Predicting the outcomes of treatment to eradicate the latent reservoir for HIV-1

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Predicting the outcomes of treatment to eradicate the latent reservoir for HIV-1

Alison L. Hill\textsuperscript{a,b,1,2}, Daniel I. S. Rosenbloom\textsuperscript{a,c,1}, Feng Fu\textsuperscript{d}, Martin A. Nowak\textsuperscript{e}, and Robert F. Siliciano\textsuperscript{a,f}

\textsuperscript{a}Program for Evolutionary Dynamics, Department of Mathematics, and Department of Organismic and Evolutionary Biology, and \textsuperscript{b}Biophysics Program and Harvard-MIT Division of Health Sciences and Technology, Harvard University, Cambridge, MA 02138; \textsuperscript{c}Department of Biomedical Informatics, Columbia University Medical Center, New York, NY 10032; \textsuperscript{d}Institute of Integrative Biology, Eidgenössische Technische Hochschule Zürich, 8092 Zurich, Switzerland; and \textsuperscript{e}Department of Medicine and \textsuperscript{f}Howard Hughes Medical Institute, The Johns Hopkins University School of Medicine, Baltimore, MD 21205

Edited by John M. Coffin, Tufts University School of Medicine, Boston, MA, and approved July 9, 2014 (received for review April 12, 2014)

Massive research efforts are now underway to develop a cure for HIV infection, allowing patients to discontinue lifelong combination antiretroviral therapy (ART). New latency-reversing agents (LRAs) may be able to purge the persistent reservoir of latent virus in resting memory CD4\textsuperscript{+} T cells, but the degree of reservoir reduction needed for cure remains unknown. Here we use a stochastic model of infection dynamics to estimate the efficacy of LRA needed to prevent viral rebound after ART interruption. We incorporate clinical data to estimate population-level parameter distributions and outcomes. Our findings suggest that \(~2,000\)-fold reductions are required to permit a majority of patients to interrupt ART for 1 year without rebound and that rebound may occur suddenly after multiple years. Greater than 10,000-fold reductions may be required to prevent rebound altogether. Our results predict large variation in rebound times following LRA therapy, which will complicate clinical management. This model provides benchmarks for moving LRAs from the laboratory to the clinic and can aid in the design and interpretation of clinical trials. These results also apply to other interventions to reduce the latent reservoir and can explain the observed return of viremia after months of apparent cure in recent bone marrow transplant recipients and an immediately-treated neonate.

Significance

HIV infection cannot be cured by current antiretroviral drugs, due to the presence of long-lived latently infected cells. New antilatency drugs are being tested in clinical trials, but major unknowns remain. It is unclear how much latent virus must be eliminated for a cure, which remains difficult to answer empirically due to few case studies and limited sensitivity of viral reservoir assays. In this paper, we introduce a mathematical model of HIV dynamics to calculate the likelihood and timing of viral rebound following antilatency treatment. We derive predictions for the required efficacy of antilatency drugs, and demonstrate that rebound times may be highly variable and occur after years of remission. These results will aid in designing and interpreting HIV cure studies.


Edited by John M. Coffin, Tufts University School of Medicine, Boston, MA, and approved July 9, 2014 (received for review April 12, 2014)
The expression $gAP\frac{qAP}{\delta}$ greatly exceeds Fig. S1 greatly exceeds 200 copies per mL. We computed the relationship approximates the expected number of SI Materials and Methods (the ratio of post-LRA to pre-LRA reservoir $\delta^{1}$ A $\delta^{2}$, the rate at which the LR remains after therapy, which determines the initial conditions of the model. The stochastic model of viral dynamics following interruption of ART and LRA tracks both latently infected resting CD4+ T cells (rectangles) and productively infected CD4+ T cells (ovals). Each arrow represents an event that occurs in the model. Alternate models considering homeostatic proliferation and turnover of the LR are discussed in SI Materials and Methods. Viral rebound occurs if at least one remaining cell survives long enough to activate and produce a chain of infection events leading to detectable infection (plasma HIV-1 RNA >200 copies per mL).

active infection of some number of other cells, or it can die from other causes without producing virions that infect other cells. In the latter case, cytotoxic T lymphocyte (CTL) killing, errors in viral reverse transcription, or other problems upstream of virion production may prevent further infection. The model only tracks the initial stages of viral rebound, when target cells are not yet limited. A full description of the model is provided in Materials and Methods and SI Materials and Methods.

