

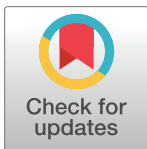
## RESEARCH ARTICLE

# Models of SIV rebound after treatment interruption that involve multiple reactivation events

Christiaan H. van Dorp<sup>1\*</sup>, Jessica M. Conway<sup>2</sup>, Dan H. Barouch<sup>3,4</sup>, James B. Whitney<sup>3,4</sup>, Alan S. Perelson<sup>1</sup>

**1** Theoretical Biology and Biophysics (T-6), Los Alamos National Laboratory, Los Alamos, New Mexico, United States of America, **2** Department of Mathematics and Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, Pennsylvania, United States of America, **3** Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, United States of America, **4** Ragon Institute of MGH, MIT, and Harvard, Cambridge, Massachusetts, United States of America

\* [chvandorp@gmail.com](mailto:chvandorp@gmail.com)



## OPEN ACCESS

**Citation:** van Dorp CH, Conway JM, Barouch DH, Whitney JB, Perelson AS (2020) Models of SIV rebound after treatment interruption that involve multiple reactivation events. PLoS Comput Biol 16(10): e1008241. <https://doi.org/10.1371/journal.pcbi.1008241>

**Editor:** Miles P. Davenport, UNSW Australia, AUSTRALIA

**Received:** February 3, 2020

**Accepted:** August 12, 2020

**Published:** October 1, 2020

**Copyright:** This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

**Data Availability Statement:** All scripts and data are available on GitHub (<https://github.com/lan/multiple-reactivation-model>).

**Funding:** Portions of this work were done under the auspices of the U.S. Department of Energy under contract 89233218CNA000001 and supported by National Institutes of Health ([www.nih.gov](https://www.nih.gov)) grants P01-AI131365 (JBW); R01-AI028433 and R01-OD011095 (ASP); AI124377, AI126603, and AI128751 (DHB); and National Science Foundation ([www.nsf.gov](https://www.nsf.gov)) grant DMS-

## Abstract

In order to assess the efficacy of novel HIV-1 treatments leading to a functional cure, the time to viral rebound is frequently used as a surrogate endpoint. The longer the time to viral rebound, the more efficacious the therapy. In support of such an approach, mathematical models serve as a connection between the size of the latent reservoir and the time to HIV-1 rebound after treatment interruption. The simplest of such models assumes that a single successful latent cell reactivation event leads to observable viremia after a period of exponential viral growth. Here we consider a generalization developed by Pinkevych *et al.* and Hill *et al.* of this simple model in which multiple reactivation events can occur, each contributing to the exponential growth of the viral load. We formalize and improve the previous derivation of the dynamics predicted by this model, and use the model to estimate relevant biological parameters from SIV rebound data. We confirm a previously described effect of very early antiretroviral therapy (ART) initiation on the rate of recrudescence and the viral load growth rate after treatment interruption. We find that every day ART initiation is delayed results in a 39% increase in the recrudescence rate (95% credible interval: [18%, 62%]), and a 11% decrease of the viral growth rate (95% credible interval: [4%, 20%]). We show that when viral rebound occurs early relative to the viral load doubling time, a model with multiple successful reactivation events fits the data better than a model with only a single successful reactivation event.

## Author summary

HIV-1 persists during suppressive antiretroviral therapy (ART) due to a reservoir of latently infected cells. When ART is stopped, HIV generally rebounds within a few weeks. However, there is a small fraction of patients that do not rebound over a period of months or years. A variety of treatments are being tested for their ability to reduce the size of the

1714654 (JMC). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

latent reservoir, to induce effective immune responses against the virus, or to prevent or prolong the time to viral rebound after ART interruption. These novel treatments are typically first tested in SIV infected macaques, and the efficacy of the treatment assessed by interrupting ART and measuring the time to viral rebound. Here, we develop and test a mathematical and statistical model that describes the process of viral rebound. The model can be used for statistical inference of the efficacy of newly developed treatments. Importantly, the model takes into account that multiple recrudescence events can precede rebound. We test the model using data from early treated SIV infected macaques.

## Introduction

HIV and SIV are able to persist despite antiretroviral therapy (ART) because of a long-lived reservoir of latently infected CD4<sup>+</sup> T cells [1]. Recent studies have shown that the latent reservoir is established very early after infection [2–4], and that the seeding of the reservoir can only be prevented when ART starts extremely early [5]. Other studies have focused on the effect of potentially curative treatment strategies that might extend remission after interruption of ART [6–8].

In all these studies an important observable is the time between treatment interruption and viral rebound, i.e. the first time the viral load (VL) becomes observable. Under the common assumption that rebound results from reactivation of latently infected cells [9–11], and that the rate at which the latent population reactivates is proportional to the size of the latent reservoir, the time to viral rebound can be used to gauge the reservoir size. Some curative strategies aim to reduce the size of the reservoir by administering latency reversing agents such as vorinostat [12], romidepsin [13], and TLR7 agonists [8], but also gene editing [14], so-called block-and-lock strategies [15], and anti-proliferative therapy [16, 17] are being considered. The time to rebound can then be used as an indication of the effectiveness of the treatment, consistent with the aforementioned assumption [9, 18, 19].

The simplest model of rebound combines an exponentially distributed waiting time for a recrudescence event with subsequent exponential growth of the VL (henceforth, this is referred to as the “single-reactivation model”). Such a model has been used to estimate the reactivation rate of cells from the reservoir in HIV-1 patients undergoing ART interruption [10]. The main conclusion of this study—reactivation occurs on average every 5–8 days—resulted in some discussion about the sensitivity of the aforementioned result to inter-patient variability of the model parameters [20, 21]. From this discussion, an interesting and slightly more complex model of viral rebound emerged [20, 21] that takes into account the possibility that multiple latently infected cells reactivate within a short time interval, and that each of these reactivation events contributes to VL growth (we hereafter refer to this model as the “multiple-reactivation model”). We refer to a reactivation event that leads to an exponentially growing and potentially observable lineage of actively infected cells as a “successful reactivation event” or “recrudescence event”, since reactivation can also lead to extinction of the lineage by chance [11, 18, 22, 23].

The occurrence of multiple recrudescence events is not merely a theoretical hypothesis, but has recently been observed *in vivo*. In one study, phylogenetic analysis has revealed that HIV-1 rebound is seeded from multiple anatomical sites [24]. In another study, treatment interruption experiments with macaques infected with a genetically barcoded SIV strain showed that many cells successfully reactivate from the latent reservoir [25]. In the latter study, the multiple-reactivation model was used to analyze the viral rebound data [25, 26], underpinning the

current interest in this model. Moreover, in a recent analysis of potentially curative treatment effects the multiple-reactivation model was used as a bridge between stochastic and deterministic reactivation domains [19]. Here we present an improvement of the multiple-reactivation model that we derive using a Poisson counting process. Although the average behavior of our improved model is only marginally different from the previous version, our approach allows us to not only model the expected viral load rebound curve, but also the deviation from this expectation. Most importantly, this enables us to derive a parametric expression for the distribution of the time-to-rebound, that can be used for parameter inference from rebound data.

We test the improved model using data from SIV infected macaques that are put on ART at different times post infection and exhibit varying viral rebound dynamics [2, 5]. We find very strong statistical evidence in favor of the multiple-reactivation model over the single-reactivation model. We attribute this superior model performance to the fact that it better explains the data from macaques that rebound soon after ART cessation and exhibit relatively slow exponential growth of the VL. We argue that whenever such data is used for inference about the effects of experimental curative treatments in delaying viral rebound, the multiple-reactivation model should be used to estimate the relevant parameters. Our refined multiple-reactivation model fits the data only slightly better than the approximation developed earlier by Pinkevych *et al.* [21]. However, using an example, we show that our model can be generalized further to include more complex features of reservoir and rebound dynamics, such as heterogeneity of the reservoir in terms of clone-specific growth rates.

## Results

We start by mathematically defining the multiple-reactivation model and deriving the mean behavior and deviation from the mean of this model. We then use these quantities to derive an approximate probability distribution of the time to viral rebound, and assess whether this approximation is reliable. This time-to-rebound distribution is then used to infer the rate of recrudescence from a heterogeneous set of SIV rebound data. This inference allows us to quantify the effect of ART initiation time on the recrudescence rate and viral growth rate, and to compare our multiple-reactivation model with the simpler single-reactivation model. We identify two mechanisms that make the multiple-reactivation model better suited for modeling rebound data than the single-reactivation model. Finally, using simulated data sets, we test how sensitive the model is to parameter and model misspecification.

### The multiple- and single-reactivation models

We start by constructing a model that predicts the short-term SIV or HIV viral dynamics following the cessation of ART including viral rebound to detectable viremia and subsequent exponential growth of the VL. In our modeling we rely on the common, central assumption that activation of latently infected cells drives viral rebound [9–11]. Specifically, we assume that the activation of a latently infected cell can be followed by viral production, which in turn may lead to infection of additional cells. Viral rebound is caused by exponential growth in resultant viral lineages. We refer to a latent cell reactivation that leads to exponential growth as a “successful reactivation event” or a “recrudescence event”, to explicitly make the distinction with reactivation events leading to a viral lineage that by chance goes extinct while the population size is still small. We provide an overview of the models we employ in this study with full details in Materials and methods. A synopsis of the parameters and variables used is given in Table 1.

