{\partial H_{0,0,1}}{\partial t} =&0\end{aligned} \tag{SM8}$$

where $H_{n_0,v_0,c_0} = H_{1,0,0}^{n_0}H_{0,1,0}^{v_0}H_{0,0,1}^{c_0}$. Using the probability generating function we can compute the probability for the cumulative number of cell infections.

We want the cumulative number of cell infections as $t \to \infty$ conditioned on infection clearance. To get at this quantity, we will take the limit

$$\lim_{t\to\infty} P_{0,0,c;0,v_0,0}(t),$$

that is, the limit as $t \to \infty$ of the probability that there are c cumulative cell infections but no infected cells or virus. From the pgf H_{n_0,v_0,c_0},

$$\begin{aligned}P_{0,0,c;0,v_0,0}(t) &= \frac{1}{c!}\frac{\partial^c}{\partial r^c}H_{n_0,v_0,c_0}(0,0,r;t)|_{r=0}\\ &= \oint_{C_z}\frac{H_{n_0,v_0,c_0}(0,0,z;t)}{z^{c+1}}\,dz\\ &= \frac{1}{2\pi}\int_0^{2\pi} H_{n_0,v_0,c_0}(0,0,e^{i\theta};t)e^{-ic\theta}\,d\theta\end{aligned}$$

using the Cauchy Gauss integral formula [SM2] in the second step, and take the contour C_z as the unit circle $z = e^{i\theta}$ in the third. Noting that $H_{n_0,v_0,c_0} = H_{1,0,0}^{n_0}H_{0,1,0}^{v_0}H_{0,0,1}^{c_0}$ with $c_0 = 0$, and that we take a virus-only inoculum, so $n_0 = 0$, and simplifying, we recover

$$P_{0,0,c;0,v_0,0}(t) = \frac{1}{\pi}\int_0^{\pi}(H_{0,1,0}(0,0,e^{i\theta};t))^{v_0}e^{-ic\theta}\,d\theta.$$

Finally, as in subsection 3.1, we assume that our inoculum of size N is mixed, with each virus infectious with probability Q. Therefore the probability of accumulating c cell infections at time t, with no circulating virus or infected cells, is

$$\begin{aligned}\text{Prob} &= \sum_{j=0}^{N} \binom{N}{j} (1-Q)^{N-j} Q^j P_{0,0,c;0,j,0}(t) \\ &= \sum_{j=0}^{N} \binom{N}{j} (1-Q)^{N-j} Q^j \frac{1}{\pi} \int_0^\pi (H_{0,1,0}(0,0,e^{i\theta};t))^j e^{-ic\theta}\, d\theta \\ &= \frac{1}{\pi} \int_0^\pi \Big(1 - Q + Q H_{0,1,0}(0,0,e^{i\theta};t)\Big)^N e^{-ic\theta}\, d\theta, \end{aligned} \tag{SM9}$$

summing the series in the last step.

Thus to compute the probability of accumulating c infected cells before clearing infection as shown in Figure SM3, we compute eq. SM9 together with eq. SM8 numerically, evaluating the limit by computing over long times and verifying convergence, and normalizing to condition on infection clearance.

REFERENCES

[1] F. Ball, *The threshold behaviour of epidemic models*, J Appl Probab, 20 (1983), pp. 227–241.
[2] J. Brown and R. Churchill, *Complex Variables and Applications*, McGraw-Hill, 8 ed., 2008.
[3] J. Conway and D. Coombs, *A stochastic model of latently infected cell reactivation and viral blip generation in treated HIV patients*, PLoS Comput Biol, 7 (2011), p. e1002033.
[4] J. M. Conway, B. Konrad, and D. Coombs, *Stochastic analysis of pre- and post-exposure prophylaxis against HIV infection*, SIAM J Appl Math, 73 (2013), pp. 904–928.
[5] O. Diekmann and J. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*, Wiley Series in Mathematical & Computational Biology (Book 5), Wiley, 2000.
[6] J. Heffernan, R. Smith?, and L. Wahl, *Perspectives on the basic reproductive ratio*, J R Soc Interface, 2 (2005), pp. 281–293.
[7] M. Nowak and R. May, *Virus Dynamics: Mathematical Principles of Immunology and Virology*, Oxford University Press, USA, 2001.
[8] J. Pearson, P. Krapivsky, and A. Perelson, *Stochastic theory of early viral infection: Continuous versus burst production of virions*, PLoS Comput Biol, 7 (2011), p. e1001058.
[9] A. Perelson, P. Essunger, Y. Cao, M. Vesanen, A. Huerley, K. Saksela, M. Markowitz, and D. Ho, *Decay characteristics of HIV-1-infected compartments during combination therapy*, Nature, 387 (1997), pp. 188–191.
[10] A. Perelson and P. Nelson, *Mathematical analysis of HIV-1 in vivo*, SIAM J Appl Math, 41 (1999), pp. 3–44.
[11] A. Perelson, A. Neumann, M. Markowitz, J. Leonard, and D. Ho, *HIV-1 dynamics in vivo: Virion clearance rate, infected cell life-span, and viral generation time*, Science, 271 (1996), pp. 1582 – 1586.
[12] R. Ribeiro, L. Qin, L. Chavez, D. Li, S. Self, and A. Perelson, *Estimation of the initial viral growth rate and basic reproductive number during acute HIV-1 infection*, J Virol, 84 (2010), pp. 6096–6102.