The initial conditions for the dynamic model depend on the number of latently infected cells left after LRA therapy. LRA efficacy is defined by the fraction $q$ of the LR that remains following treatment. The model tracks each latent and active cell to determine whether viral rebound occurs, and if so, how long it takes. Importantly, no single activated cell is guaranteed to reestablish the infection, as it may die before infecting other cells. Even if it does infect others, those cells likewise may die before completing further infection. This possibility is a general property of stochastic models, and the specific value for the establishment probability depends on the rates at which infection and death events occur. Our goal is to calculate the probability that at least one of the infected cells remaining after therapy escapes extinction and causes viral rebound, and if so, how long it takes. If all cells die, then rebound never occurs and a cure is achieved. As the model only describes events after completion of LRA therapy, our results are independent of the therapy protocol or mechanism of action.

Using both stochastic simulations and theoretical analysis of this model, we find that the probability and timing of rebound relies on four key parameters: the decay rate of the LR in the absence of viral replication ($\delta$), the rate at which the LR produces actively infected cells ($A$), the probability that any one activated cell will produce a rebounding infection before its lineage dies ($P_{E0}$), and the net growth rate of the infection once restarted ($r$). Estimates of these four parameters are provided in Table 1 and Fig. S1. After therapy, the rate at which the LR produces actively infected cells is reduced to $qA$. The probability that an individual successfully clears the infection is

$$P_{CR}(q) \approx e^\frac{-4qAP}{\delta^2}.$$  \[1\]

The expression $gAP\frac{qAP}{\delta}$ approximates the expected number of the product of the pre-LRA reservoir size and the per-cell activation rate), and $q$ (the ratio of post-LRA to pre-LRA reservoir size), it does not independently influence outcomes. Both of these formulas provide an excellent match to the explicit simulation of the model (Fig. 2). The key assumption required for the analysis is that $r$ greatly exceeds $\delta$; because viral doubling times during rebound are measured on the order of a few days, whereas LR decay is measured on the order of many months or years, this assumption is expected to hold. Likelihood-based inference can therefore proceed by efficient computation of rebound probabilities (using [86]), rather than by time-consuming stochastic simulation.

Outcomes depend only on the four parameters above even in more complex models of viral dynamics, including other features of T-cell biology and the HIV lifecycle (SI Materials and Methods). Alternate models studied include explicit tracking of free virus with varying burst sizes, an eclipse phase during which an infected cell produces no virus, proliferation of cells upon reactivation, maintenance of the LR by homeostatic proliferation, and either a constant or Poisson-distributed number of infected cells produced by each cell (Figs. S2–S7). If proliferation of latently infected cells is subject to high variability, e.g., by bursts of proliferation, then rebound time and cure probability increase slightly beyond the predictions of the basic model (Figs. S6 and S7). No other modification to the model altered outcomes. Outcomes of LRA therapy therefore are likely to be insensitive to details of the viral lifecycle; accordingly, few parameters must be estimated to predict outcomes.

**Predicted Prospects for Eradicating Infection or Delaying Time to Rebound.** Using best estimates of parameters derived from previously reported data (Table 1), we can explore the likely outcomes of interventions that reduce the LR. The best outcome of LRA therapy, short of complete and immediate eradication, is that so few latently infected cells survive that none reactivate and start a resurgent infection during the patient’s lifespan. In this case, LRA has essentially cleared the infection and cure is achieved. We simulated the model to predict the relationship between LRA efficacy and clearance (Fig. 2A). We find that the LR must be reduced 10,000-fold before half of patients are predicted to clear the infection.