Mathematically, the multiple-reactivation model is a combination of a stochastic Poisson counting process  $N_t$  with rate or intensity  $\lambda$  and deterministic exponential viral growth  $v_0 e^{gt}$

Table 1. An overview of the parameters and variables.

symbol	unit	description
$N_t$	-	number of recrudescence events at time $t$ post treatment interruption
$V_t$	copies mL <sup>-1</sup>	VL at time $t$ post treatment interruption
$T_i$	d	time of the $i$ -th successful reactivation event.
$g$	d <sup>-1</sup>	exponential growth rate of the VL
$v_0$	copies mL <sup>-1</sup>	average initial viral concentration caused by a single successful reactivation
$\lambda$	d <sup>-1</sup>	recrudescence rate, the number of latently infected cells that successfully reactivate per day.
$K(\theta)$	-	cumulant-generating function of the viral load $V_t$ .
$\kappa_1$	copies mL <sup>-1</sup>	first cumulant, equal to the expectation of the VL ( $\mathbb{E}[V_t]$ ).
$\kappa_2$	copies <sup>2</sup> mL <sup>-2</sup>	second cumulant, equal to the variance of the VL ( $\text{Var}[V_t]$ ).
$\tilde{V}_t$	copies mL <sup>-1</sup>	conditionally deterministic approximation of the viral load process $V_t$ .
$\ell$	copies mL <sup>-1</sup>	limit of detection of the VL assay.
$\tau$	d	viral rebound time, satisfies the equation $V_\tau = \ell$
$S(t)$	-	fraction of subjects in remission at time $t$ post treatment interruption
$f(t; \lambda, g, v_0, \ell)$	d <sup>-1</sup>	approximation of the probability density function of the rebound time $\tau$
$\hat{T}_1$	d	first recrudescence time extrapolated from the rebound time using the growth rate and the initial VL, assuming simple exponential growth.
$t_{\text{ART}}$	d	days post infection that ART was initiated
$G_i$	d <sup>-1</sup>	random, clone-specific exponential growth rate.
$\sigma_G$	d <sup>-1</sup>	standard deviation of the growth rates $G_i$ of the clones comprising the SIV reservoir

See also Table 2 for additional parameters of the Bayesian mixed-effects model.

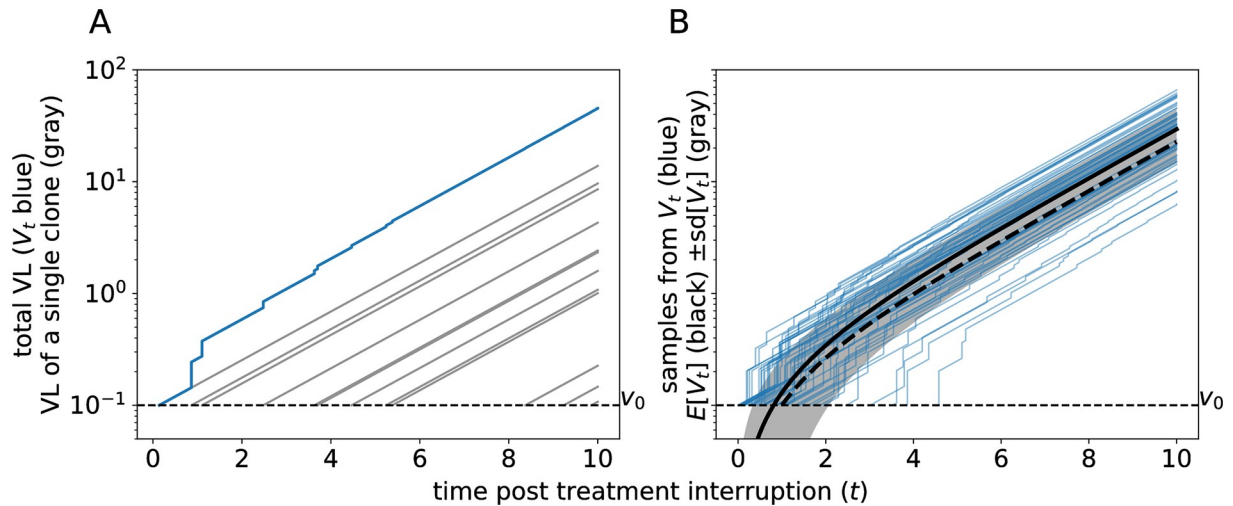
<https://doi.org/10.1371/journal.pcbi.1008241.t001>

with growth rate  $g$  and initial value  $v_0$ . The Poisson process  $N_t$  counts the number of latently infected cells that have been reactivated and successfully establish a lineage of exponentially growing infected cells at time  $t$  after treatment interruption. Under the assumption that such successful reactivation events start occurring after therapy interruption at time  $t = 0$ , we have  $N_0 = 0$  and  $N_t \sim \text{Poisson}(\lambda t)$ . The exponential curve  $v_0 e^{gt}$  describes the contribution to the total VL of such a lineage. The total VL at time  $t$  is the weighted sum of such exponential functions:

$$V_t = v_0 \sum_{i=1}^{\infty} \mathbb{1}_{[T_i, \infty)}(t) e^{g(t-T_i)} \quad (1)$$

Here, the indicator function  $\mathbb{1}_{[T_i, \infty)}(t)$  equals 1 if  $t \geq T_i$  and 0 otherwise. The random times  $T_i$  are the jump times of the Poisson process, corresponding to the times that different latently infected cells successfully reactivate. An example realization of the random process  $V_t$  given by Eq 1 is shown in Fig 1A. Notice that there might be some delay between the moment of reactivation and successful reactivation. For instance, it might be possible that the reactivation of a latently infected cell happens before treatment interruption. The times  $T_i$  correspond to the moments that lineages initiated by reactivation become large enough, the meaning of which we explore in the Discussion.

In previous analyses of Eq 1 by Pinkevych *et al.* [21] and others [19, 25, 26], the dynamics of the process  $V_t$  after the initial reactivation event  $T_1 = t_0$  was simplified using a deterministic approximation. The subsequent recrudescence times  $T_2, T_3, \dots$  were assumed to be exactly  $1/\lambda$  days apart, which is the average time between two succeeding jumps of the Poisson process.



**Fig 1. Simulations of the multiple-reactivation model.** (A) Graphical representation of Eq 1. The gray lines indicate the exponential growth curves of individual clones that originated from a single successful reactivation from the latent reservoir. The blue curve represents the total VL, i.e. the sum of the gray lines. (B) Comparison between the expectation of the process  $V_t$  (in black) and realizations sampled from this process (in blue). The mean  $\pm$  standard deviation (sd) of  $V_t$  is shown as a gray band. The dashed thick curve corresponds to the approximation  $\tilde{V}_t$  with  $t_0 = 1/\lambda$ . Parameters:  $g = 0.5 \text{ d}^{-1}$ ,  $\lambda = 1.0 \text{ d}^{-1}$ , and  $v_0 = 0.1 \text{ copies mL}^{-1}$ .

<https://doi.org/10.1371/journal.pcbi.1008241.g001>

With the aid of some further simplifications (see [Materials and methods](#) [19, 21, 25]), the following expression was obtained for the total VL at time  $t \geq t_0$  after ART suspension:

$$\tilde{V}_t \equiv v_0 \frac{e^{g(t-t_0)} - e^{-g/\lambda}}{1 - e^{-g/\lambda}} \quad (2)$$

where the tilde over the  $V$  is used to indicate that this is an approximation. In the Materials and methods section, we use the cumulant-generating function (CGF), together with some basic facts about the Poisson process to derive a functional form for the expectation of  $V_t$ , which is given by

$$\mathbb{E}[V_t] = \frac{v_0 \lambda}{g} (e^{gt} - 1) \quad (3)$$

Importantly, we no longer have to constrain recrudescence times to be  $1/\lambda$  days apart. Moreover, the same CGF technique allows us to find all other cumulants (or moments) of the distribution of  $V_t$ . For instance, we show the variance is given by

$$\text{Var}[V_t] = \frac{\lambda v_0^2}{2g} (e^{2gt} - 1)$$

and the third cumulant, which has the same sign as the skewness, equals  $\kappa_3(t) = \frac{\lambda v_0^3}{3g} (e^{3gt} - 1)$ . The expected trajectory of  $V_t$  and the standard deviation are shown in [Fig 1B](#). To compare the difference between Eqs 3 and 2, the graph of  $\tilde{V}_t$  is shown as a thick dashed curve in [Fig 1B](#). This example shows that the approximation  $\tilde{V}_t$  slightly under-estimates the expected VL (the thick black line in [Fig 1B](#)). However, the primary advantage of our improvement comes from the additional statistical properties of viral rebound dynamics that it allows us to compute, which is useful for the estimation of parameters such as the recrudescence rate  $\lambda$  (see below).