If LRA therapy fails to clear the infection, the next-best outcome is extension of the time until rebound, defined as plasma HIV-1 RNA ≥ 200 copies per mL. We computed the relationship between LRA efficacy and median time until rebound among patients who do not clear the infection (Fig. 2B). Roughly a 2,000-fold reduction in LR size is needed for median rebound times of 1 y. Only modest (about 2-fold) increases in median rebound time are predicted for up to 100-fold reductions in LR size. In this range, the rebound time is independent of latent cell lifespan (decay rate $\delta$) and is driven mainly by the reactivation rate ($A$) and the infection growth rate ($r$). The curve inflects upward (on a log scale) at ~100-fold reduction and eventually reaches a ceiling as...
cleansing of the infection becomes the dominant outcome. The upward inflection results from a change in the forces governing viral dynamics. If the reservoir is large (little reduction), then cells activate frequently, and the dominant component of rebound time is the time that it takes for virus from the many available activated cells to grow exponentially to rebound levels; the system is in a growth-limited regime. If the reservoir is small (large reduction), the dominant component is instead the expected waiting time until activation of the first cell fated to establish a rebounding lineage; the system is in an activation-limited regime. Because waiting time is roughly exponentially distributed, times to rebound in this regime can vary widely among patients on the same therapy, even with identical values of the underlying parameters.

Survival curves, plotting the fraction of simulated patients maintaining virologic suppression over time, demonstrate the extreme interpatient variability and long follow-up times required for LRA therapy (Fig. 2C). For less than 100-fold reductions in LR size, simulated patients uniformly rebound within a few months because rebound dynamics are not in the activation-limited regime. If therapy decreases LR size 1,000-fold, then ~55% of patients are predicted to delay rebound for at least 6 mo. However, of these patients, 47% suffer rebound in the following 6 mo. Higher reservoir reductions lead to clearance in many patients. In others, rebound may still occur after years of apparent cure, posing a challenge for patient management.

Earlier work suggested a shorter reservoir half-life of 6 mo (35), indicating that dramatic decreases in LR size would occur after 5 y or more of suppressive ART even in the absence of LRA therapy. We consider the prospects for HIV eradication or long treatment interruptions with this faster decay rate. In this optimistic scenario, only 1,500-fold reductions are needed for half of patients to clear the LR, and rebound becomes highly unlikely after a few years. Alternatively, in a worst-case scenario where latent cell death is perfectly balanced by homeostatic proliferation such that the reservoir does not decay at all (δ = 0), much higher efficacies are needed to achieve beneficial patient outcomes (Fig. 3).

**Setting Treatment Goals with Uncertainty Considerations.** We conducted a full uncertainty analysis of the model, by simultaneously varying all parameters over their entire ranges (Table 1 and Fig. S1). For each simulated patient, values for the three parameters δ, A, and r were sampled independently from their respective distributions, whereas \( P_{\text{Ext}} \) was sampled from a conditional distribution that depends on \( r \) (Materials and Methods). Results for this simulated cohort are similar to those for the point estimates, with greater interpatient variation in outcomes (Fig. 3A). This variation makes the survival curves less steep: Cure is slightly more likely at low efficacy, but slightly less likely at high efficacy. As expected from Eq. 1, cure is more likely for patients with lower A or \( P_{\text{Ext}} \) values and higher δ values. If therapy provides only 10 to 100-fold LR reductions, a subset of patients may delay rebound for several months.

Using these cohort-level predictions, we can set efficacy goals for the reservoir reduction needed to achieve a particular likelihood of a desired patient outcome. Fig. 4 provides target LRA efficacies for which 50% of patients are predicted to remain rebound-free for a specified interruption time. Reductions of under 10-fold afford patients only a few weeks to a month off treatment without rebound. For 1-y interruptions, a 1,000–3,000-fold reduction is needed. To achieve the goal of eradication (cure) 4-log reduction is required. This value increases to 4.8 logs to cure 75% of patients, and to 5.8 logs for 95% of patients.