We term this model the “multiple-reactivation model” since viral load is modeled as the sum of viral lineages generated by multiple recrudescence events. In the following we contrast predictions from our refined multiple-reactivation model with the “single-reactivation model”, in which rebound VL is assumed to be associated with the viral lineage resulting from a single latent cell activation only [10]. Hence, for this single-reactivation model we consider only the first recrudescence event at time  $T_1 \sim \text{Exp}(\lambda)$ , and ignore the effects of any subsequent reactivations from the reservoir. Given that  $T_1 = t_0$ , we therefore get the following simple expression for  $t \geq t_0$ :

$$V_t^{\text{single}} \equiv v_0 e^{g(t-t_0)}$$

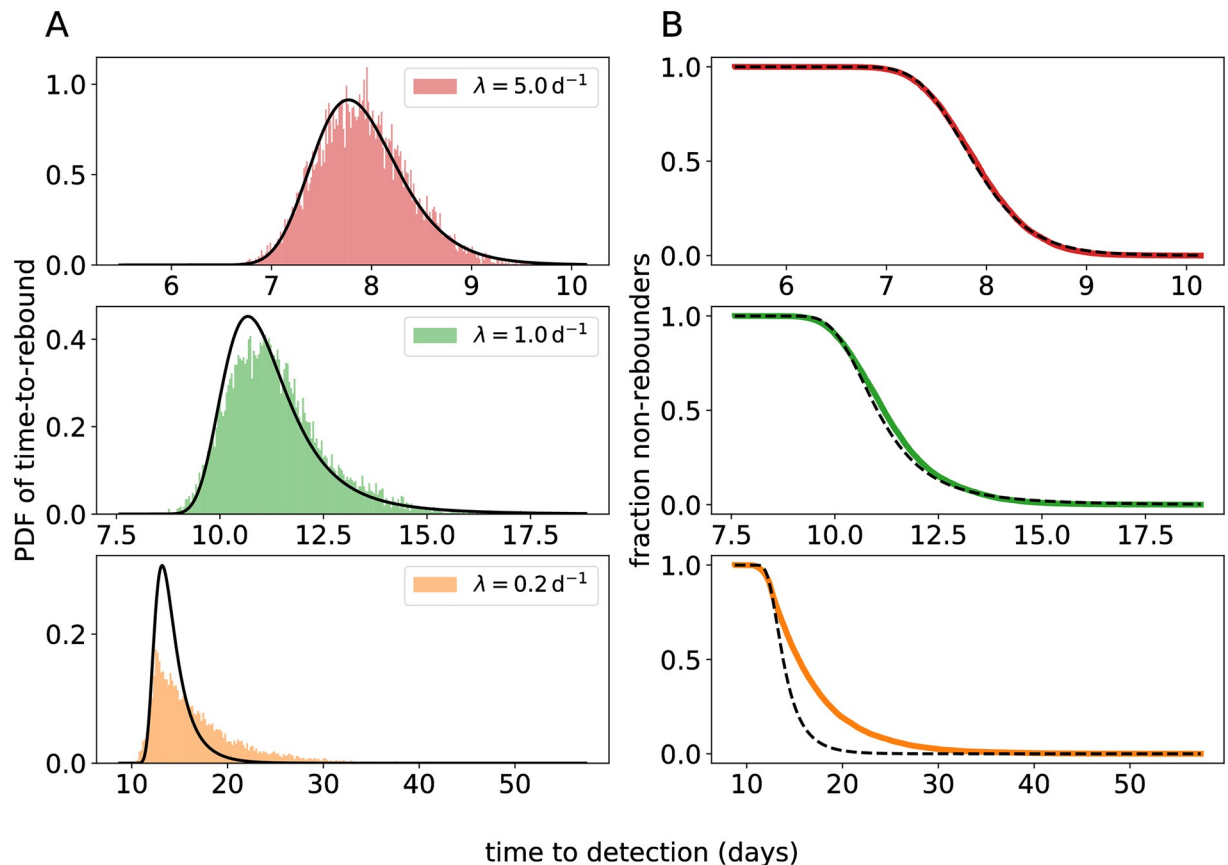
## The distribution of time-to-rebound

In treatment-interruption experiments, the main quantity of interest is the time-to-rebound, which we denote by  $\tau$ . In order to properly infer the recrudescence rate  $\lambda$ —a proxy for the replication competent reservoir size—from viral rebound data, a statistical model that expresses the likelihood of the time-to-rebound in terms of the model parameters is desirable. The predicted distributions of the time-to-rebound under the multiple-reactivation model and single-reactivation model naturally differ, because multiple reactivation events, early after treatment interruption, skew the time-to-rebound towards lower values. This means that because of these multiple recrudescence events prior to viral rebound, each of which causing a jump in the viral load, the growth of the still unobservable VL is faster than exponential growth at rate  $g$ . We refer to this as “early faster-than-exponential growth”. Using an exponential distribution for the first recrudescence time  $T_1$ , and the approximation  $\tilde{V}_t$  given in Eq 2, the rate of successful reactivation  $\lambda$  [21, 25], and the initial contribution  $v_0$  of such a reactivation event [26] have been estimated with likelihood-based methods. However, this conditionally deterministic approximation does not take the uncertainty due to secondary recrudescence events occurring at different intervals into account, a shortcoming which we fix with our fully stochastic model.

Using a diffusion approximation of the process  $V_t$  allows us to derive a convenient parametric form of the distribution of the time-to-rebound (given by Eq 6 in Materials and methods). The time-to-rebound ( $\tau$ ) is defined as the first time the virus load crosses a threshold  $\ell$  corresponding to the limit of detection (LoD; typically 50 RNA copies per mL) of the assay used to measure SIV or HIV RNA. Our parametric distribution depends on  $\ell$  and the parameters  $v_0$ ,  $\lambda$ , and  $g$  and can be used to estimate these parameters directly from time-to-rebound data using methods such as maximum likelihood. In order to test if the diffusion approximation is justified, we simulated the process  $V_t$  and compared the empirical distribution of the time to rebound with the parametric approximation (see Fig 2). When successful reactivation is fast ( $\lambda \geq 1 \text{ d}^{-1}$ ), the simulations and our approximation are in excellent agreement (by visual inspection; Fig 2 top and middle panels).

However, when the successful-reactivation rate is small ( $\lambda = 0.2 \text{ d}^{-1}$ ), the diffusion approximation breaks down (Fig 2 bottom panels), as the time to rebound is mostly determined by the first successful reactivation, and hence by the exponentially distributed initial recrudescence time  $T_1$ . Further, the distribution of the diffusion approximation of  $V_t$  at time  $t$  is a Gaussian  $\mathcal{N}(\kappa_1(t), \kappa_2(t))$ , which is symmetric. As  $\kappa_3(t) > 0$  for  $t > 0$  the exact distribution of  $V_t$  is in fact right-skewed. When we fit the multiple-reactivation model to data below, we account for these discrepancies by explicitly modeling the first reactivation time  $T_1$  as a so-called latent variable of the statistical model, which is exponentially distributed with rate  $\lambda$ . The diffusion



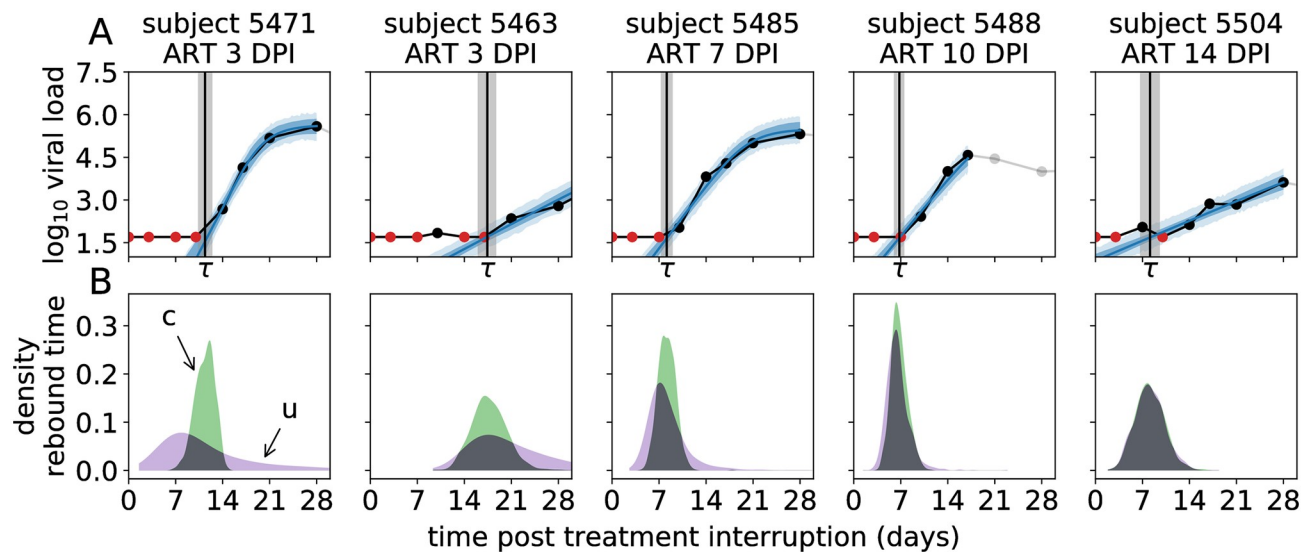


**Fig 2. Comparison between the approximation for the time-to-rebound distribution and simulated rebound times.** The simulated empirical distributions are shown in color, and our approximation is shown in black. (A) The probability density function (PDF; defined by Eq 6). (B) The survival function (i.e. the fraction of subjects  $S(t)$  that do not have a detectable VL at time  $t$ ). For the top, middle, and bottom panels different values of  $\lambda$  are used ( $\lambda = 5 \text{ d}^{-1}$ ,  $1 \text{ d}^{-1}$ , and  $0.2 \text{ d}^{-1}$  respectively). Notice the different time scale on the horizontal axes. For the remaining parameters, we used the values:  $g = 0.5 \text{ d}^{-1}$ ,  $v_0 = 0.1 \text{ copies mL}^{-1}$ , LoD  $\ell = 50 \text{ copies mL}^{-1}$ .