**Model Applications and Comparison with Data.** The current ability to test the model against clinical data is limited both by the dynamic range of assays measuring LR size and by the low efficacy of investigational LRA treatments. However, we can compare our predictions to results observed for non-LRA-based interventions that lead to smaller LR size and prolonged treatment interruptions (Fig. 4). A 2010 study of early ART initiators who eventually underwent treatment interruption found a single patient with LR size ~1,500-fold lower than a typical patient (0.0064 infectious units per million resting CD4+ T cells versus an average of one per million) in whom rebound was delayed until 50 d off treatment (36). The well-known “Berlin patient” (22) has remained off treatment following a stem cell transplant since 2008, and a comprehensive analysis of his viral reservoirs found HIV DNA levels at least 7,500-fold lower than typical patients in the most sensitive assay (37). The two recently reported “Boston patients” also interrupted treatment, following transplants causing at least a 3 to 4 log decrease in viral reservoirs (23); they have since both rebounded, at ~3 and 8 mo post interruption. In the case of the “Mississippi baby”, infection was discovered and treated within 30 h of birth, and ART continued until interruption at around 18 mo. Virus remained undetectable for 27 mo, when viral rebound occurred, assuming the accuracy of widely reported claims (e.g., ref. 38). At the time of treatment cessation, the LR size was likely at least 300-fold lower than that of a typical adult [based on less than 0.017 infectious units per million resting CD4+ T cells at age 30 mo (39), and scaled on a weight basis relative to adults]. These few available cases demonstrate that our model is not inconsistent

![Fig. 2.](image-url)

**Table 1.** Estimated values for the key parameters of the stochastic viral dynamics model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Estimation method</th>
<th>Ref(s).</th>
<th>Best estimate</th>
<th>Distribution*</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR decay rate</td>
<td>( \delta )</td>
<td>Long-term ART, ( \delta = \ln(2)/12 )</td>
<td>(6, 7)</td>
<td>( 5.2 \times 10^{-4} \text{ d}^{-1} )</td>
<td>( \delta \sim \mathcal{N}(5.2, 5.6) \times 10^{-4} \text{ d}^{-1} )</td>
</tr>
<tr>
<td>LR exit rate</td>
<td>A</td>
<td>Viral rebound after ART interruption</td>
<td>(8, 69)</td>
<td>57 cells per day</td>
<td>( \log_{10}(A) \sim \mathcal{N}(1.76, 1.0) )</td>
</tr>
<tr>
<td>Growth rate</td>
<td>( r )</td>
<td>Population genetic modeling</td>
<td>(71, 72)</td>
<td>0.4 d^{-1}</td>
<td>( \log_{10}(r) \sim \mathcal{N}(-0.40, 0.19) )</td>
</tr>
<tr>
<td>Establishment probability</td>
<td>( P_{\text{Ext}} )</td>
<td>Population genetic modeling</td>
<td>(71, 72)</td>
<td>0.069</td>
<td>Composite distribution</td>
</tr>
</tbody>
</table>

*Notation \( X \sim \mathcal{N}(\mu, \sigma) \) means that \( X \) is a random variable drawn from a normal distribution with mean \( \mu \) and SD \( \sigma \).
with current knowledge. When survival curves for larger cohorts become available, Bayesian methods can be used to update estimates in Table 1 and reduce uncertainty of future predictions.

Discussion

Our model is, to our knowledge, the first to quantify the required efficacy of LRAAs for HIV-1 and set goals for therapy. For a wide range of parameters, we find that therapies must reduce the LR by at least 2 orders of magnitude to meaningfully increase time to rebound after ART interruption (upward inflection in Figs. 2B and 3 A, II; B, II; C, II), and that reductions of approximately 4 orders of magnitude are needed for half of patients to clear the infection (Figs. 3 A, I; B, I; C, I; and 4). Standard deviations in rebound times of many months are expected, owing to substantial variation in reactivation times after effective LRA therapy brings the infection to an activation-limited regime.

Though the efficacy required for these beneficial outcomes likely exceed the reach of current drugs, our results permit some optimism: We show for the first time, to our knowledge, that reactivation of all cells in the reservoir is not necessary for cessation of ART. This is because some cells in the LR will die before reactivating or, following activation, will fail to produce a chain of infections leading to rebound. On a more cautionary note, the wide distribution in reactivation times necessitates careful monitoring of patients, as rebound may occur even after long periods of viral suppression.

Even without any reservoir reduction, variation in infection parameters and chance activation together predict delays in rebound of at least 2 mo in a small minority of patients (Figs. 3 A, III and 4), consistent with ART interruption trials such as SPARTAC (40). More detailed (and possibly more speculative) models including immune responses may be needed to explain multiyear posttreatment control, such as seen in the VISCONTI cohort (24).