<https://doi.org/10.1371/journal.pcbi.1008241.g002>

approximation for the time-to-rebound distribution (Eq 6) is then used to model the difference  $\tau - T_1$ , and we set an initial condition  $V_{T_1} = v_0$ . This ensures that the model can be used for inference irrespective of whether remission is short or long, and even with data sets containing heterogeneous rebound times, as we will demonstrate below.

In the S1 Text we explore two other approximations of the rebound-time distribution that behave better for small values of  $\lambda$ . First, we replaced the Gaussian distribution  $\mathcal{N}(\kappa_1, \kappa_2)$  with a Gamma distribution, for which we matched the mean and variance with  $\kappa_1$  and  $\kappa_2$ , respectively. Like the distribution of  $V_b$ , the Gamma distribution has a positive skewness, which results in a greater similarity between the approximate rebound-time distribution and simulations when the recrudescence rate is small (S3 Fig). Second, instead of diffusion, we applied the so-called WKB approximation to the process  $V_b$ , which gave even better results for small recrudescence rates than the Gamma-law approximation (S4 Fig). Unfortunately, both these improved approximations are more difficult to implement in standard parameter-inference frameworks. For this practical reason, we use the more tractable diffusion approximation in our data analysis below.



**Fig 3. Representative examples of the fits of the mixed-effects model to the VL rebound time series.** (A) The top panels show the VL data (black dots connected by black lines, with red dots for left-censored observations; the grey dots are ignored) taken from macaques where ART was started at different days post infection (DPI), and the model prediction (blue lines: posterior mean; dark blue band: 50% credible interval (CrI), light blue band: 50% posterior predictive interval). The estimated time-to-rebound ( $\tau$ ) is given by the vertical black line (gray band: 50% CrI). (B) The bottom panels show posterior predictive distributions of the time-to-rebound. The green distributions (c) are conditioned on the estimated time of the initial recrudescence event, the purple distributions (u) are unconditional. Model fits and posterior predictive distributions for all 25 macaques are shown in [S1 Fig](#).

<https://doi.org/10.1371/journal.pcbi.1008241.g003>

## Analysis of SIV rebound data

To assess the performance of the multiple-reactivation model and our diffusion approximation with respect to actual data, we employ the results of treatment interruption experiments with the macaque SIV model [2, 5]. This data set consists of longitudinal VL measurements from macaques for whom treatment was initiated early and at varying time points after SIV challenge in different groups of animals. The time of ART initiation has been shown to be a predictor for the time-to-rebound with early SIV treatment leading to delayed rebound [2]. Moreover, in the same study [2] it was found that the rate of exponential growth of the VL after viral rebound is decreased when ART is initiated later, perhaps because immune responses develop due to higher antigen concentrations. Hence, the data set contains SIV rebound time series with varying exponential growth rates and rebound times.

Of the 36 macaques in the data set,  $n = 25$  showed viral rebound during the 16 week observation period after treatment interruption. As VL is measured at discrete times, the actual time of viral rebound ( $\tau$ ) has to be interpolated from these VL measurements. In addition, estimating the recrudescence rate ( $\lambda$ ) requires that we also estimate the viral growth rate ( $g$ ), and since we expect  $g$  and  $\lambda$  to be correlated, additional data that informs the growth rate helps to estimate both  $g$  and  $\lambda$  more accurately. In order to infer  $\tau$  and estimate  $g$  from the VL time series for each macaque, we fit a logistic growth model [2] to the initial VL data points of the time series. We manually selected time points that are consistent with logistic growth (the gray data points in [Fig 3A](#) and [S1 Fig](#) were excluded). We opted for logistic instead of exponential growth because fitting an exponential growth model to non-linear rebound data ([Fig 3A](#) and [S1 Fig](#)) can result in an under-estimation of the growth rate [19]. Because for many macaques the number of observations that can inform these estimates is limited, we used a mixed effects model to estimate the growth rate  $g$ , using the time of ART initiation ( $t_{\text{ART}}$ ) as a covariate. Similarly,  $t_{\text{ART}}$  is used as a covariate for estimating the recrudescence rate  $\lambda$ , which again has a



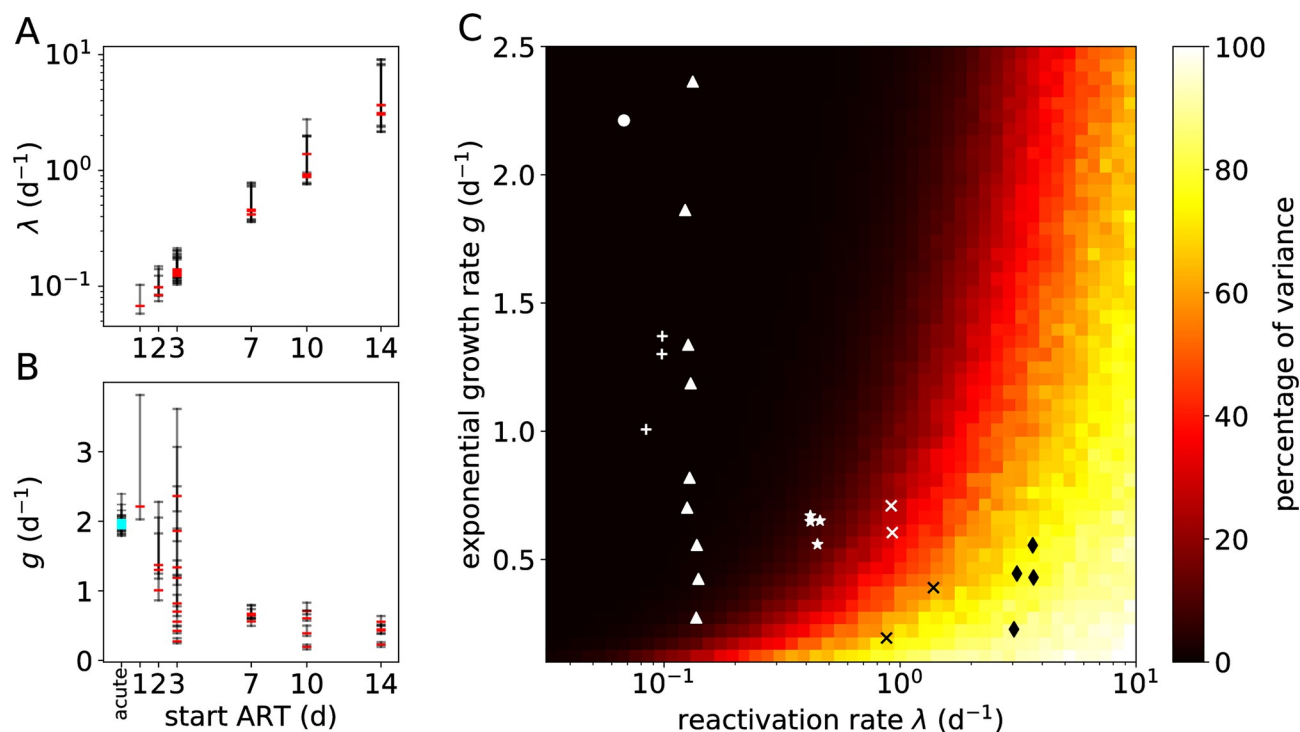
random effect for each macaque. The statistical model is defined in full detail in the Materials and methods.

Following Pinkevych *et al.* [26], we fit the logistic growth model to the VL data in a Bayesian framework using MCMC (Fig 3A and S1 Fig). This way we are able to naturally factor the rebound time distribution for the multiple-reactivation model (Eq 6) into the likelihood. We have to fit the model to the data from all 25 macaques simultaneously, as it contains fixed and random effects. We write  $\alpha_\lambda$  (and  $\alpha_g$ ) for the fixed effect of  $t_{\text{ART}}$  on  $\lambda$  (and  $g$ , respectively; see Eq 9 in Materials and methods). In accordance with previous analyses [2], we find that the time of ART initiation is a strong predictor of both the rate of reactivation and the exponential growth rate after rebound. The posterior probability  $\mathbb{P}[\alpha_\lambda < 0] < 10^{-3}$  strongly suggests that  $\alpha_\lambda > 0$ , i.e. that rebound occurs more rapidly when ART is initiated later. For the fixed effect  $\alpha_g$  of the time of ART initiation on the growth rate  $g$  we find the posterior probability  $\mathbb{P}[\alpha_g > 0] = 0.002$ , suggesting that it is highly likely that the growth rate after rebound will slow down with later ART initiation. Here the statistical significance of the effect of treatment initiation time is much larger than found previously [2]. This increased significance is due the inclusion of data from additional macaques [5], as exclusion of this data gave a posterior probability  $\mathbb{P}[\alpha_g > 0] = 0.36$ .