Our analysis characterizing the required efficacy of LRA therapy does not rely on the specific mechanism of action of these drugs, only the amount by which they reduce the reservoir. We have assumed that, after ART/LRA therapy ends, cell activation and death rates return to baseline. We have also assumed that the reservoir is a homogeneous population with constant activation and death rates. The presence of reservoir compartments with different levels of LR penetration does not alter our results, as they are stated in terms of total reservoir reduction. If, however, these compartments vary in activation or death rates (41), or if dynamics of activated cells depends on their source compartment, then our model may need to be modified. Moreover, if spatial population structure affects viral replication, viral dynamics above the detection limit (from which we estimated parameters $r$ and $A$) may not correspond straightforwardly to the infection/death rates in early infection, due to local limitations in target cell density (42). Spatial restrictions on viral transmission may be particularly important in densely packed lymphoid tissue (43). Without a clear understanding of multiple compartments constituting the LR, we have considered the simplest scenario which may fit future LRA therapy outcomes.

Throughout this paper, we assume that combination ART is sufficiently effective so that viral replication alone cannot sustain the infection after all latent virus is cleared. Studies of treatment intensification (44, 45), viral evolution during ART (46, 47), and in vitro antiviral efficacy (48, 49) all support this assumption. Moreover, HIV persistence is widely believed to result solely from the long lifespan or proliferative ability of latently infected cells (3, 50). If this assumption is violated, e.g., by the presence of latency-reserved-drug-protected compartments (43, 51, 52), then any curative strategy predicted solely on latency reversal would be futile.

Our model also highlights the importance of measuring specific parameters describing latency and infection dynamics. Despite the field’s focus on measuring LR size with increasing accuracy (53), our results suggest that the rate at which latently infected cells activate, and the fraction of these that are expected to establish a rebounding infection, are more predictive of LRA outcomes. Among all parameters that determine outcome, the establishment probability is least understood, as it cannot be measured from viral load dynamics above the detection limit. Simply because an integrated provirus is replication competent and transcriptionally active does not mean that it will initiate a growing infection: As with all population dynamics, chance events dominate early stages of infection growth (34, 54). HIV-1 transcription is itself a stochastic process, governed by fluctuating concentrations of early gene products (55). Sensitive assays of viral outgrowth may pave the way toward understanding the importance of these chance events to early infection; for instance, fluorescent imaging studies of adenovirus have shown that a large majority of in vitro infections seeded by single productively infected cells die out early, before rapid growth and plaque formation can occur (54). Keeping other parameters constant, assuming a worst-case (highest) value for the establishment probability raises the reservoir reductions required for cure or a desired extended rebound time by 0.8 logs. Regardless of the exact probability, the stochastic nature of HIV-1 activation and infection dynamics implies that even similarly situated patients may experience divergent responses to LRA.

The model can also advise aspects of trial design for LRAs. Survival curves computed from [S8] can be used to predict the probability that a patient is cured, given that they have been off