The estimates of  $\lambda$  and  $g$  for individual macaques are shown in Fig 4A and 4B as a function of ART initiation time, and also listed in S1 Table. The recrudescence rate is clearly influenced by the ART initiation time. We predict that each day ART is delayed, the recrudescence rate is increased by 39%, with a 95% credible interval (CrI) of [18%, 62%]. Even though we find that the time ART starts is a significant predictor for the growth rate  $g$  after rebound, the standard deviation of the growth rate's random effects ( $\sigma_g$ ) is about 5 times larger than that of the reactivation rate ( $\sigma_\lambda$ ; see S1 Table). Nonetheless, we can estimate that each day ART is delayed, the growth rate decreases by 11% (95% CrI: [4%, 20%]).

The latter observation is remarkably consistent with acute VL dynamics. As none of the macaques that were treated on day 0 (6h post-infection) showed viral rebound after ART cessation [5], we could not directly estimate the viral growth rates for these animals. However, we could compare the estimated growth rates after viral rebound with growth rates in the acute phase. Using again a simple random-effects logistic growth model, we were able to estimate viral growth rates during acute infection for the 13 out of the 25 macaques that showed observable viremia prior to ART initiation. These estimates are added to Fig 4B (cyan markers, located at “acute”). Our estimates for the acute growth rates are slightly higher than reported previously [2], possibly due to the use of a logistic growth model (see Materials and methods). Using our estimates from the rebound data of the population-level growth rate ( $\mu_g$ , see Table 2) and fixed effect of  $t_{\text{ART}}$  ( $\alpha_g$ ), we extrapolated the population-level growth rate for subjects treated at day 0 (using Eq 9). Our estimate of the population-level growth rate for the acute infection ( $\hat{\mu}_g = 0.67 \log \text{d}^{-1}$ ; S1 Table) falls within the 50% CrI of the extrapolated growth rate ([0.28, 0.76]  $\log \text{d}^{-1}$ ). This suggests that viral dynamics after rebound in very early treated subjects resembles acute infection dynamics.

We then used model selection theory to compare the multiple-reactivation model to the single-reactivation model (see Methods). Using the Watanabe-Akaike information criterion (WAIC; see Materials and methods and [27]) for model comparison, we find “very strong evidence” (sec. [28]) in favor of the multiple-reactivation model ( $\Delta\text{WAIC} = 11.5$ ). The superior performance of the multiple-reactivation model can be explained by two mechanisms that were mentioned above: (i) the stochasticity of secondary recrudescence events and (ii) early faster-than-exponential growth. We will now look closer into the effects of these mechanisms in the context of our SIV data set.



**Fig 4. Estimates of recrudescence and growth rates from the SIV rebound data and the percentage of the variance of the time-to-rebound.** (A) Point estimates (posterior modes; red) and 50% CrIs (black) of  $\lambda$  for each macaque as a function of the time ART was initiated. (B) Estimates of  $g$ . The cyan markers denote estimates of the growth rate for acute infections of 13 of the 25 macaques. These acute VL growth rates cluster around  $2 \text{ d}^{-1}$ . (C) Proportion of the total variance due to secondary reactivation events. The heat map shows  $\text{Var}[\tau_1]/\text{Var}[\tau_0] \cdot 100\%$ , where  $\tau_i := \inf\{t : V_t \geq \ell, V_0 = i \cdot v_0\}$  is the rebound time ( $i = 0$ ) or the time between the first successful reactivation and rebound ( $i = 1$ ). Additional parameters are  $v_0 = 0.1 \text{ copies mL}^{-1}$  and  $\ell = 50 \text{ copies mL}^{-1}$ . The markers indicate the estimates from macaque SIV rebound experiments in which the macaques were treated, starting  $t_{\text{ART}}$  days after infection, with  $t_{\text{ART}}$  equal to 1 day (●), 2 days (+), 3 days (▲), 7 days (★), 10 days (×), or 14 days (◆).

<https://doi.org/10.1371/journal.pcbi.1008241.g004>

**Uncertainty due to secondary recrudescence events.** According to the multiple-reactivation model, successful reactivation events that follow the first event lead to faster-than-exponential growth of the VL during the early stages of rebound (Fig 1). However, these secondary reactivation events can only contribute noticeably to the viral load when the VL is still relatively low and close to the initial value  $v_0$ . This most likely happens when the reactivation rate is large or the exponential growth rate is small. In order to quantify these effects, we can

**Table 2. Prior distributions of the Bayesian mixed-effects model.**

parameter	description	prior	hyper-prior
$\sigma$	VL measurement error	$ \mathcal{N}(0, 0.5) $	-
$\epsilon_g$	random effect VL growth rate ( $g$ )	$\mathcal{N}(\mu_g, \sigma_g)$	$\mu_g \sim \mathcal{N}(0, 1),$ $\sigma_g \sim  \mathcal{N}(0, 1) $
$\alpha_g$	fixed effect of $t_{\text{ART}}$ on $g$	$\mathcal{N}(0, 1)$	-
$\epsilon_\lambda$	random effect recrudescence rate ( $\lambda$ )	$\mathcal{N}(\mu_\lambda, \sigma_\lambda)$	$\mu_\lambda \sim \mathcal{N}(0, 1),$ $\sigma_\lambda \sim  \mathcal{N}(0, 1) $
$\alpha_\lambda$	fixed effect of $t_{\text{ART}}$ on $\lambda$	$\mathcal{N}(0, 1)$	-
$\log_{10}(K)$	carrying capacity VL	$\mathcal{N}(\mu_K, \sigma_K)$	$\mu_K \sim \mathcal{N}(5, 2),$ $\sigma_K \sim  \mathcal{N}(0, 2) $
$\log(v_0)$	initial VL equivalent	$\mathcal{N}(-1, 1)$	-

The notation  $x \sim |D|$  for probability distribution  $D$  means that  $x$  is positive and that  $D$  is truncated at zero. The normal distribution is parameterized with the mean and standard deviation. The time  $t_{\text{ART}}$  denotes the number of days post infection at which antiretroviral treatment was initiated.

<https://doi.org/10.1371/journal.pcbi.1008241.t002>

decompose the variance of the time-to-rebound ( $\tau$ ) as the sum of the variance of the first reactivation time ( $T_1$ ), and the variance due to all subsequent reactivation events. The proportion of the total variance that is due to secondary reactivation events is shown in Fig 4C for different values of the growth rate  $g$  and the recrudescence rate  $\lambda$ . The dark region in this heat map corresponds to a part of the model's parameter space where it is indistinguishable from the single-reactivation model. On the contrary, the light region corresponds to parameter combinations for which most of the variance in the time-to-rebound is due to the secondary successful reactivation events. In this parameter regime the model is most relevant.

The superior performance of the multiple-reactivation model can be explained by having data from macaques with a high recrudescence rate  $\lambda$  and a small exponential growth rate  $g$ . Point estimates (i.e., modes of the marginal posterior distributions) of  $\lambda$  and  $g$  for each macaque are projected onto the heat map in Fig 4C. The macaques that were treated starting at 7, 10 and 14 days after infection fall into the parameter domain where the multiple-reactivation model is most relevant.

To further assess the effect of multiple reactivation events for each macaque, we sampled from the posterior predictive distribution of the time-to-rebound (Fig 3B, purple distributions). This distribution indicates when viral rebound is most likely to take place, given estimates for the growth rate  $g$ , the rate of successful reactivation  $\lambda$  and the initial VL  $v_0$  when exponential growth begins. The actual estimates for the rebound time  $\tau$  (Fig 3A and S1 Fig, black vertical lines) correspond well with the posterior predictive distributions, as all 25 estimates of  $\tau$  fall within the 2.5 and 97.5 percentiles of the posterior predictive distributions and 21 out of 25 estimates fall within the interquartile range. In the model, we explicitly estimate the first recrudescence time  $T_1$  (see Materials and methods) and hence, we can also sample from the posterior predictive distribution of  $\tau$  conditioned on  $T_1$  (Fig 3B, green densities). These second posterior predictive distributions indicate the uncertainty in the rebound time due to secondary recrudescence events (in addition to uncertainty in the parameter estimates). Hence, by comparing the conditional (Fig 3B, green) and unconditional (Fig 3B, purple) posterior predictive distributions of  $\tau$ , we see what effect multiple recrudescence events have on the uncertainty of the rebound time. For early treated macaques ( $\text{ART} \leq 3$  days post infection), most uncertainty in the rebound time comes from the first successful reactivation, as illustrated by the purple densities being much wider than the green densities. On the other hand, for the macaques treated later the subsequent reactivation events determine the rebound time distribution, as illustrated by the purple and green densities overlapping.