---

Fig. 3. Predicted LRA therapy outcomes, accounting for uncertainty in patient parameter values. (A) Full uncertainty analysis where all viral dynamics parameters are sampled for each patient from the distributions provided in Table 1. (B) A best-case scenario where the reservoir half-life is only 6 mo ($\delta = 3.8 \times 10^{-3} d^{-1}$). All patients have the same underlying viral dynamic parameters, otherwise given by the point estimates in Table 1. (C) A worst-case scenario where the reservoir does not decay because cell death is balanced by homeostatic proliferation ($\delta = 0$). (I) Probability that the LR is cleared by LRA. Clearance occurs if all cells in the LR die before a reactivating lineage leads to viral rebound. LRA log-efficacy is the number of orders of magnitude by which the LR size is reduced following LRA therapy, $-\log_{10}(\delta/q)$. (II) Median viral rebound times (logarithmic scale) among patients who do not clear the infection. (III) Survival curves (Kaplan–Meier plots) show the percentage of patients who have not yet experienced viral rebound, plotted as a function of the time (logarithmic scale) after treatment interruption. All simulations included $10^3$ to $10^6$ patients.
Fig. 4. Efficacies required for successful LRA therapy. The target LRA log-ef
ciency is the treatment level (in terms of log reduction in LR size) for which at least 50% of patients still have suppressed viral load after a given treat
ment interruption length (blue line). Shaded ranges show the results for the middle 50% (dark gray) and 90% (light gray) of patients. Lifetime indicates
the LR is cleared. Annotations on the curve represent data points for case
studies describing large reservoir reductions and observing rebound times
after ART interruption. From left to right, they represent a case of early ART
initiation in an adult [the “Chun patient” (C) (36)], two cases of hematopa
tic stem cell transplant with wild-type donor cells [the two Boston
patients (Bo.1 and Bo.2) (23)], a case of early ART initiation in an infant [the
Mississippi baby (Mi.) (25), assuming, as recently reported, rebound after
27 mo], and a case of hematopoietic stem cell transplant with Δ32 CCSR donor
cells [the Berlin patient (Be.) (22, 37)]. For the Chun patient, the annotations
represent the maximum likelihood estimate for LR reduction (diamond), as
well as 95% confidence intervals (vertical bar). For the Boston, Berlin, and
Mississippi patients, vertical arrows indicate that only a lower bound on
treatment efficacy is known (LR size was below the detection limit) and that
the true value may extend further in the direction shown. For the Berlin
patient, the horizontal arrow indicates that rebound time is at least 5 y
(rebound has not yet occurred).

Treatment interruption goal
1 wk 1 mth 3 mths 1 yr 10 yrs Lifetime
0 1 2 3 4
Target LR Log-reduction

In this notition, Y and Z represent individual actively or latently infected
cells, respectively, Θ represents no cells, and the arrows represent one type
of cell becoming the other type. A latently infected cell can either activate
(at rate a) or die (at rate d). An actively infected cell can either die (at rate
d) or produce a collection of virions (at rate b) that results in the infection of
c other cells, where c is a Poisson-distributed random variable with param-
eter λ(p,c) = (exp(−−)κ)(λ). After an infection event, the original cell dies.
Each event occurs independently within a large, constant target cell popu-
lation. As the model does not include limitations on viral growth, it
describes only the initial stages of viral rebound. Because clinical rebound
thresholds (plasma HIV RNA >500 copies per mL) are well below typical
set points (108 to 1010 copies per mL), this model suffices to analyze re-
bound following LRA therapy and ART interruption. We do not explicitly
track free virus, but assume it to be proportional to the number of
infected cells. This assumption is valid because rates governing produc-
tion and clearance of free virus greatly exceed other rates, allowing a separa-
tion of time scales. As we are not interested in blips or other intradyal viral
dynamics, this assumption does not influence our results. A method for
calculating the proportionality between free virus and infected cells is
provided in SI Materials and Methods.

The growth rate of the infection is r = bλ(c) − d. The total death rate of
infected cells is d = b + d, and the basic reproductive ratio (mean offspring
for a single infected cell) is R0 = bλ(c)/d. The establishment prob-
ability P0,0 is the solution to R0(1 − e−ϕP0,0) − P0,0 = 0. The total LR decay rate in
the absence of viral replication is λc = cλ. If there are Z cells in the LR, then
the number of cells reactivating per day is A = zλ.

Analysis of the model to determine the four key parameters (a, A, r, P0,0)
and rapidly compute survival curves is provided in SI Materials and Methods.
A script for computation of survival curves is also provided at www.
danielrosenbloom.com/reboundtimes.

Parameter Estimation. The half-life of latently infected cells has been esti-
- mated to be approximately r−1 = 44 mo (3, 4). The resulting value of b =
ln(2)/r−1, ln(2)/44, is centered at 5.2 × 10−4 d−1, and we construct a distribution of
values based on ref. 3. This value represents the net rate of LR decay
during suppressive therapy, considering activation, death, homeostatic pro-
- liferation, and (presumably rare) events when activated CD4+ T cells reenter
a memory state. The net infection growth rate r describes the rate of ex-
ponential increase in viral load once infection has been reseeded. The LR
reactivation rate A is the number of cells exiting the LR per day, before LRA
therapy. A and r were jointly estimated from the dynamics of viral load
during treatment interruption trials in which there was no additional reser-
voir-reducing intervention (5, 58); in particular, infection growth imme-
diately following rebound is sensitive to r, whereas the time to rebound is
sensitive to A. Absent reservoir reduction, observed rebound dynamics are
innocuous to P0,0 and so this estimated P0,0 is free from population genetic
models (59, 60) relating observed rates of selective sweeps and emergence of
drug resistance to variance in the viral offspring distri-
bution (SI Materials and Methods).