**Early faster-than-exponential growth.** When the recrudescence rate is large, and before the VL has become detectable, the multiple-reactivation model predicts that the VL grows faster than exponentially (Fig 1). To demonstrate the effect of this faster-than-exponential growth, we can use the regular exponential growth model to extrapolate what the first reactivation time would have been under the single-reactivation model. This time is denoted  $\hat{T}_1$ , and can easily be calculated using the model's parameters as  $\hat{T}_1 = \tau - \frac{1}{g} \log(\ell/v_0)$ . The marginal posterior densities of the first reactivation time  $T_1$  and the extrapolated initial recrudescence time  $\hat{T}_1$  are nearly identical for the early treated macaques (S2 Fig). However, for the macaques that are treated later, the extrapolated recrudescence time becomes negative (i.e. successful reactivation is predicted to occur before treatment interruption), while, according to our models, the first recrudescence time  $T_1$  has to be positive. This shows that given the estimates of  $g$ ,  $v_0$ , and  $\tau$ , faster-than-exponential growth as predicted by the multiple-reactivation model is required to explain the VL data.

To identify which of the two mechanisms (uncertainty due to secondary recrudescence events or early faster-than-exponential growth) described above is the most important for

explaining the difference in WAIC between the single and multiple-reactivation model, we also fit the “conditionally deterministic” multiple-reactivation model (see [Materials and methods](#) and [21]) to our SIV data set. Recall that in this model all secondary recrudescence events occur at fixed intervals. Again using WAIC, we find that our fully stochastic multiple-reactivation model fits the data better than the conditionally deterministic version, but only with limited statistical significance ( $\Delta\text{WAIC} = 2.1$  and see [S2 Table](#)). As our fully stochastic model differs from the conditionally deterministic model in that it describes uncertainty in the rebound time due to secondary recrudescence events, we find that for this data set the randomness of these secondary events may be of limited importance.

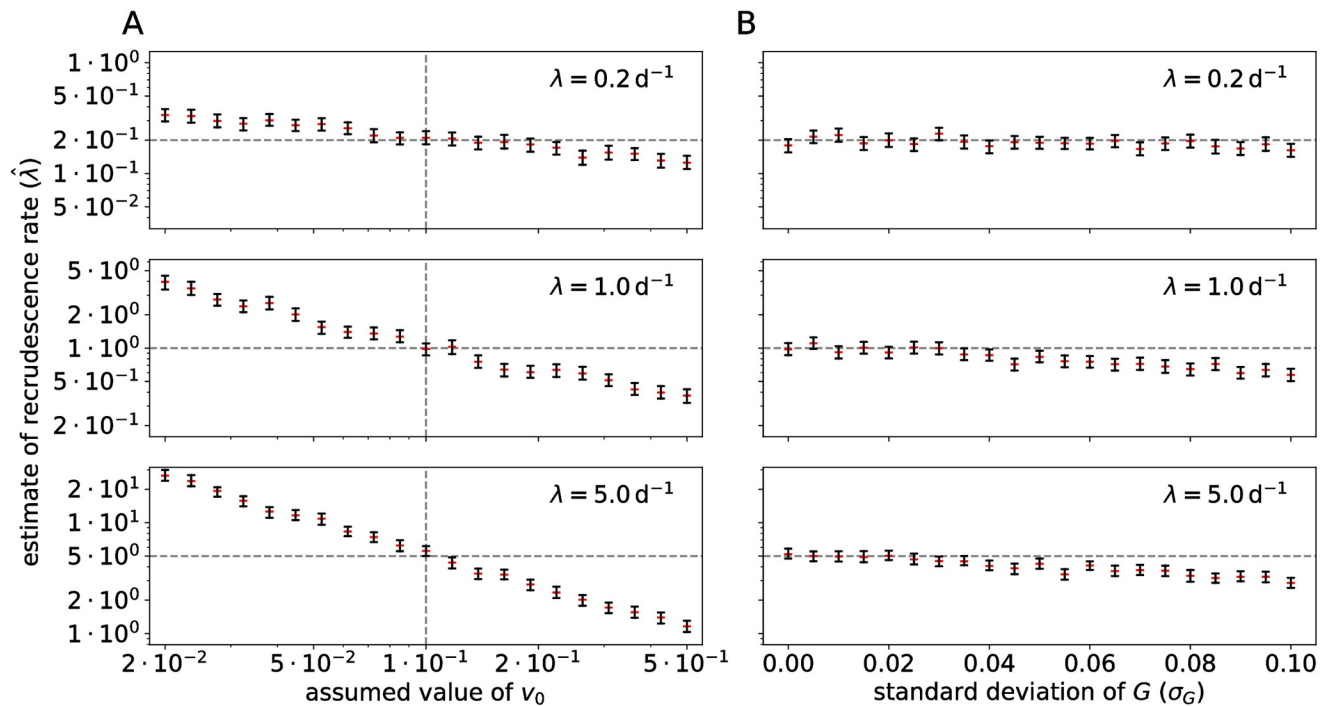
### Sensitivity to parameter and model misspecification

Next, we investigated the effects of uncertainty in the initial viral load parameter ( $v_0$ ) and heterogeneity of the exponential viral growth rate, which can exist when the reservoir is comprised of a variety of phenotypically distinct SIV clones, on the estimates of the recrudescence rate  $\lambda$ .

**Uncertainty in the initial viral load equivalent.** The meaning of the parameter  $v_0$  is biologically ambiguous. Previously, this parameter has been described as the initial “plasma viral load equivalent” [26] and estimates of  $v_0$  are at least roughly compatible with the number of virions produced by one productively infected cell, the clearance rate of virions [29] and the blood volume of a macaque [26]. Another interpretation is linked to extinction probabilities of a recently reactivated lineage. In this case  $v_0$  is the viral load at which extinction of an exponentially growing lineage is extremely unlikely [18, 20]. The actual model is agnostic with respect to the interpretation of  $v_0$ , which can be thought of as the effect size of the multiple-reactivation model. This means that  $v_0$  simply provides a measure of the effect of each independent recrudescence event, each possibly originating from a separate anatomical site [24], on the VL dynamics and time-to-rebound. However, as it is difficult to estimate both the recrudescence rate  $g$  and the initial viral load  $v_0$  simultaneously, we investigated the effect of a misspecified  $v_0$  on the estimate of the recrudescence rate  $\lambda$ .

To assess the bias due to misspecification of  $v_0$ , we simulated large data sets ( $n = 200$ ) with various ground-truth parameter values and fit our model to the synthetic data. For simplicity, we used an exponential growth model instead of logistic growth, and removed the random effects from the statistical model. Hence, all simulated subjects share the same parameter values. The ground-truth  $v_0$  was kept constant to  $0.1 \text{ copies mL}^{-1}$ , while in the statistical model, the assumed constant value of  $v_0$  was varied from  $0.02$  to  $0.5 \text{ copies mL}^{-1}$ . Assuming an erroneous value of  $v_0$  resulted in a sizable bias in the estimate of  $\lambda$  ([Fig 5A](#)). When  $v_0$  is assumed smaller than the ground truth value, the model requires a larger recrudescence rate in order to fit the data, and vice versa. This is especially clear when the ground-truth reactivation rate is large ( $\lambda = 5 \text{ d}^{-1}$ ). This is as expected, because again,  $v_0$  can be interpreted as the effect size of the multiple-reactivation model, and becomes more important when secondary recrudescence events are more frequent.

**Within-host heterogeneity of the exponential growth rate.** Throughout the paper, we have made the strong assumption that within one subject all successfully reactivated lineages have the same exponential growth rate  $g$ . In natural infections, ART is only rarely started during hyper-acute infection and this means that the latent reservoir consists of a diverse archive of proviral sequences [30], probably varying in their growth rate due to intrinsic fitness costs of mutations or escape mutations from immune responses [31]. To measure the effect of this potential model misspecification, we performed a sensitivity analysis with simulated data sets as before. In this case, we had to generalize the multiple-reactivation model ([Eq 1](#)) and replace



**Fig 5. Sensitivity of the multiple-reactivation model to misspecification.** (A) A misspecified initial viral load  $\nu_0$  can lead to a biased estimate of the recrudescence rate  $\lambda$ . Rebound data sets ( $n = 200$ ) were simulated by sampling from the viral load process (Eq 1) using different values of the recrudescence rate  $\lambda$  (horizontal dashed lines), and different assumed values of  $\nu_0$  (horizontal axis). The ground truth value of  $\nu_0$  equals 0.1 copies  $\text{mL}^{-1}$  (vertical dashed lines). Shown are the 95% CrIs of the estimate  $\hat{\lambda}$  of  $\lambda$  (black bars) and the posterior medians (red). (B) Intra-host variation in the exponential growth rate of the VL can lead to a biased estimate of the recrudescence rate  $\lambda$ . Data sets of rebound time series were now simulated from a viral load process with within-host variability of the growth rate  $G_i$  (Eq S12 in S1 Text, S5 Fig), using different values of the recrudescence rate and the standard deviation of the viral growth rate ( $\sigma_G$ , horizontal axis), ranging from 0% to 20% of the most likely growth rate  $g = 0.5$ .

<https://doi.org/10.1371/journal.pcbi.1008241.g005>

the growth rate  $g$  with a clone-specific random variable  $G_i$ , representing the random growth rate of the  $i$ -th clone (see Eq S12 in S1 Text). Example trajectories of this generalized model are shown in S5 Fig. This generalized multiple-reactivation model requires that we specify a distribution for the random growth rates  $G_i$ . In the S1 Text we developed a simple example providing a model for an SIV reservoir in which the frequency of a clone is proportional to its fitness (see the inset of S5 Fig). The most fit and abundant clone has growth rate  $g$  and we write  $\sigma_G$  for the standard deviation of the random growth rate  $G_i$ .

As in the sensitivity analysis described above, we simulated large data sets ( $n = 200$ ), varying the recrudescence rate  $\lambda$  and the standard deviation  $\sigma_G$  of the random growth rates. We then estimated the model parameters  $g$  and  $\lambda$  with the simplified statistical model (i.e. exponential growth instead of logistic growth and no random effects). The initial viral load  $\nu_0$  was kept constant to the true value. The estimated reactivation rates are shown in Fig 5B. This shows that a non-zero standard deviation in the within-host growth rate introduces a bias in the estimate of  $\lambda$ . When  $\sigma_G$  is large, the estimate of the recrudescence rate is smaller than the ground-truth value. We can understand this intuitively, because clones that reactivate early might have a growth rate that is significantly smaller than the maximum growth rate  $g$ , which delays the time of rebound, while the observed growth rate is dominated by fitter clones that have successfully reactivated after the first clone. This effect is most pronounced when the ground-truth recrudescence rate is intermediate or large ( $\lambda = 1$  or  $5 \text{ d}^{-1}$ ). Again, this is in line with



expectations, because when the recrudescence rate is small, the growth rate of the total VL is mostly determined by the clone that successfully reactivates first.

In the case of the SIVmac251-infected macaques analyzed here that are treated within 2 weeks of infection, we expect that the phenotypic variation in the reactivating strains is limited and that using a constant growth rate  $g$  is a valid simplification. However, when a data set contains subjects that are put on ART relatively late in the acute infection, or during chronic infection, recrudescence rates estimated with the multiple-reactivation model will likely be biased towards lower values, due to longer rebound times. We therefore investigated if our parametric rebound-time distribution can be adjusted to account for situations when  $\sigma_G > 0$ . In the [S1 Text](#), we derive the CGF for the generalized multiple-reactivation model described above (Eq S13). In particular, the first and second cumulants can be used to derive an approximate survival function for the fraction of subjects in remission, which is in excellent agreement with simulated rebound times ([S6 Fig](#)). This shows that our probabilistic methodology can be used to extend the multiple-reactivation model to account for important biological aspects as heterogeneity of the reservoir.

## Discussion

We carefully analyzed a model for SIV and HIV rebound after treatment interruption developed by Pinkevych *et al.* [21] and Hill *et al.* [20] that takes into account the potential effect of the reactivation of multiple latently infected cells on the rebound time. In doing so, we were able to derive a relatively simple statistical model that can be used for the inference of the rate of recrudescence after treatment cessation, the viral growth rate after recrudescence, and perhaps ultimately the efficacy of novel HIV treatments in delaying viral rebound. Moreover, using our mathematical formulation, the model can be compared to similar models of viral rebound in a statistically rigorous manner. We were able to find strong statistical evidence ( $\Delta\text{WAIC} = 11.5$ ) in favor of the multiple-reactivation model over a simple model with only one reactivation event using previously published data from treatment-interruption experiments performed in SIV-infected macaques [2, 5]. We argued that the multiple-reactivation model is most relevant for data sets that contain subjects with early viral rebound, as our SIV data set. This is often the case for human data sets as well. For example in a pooled data set of six ACTG studies [32], 6–63% of subjects showed detectable viremia within a week, and 21–74% within 2 weeks of ART cessation [11].

Our method captures the uncertainty in SIV rebound times that is due to the stochastic nature of any recrudescence events that follow the initial activation of a latently infected cell that led to remission failure. This feature is not present in the approximation derived by Pinkevych *et al.* [21]. This novel aspect slightly improves the model's ability to describe experimental data; when we compared our fully stochastic multiple-reactivation model with the conditionally deterministic model in the context of our SIV rebound data set, we found a small  $\Delta\text{WAIC}$  of 2.1 in favor of the fully stochastic model. This indicates that the most important advantage of the multiple-reactivation model is the ability to explain fast rebound due to early faster-than-exponential viral growth.

Our fully stochastic multiple-reactivation model suffers from some of the same limitations as previous approximations [19, 25, 26]. The exact biological meaning of the initial viral load parameter  $v_0$  is ambiguous, and as we have shown with our sensitivity analysis, the estimate of the recrudescence rate is biased when the value of  $v_0$  is misspecified. In our Bayesian data analysis, we resolved this issue by choosing a broad prior distribution for  $v_0$ , such that uncertainty in this parameter is propagated to uncertainty in the recrudescence rate  $\lambda$ . However, the model is still sensitive to the exact location and spread of this prior distribution. In addition, we



found that the multiple-reactivation model is also sensitive to within-host variation in the viral growth rate. Even though this will likely not affect our estimates, because early treatment limits the heterogeneity of the reservoir, this bias should be taken into account when the model is applied to treatment interruption experiments with later-treated subjects, in particular in most human studies.

By specifying the model in terms of the recrudescence rate  $\lambda$ , the recrudescence times  $T_i$ , the initial viral load equivalent  $v_0$ , and the exponential growth rate  $g$ , we have combined all complex dynamics of reactivation and the initial stochastic growth into a single abstract recrudescence event. *In vitro* experiments have pointed out that this may be an oversimplification [33]. It is likely that a reduced exponential growth rate, for instance due to a therapeutic vaccine, also influences the rate of recrudescence, because the chances of successful reactivation are dependent on the fitness of the clone, which will be influenced by the immune response. Therefore our parameters  $\lambda$  and  $g$  are *a priori* dependent. A possible solution would be to parameterize the model in terms of the reactivation rate instead of the recrudescence rate, and add a parameter that determines the probability of successful reactivation. This parameter is known as the “establishment probability”, and depends on the viral dynamics in a non-trivial manner [18]. For the aims of our current analysis, the exact relation between the reactivation and recrudescence rate are not important. However, when the multiple-reactivation model is applied to novel HIV therapies that aim to (indefinitely) extend remission, it can be important to distinguish the effects of therapies that reduce viral fitness, such as therapeutic vaccination [7] or broadly neutralizing antibodies [8, 34, 35], and therapies that reduce the reactivation rate, such as latency reversing agents [36].

In the presented model formulation and inference, we have ignored the period of drug washout after treatment interruption. While pharmacokinetics and dynamics may be important for precisely estimating the reactivation rate, and for instance the value of  $v_0$  [26], taking a drug washout time of 0 days is a conservative assumption for the purpose of this study. Indeed, incorporating a drug washout decreases the time available for exponential growth and hence multiple reactivation events that lead to faster-than-exponential growth become more important for rapidly rebounding macaques. We verified this by repeating the analyses with a fixed drug washout period of 1 day, during which recrudescence is not allowed to occur. Compared to the single-reactivation model, the evidence in favor of the stochastic multiple-reactivation model increased ( $\Delta\text{WAIC} = 14.3$ ), and compared to the conditionally deterministic multiple-reactivation model results were as before ( $\Delta\text{WAIC} = 2.2$ ).

Based on our estimates of the effect of the ART initiation time on the recrudescence rate ( $\alpha_r$ ), we predict that each day that ART initiation is delayed, the recrudescence rate increases by 36%. Recently, the aforementioned genetically barcoded SIV rebound experiments [25] have been repeated with ART initiated at day 10 and 27 post infection as opposed to day 4 [37]. These barcoded experiments could in principle give a much better estimate of the recrudescence rate, because for each macaque multiple successful reactivation events can be observed by counting the frequencies of different SIVmac239M clonotypes. In the same study, the size of the reservoir was also estimated more directly by measuring cell-associated (CA) SIV DNA in peripheral blood mononuclear cells (PBMCs). Surprisingly, while the estimated size of the reservoir based on SIV CA-DNA at the time of treatment interruption is increased more than a 100-fold when ART is started at day 10 instead of day 4 post infection, the rate of successful reactivation (inferred by counting clonotype frequencies) only increases 3.6-fold, which would amount to a 25% increase per day. This rate falls within the 95% CrI of our estimate (*viz.* [18%, 62%]). When treatment was initiated even later (day 27), the frequency of CA-DNA at the time of treatment interruption appeared to plateau at the same level as the day-10 treated macaques. Surprisingly, the inferred recrudescence rate dropped to only a

2-fold increase compared to the day-4 treated macaques. This strongly suggests that our result cannot be extrapolated to ART initiation beyond hyper-early infection, making it difficult to compare these results to most human studies, because treatment almost never starts this early for human subjects. For the early treated human cohort studies that do exist (e.g. [4]), the comparison between macaque and human data can be aided by the fact that macaques are challenged with a much higher infectious dose, leading to a shorter eclipse phase compared to humans.