Simulation of the Model. We use the Gillespie algorithm to track the num-
- ber of latently and actively infected cells in a continuous time stochastic
- process. The initial number of latent cells is Z(0) = Binomial(N0, q), where N0 is
pre-treatment LR size and q is efficacy of LRA treatment (fraction of cells
remaining). The initial number of actively infected cells Y(0) is then chosen
from a Poisson distribution with parameter dA/Z(0)/dY (corresponding to the
immigration–death equilibrium of the branching process). The simulation
proceeds until the number of actively infected cells reaches the threshold for
clinical detection given by a viral load of 200 copies per mL (equivalent to
Y = 3 × 105 cells total) or until no active or latent cells remain. Because sto-
- chastic effects are important only for small Y, we switch to faster deter-
ministic numerical integration when Y reaches a level where extinction
probability is very low (<10−5). For each q value we perform 104 to 105
- simulations.

Simulations are seeded with values of the key parameters (b, A, r, P0,0),
which may be either the point estimates or random numbers sampled from
the distributions in Table 1. We then back out values of model-specific
parameters consistent with the sampled key parameters. In general, we use
a pretherapy LR size of N0 as 106 cells to get a = A/N0. We then have d = a.a. As detailed in SI Materials and Methods, sampling P0,0 requires first

Materials and Methods

Basic Stochastic Model. The basic model of reservoir dynamics and rebound
tracks two cell types: productively infected activated CD4+ T cells and latently
infected resting CD4+ T cells. The model can be described formally as a two-
type branching process, in which four types of events can occur (Fig. 1):

Z → Y rate constant: a
Z → Θ rate constant: d
Y → CF rate constant: b × p(c)
Y → Θ rate constant: d

δ

with

Y = Y(0) + C + F + Θ
Z = Z(0) + A + B + Θ
C = C + A
F = F + B

Each event occurs independently within a large, constant target cell popu-
lation. As the model does not include limitations on viral growth, it
describes only the initial stages of viral rebound. Because clinical rebound
thresholds (plasma HIV RNA >500 copies per mL) are well below typical
set points (108 to 1010 copies per mL), this model suffices to analyze re-
bound following LRA therapy and ART interruption.
sampling the variance-to-mean ratio of the viral offspring distribution. Then using this relationship, we can estimate 

\[ \text{VRM} = \frac{\text{VRM}_d + \text{VRM}_d}{2} \]

as a measure of the viral burden in the reservoir. We find that this estimate is consistent with the results of simulations, which suggest that the viral reservoir is 

\[ \text{VRM}_d = \text{VRM}_d + \text{VRM}_d \]

for all values of the viral offspring distribution. Therefore, we conclude that the viral reservoir is a significant contributor to the viral burden in the reservoir.

**ACKNOWLEDGMENTS.**

We thank Y.-C. Ho, S. A. Rabi, L. Shan, and G. Laird for insightful discussions and for sharing data, and we thank A. Perelson for helpful comments on an earlier version of this manuscript. This work was supported by the Martin Delaney Collaboratory of AIDS Researchers for Eradication and the Delaney AIDS Research Enterprise Collaboratory (National Institutes of Health Grants AI096113 and U19AI096109), Grant 108165-50-RGRL from the American Foundation for AIDS Research Consortium on HIV Eradication, the Johns Hopkins Center for AIDS Research, and the Howard Hughes Medical Institute. M.A.N. was supported by the John Templeton Foundation, F.F. was funded by the European Research Council Advanced Grant PBDR268540, and A.L.H. and D.I.S.R. were supported by Bill and Melinda Gates Foundation Grand Challenges Explorations Grant OPP1044503.