When we consider the 25 macaques used for this study, a large qualitative difference seems to exist between the animals treated within 3 days and those animals treated after 7 days (S1 Fig). This can potentially be explained by the fact that in the early-treated macaques, no SIV-specific antibody, CD4<sup>+</sup>, or CD8<sup>+</sup> T-cell responses could be detected [2], contrary to macaques treated from day 7 onward. One could even argue that as the time of ART initiation approaches the time of infection, the viral rebound dynamics after ART interruption starts to resemble those of an acute infection (Fig 4B). Similar patterns have been found for HIV-1, where ART initiation during acute HIV infection can lead to an incomplete HIV-specific humoral immune response, as measured by diagnostic assays [38, 39]. On the other hand, patients treated during Fiebig stage I or II have been shown to develop detectable HIV-specific CD8<sup>+</sup> T-cell responses [40]. Although these responses are lower in magnitude and breadth than CD8<sup>+</sup> T-cell responses from untreated individuals, they show enhanced differentiation into the effector-memory T-cell phenotype, leading to a more functional CD8<sup>+</sup> memory T-cell pool compared to patients for whom treatment was initiated later. The effects of early ART on the formation of immunological memory and the subsequent impact on viral rebound dynamics could be resolved by experimentally filling the gap between macaques treated at day 3 and day 7, ideally incorporating immunological assays and using a barcoded strain. In order to extrapolate beyond ART initiation within two weeks, we will likely need models that explicitly incorporate immune responses and mechanisms like CD8<sup>+</sup> T-cell exhaustion [41].

Mathematical models are required to bridge the gap between experimental observations made during treatment interruption experiments and the effect induced by novel curative treatments. A more accurate mathematical model will therefore increase the precision by which we can estimate reactivation rates—and importantly the uncertainty of these estimates—and infer the efficacy of such treatments. Here we showed that with the right mathematical tools, models of rebound dynamics can easily be refined, and used to measure parameters relevant for recrudescence. As we exemplified by incorporating within-host heterogeneity of the exponential growth rate, we envisage that our framework can be extended to include many other biological aspects, such as the pharmacodynamics of antiretrovirals or monoclonal antibodies [34] and detailed reactivation mechanics. Hopefully, this will lead to a more accurate understanding of SIV and HIV rebound kinetics and the efficacy of novel HIV therapies.

## Materials and methods

### Data

The collection of the data is described in detail by Whitney *et al.* [2, 5]. In short, 36 rhesus macaques were infected with 500 TCID<sub>50</sub> of SIVmac251. Combination antiretroviral treatment (a cocktail of tenofovir, emtricitabine, and dolutegravir) was initiated at various times post infection (6 hours, 1, 2, 3, 7, 10, and 14 days). Treatment continued for 24 weeks, and the viral load (VL) was monitored for 16 weeks after treatment interruption, while taking weekly measurements with a limit of detection of 50 RNA copies per mL.

## The cumulants of the process $V_t$

The VL is modeled by the process  $V_t$  given by Eq 1, where  $\{T_i : i = 1, 2, \dots\}$  are the jump times of the Poisson process  $N_t$ , with each jump reflecting a successful reactivation event from the reservoir. The derivation of the cumulants of the process  $V_t$  makes use of the fact that conditioned on  $N_t = n$ , the random times  $\{T_1, \dots, T_n\}$  are independent and uniformly distributed on the interval  $[0, t]$  (see e.g. [42]). This simply means that if one knows that  $t$  days after treatment interruption exactly  $n$  latently infected cells successfully reactivated, then there was no *a priori* preference for when these reactivation events took place within the time window. Or course, this is only true under the assumption that successful reactivation events can be accurately modeled by a time-homogeneous Poisson process. An overview of the parameters and variables used is given in Table 1.

The cumulant-generating function (CGF) of  $V_t$  is defined as the logarithmic moment-generating function  $K(\theta) = \log \mathbb{E}[\exp(\theta V_t)] = \kappa_1 \theta + \frac{1}{2} \kappa_2 \theta^2 + \dots$  where the first cumulant  $\kappa_1 = \mathbb{E}[V_t]$  and the second cumulant  $\kappa_2 = \text{Var}[V_t]$ . First, assume that  $N_t = n$  so that

$$\begin{aligned} \mathbb{E}[\exp(\theta V_t)] &= \mathbb{E}\left[\exp\left(\theta v_0 \sum_{i=1}^n e^{g(t-T_i)}\right)\right] \\ &= \prod_{i=1}^n \mathbb{E}[\exp(\theta v_0 e^{g(t-T_i)})] = \left(\frac{1}{t} \int_0^t \exp(\theta v_0 e^{gs}) ds\right)^n \end{aligned}$$

where the expectations are conditional on  $N_t = n$ . In this derivation the second equality follows from independence, and the third from the identical uniform distributions of the  $T_i$ . Next, we drop the condition  $N_t = n$ , and use instead  $N_t \sim \text{Poisson}(\lambda t)$  and hence

$\mathbb{P}[N_t = n] = e^{-\lambda t} (\lambda t)^n / n!$ . Using the law of total probability, we get

$$\begin{aligned} \log \mathbb{E}[\exp(\theta V_t)] &= \log \left( e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} \left( \frac{1}{t} \int_0^t \exp(\theta v_0 e^{gs}) ds \right)^n \right) \\ &= \lambda \int_0^t \exp(\theta v_0 e^{gs}) ds - \lambda t \end{aligned}$$

Suppose that  $m > 0$ . The  $m$ -th cumulant  $\kappa_m$  is now given by

$$\begin{aligned} \kappa_m &= \frac{d^m}{d\theta^m} K(\theta) \Big|_{\theta=0} = \lambda \int_0^t (v_0 e^{gs})^m \exp(\theta v_0 e^{gs}) ds \Big|_{\theta=0} \\ &= \lambda v_0^m \int_0^t e^{mgs} ds = \frac{\lambda v_0^m}{mg} (e^{mgt} - 1) \end{aligned} \quad (4)$$

Notice that we could have used the moment generating function instead of the CGF, although calculating Eq 4 would have been more involved. A derivation of formulae for the cumulants of more general Poisson processes can be found in e.g. Privault [43].

Above we have focused on the statistics of the process  $V_t$  with initial condition  $V_0 = 0$ . However, below we require arbitrary initial conditions  $V_0 = v \geq 0$ . Fortunately our results easily generalize to this situation. A VL process that starts at level  $v$  at time  $t = 0$  can be written as  $ve^{gt} + V_t$  where  $V_t$  denotes the usual process with initial state  $V_0 = 0$ . Because  $ve^{gt}$  is deterministic, the cumulant generating function of  $ve^{gt} + V_t$  is simply given by  $\log \mathbb{E}[\exp(\theta ve^{gt} + \theta V_t)] = \theta ve^{gt} + K(\theta)$ , where  $K(\theta)$  is again the CGF of  $V_t$ . Therefore, when the initial condition equals  $V_0 = v$ , only the first cumulant (the mean) of  $V_t$  changes from  $\kappa_1$  to  $ve^{gt} + \kappa_1$ , and all other cumulants remain unaffected.

## The derivation of the approximation $\tilde{V}_t$

In the analysis of Pinkevych *et al.* [21], the Poisson process  $N_t$  is replaced by a process  $\tilde{N}_t$  with jump times  $T_1, T_2, \dots$  that is deterministic conditioned on  $T_1 = t_0$ ; the time of the first successful reactivation event. The subsequent recrudescence times  $T_2, T_3, \dots$  are spaced at regular intervals, with  $T_{i+1} - T_i = 1/\lambda$  for  $i \geq 1$ . The number of successful reactivation events at time  $t > t_0$  is then given by  $1 + \lfloor \lambda(t - t_0) \rfloor \equiv \tilde{N}_t \mid \{T_1 = t_0\}$ , where  $\lfloor x \rfloor$  denotes the largest integer  $\leq x$ . Eq 2 can now be derived as follows:

$$\sum_{i=1}^{\lfloor \lambda(t-t_0) \rfloor + 1} v_0 e^{g(t-t_0 - \frac{i-1}{\lambda})} = v_0 e^{g(t-t_0)} \frac{1 - e^{-g/\lambda(\lfloor \lambda(t-t_0) \rfloor + 1)}}{1 - e^{-g/\lambda}}$$

$$\approx v_0 \frac{e^{g(t-t_0)} - e^{-g/\lambda}}{1 - e^{-g/\lambda}} \equiv \tilde{V}_t$$

The first step follows from the identity for a geometric progression, and in the second step the approximation  $\lfloor \lambda(t - t_0) \rfloor \approx \lambda(t - t_0)$  is used.

Another way to approximate the stochastic process  $V_t$  is to assume that  $\lambda$  is very large, so that latently infected cells are continuously reactivated. Each of these reactivations adds  $v_0$  SIV RNA copies  $\text{mL}^{-1}$  to the total VL. In this case it becomes feasible to use an ordinary differential equation (ODE) to describe the VL dynamics. The initial value problem (IVP) for the large- $\lambda$  approximation  $\bar{V}_t$  of  $V_t$  is given by

$$\frac{d}{dt} \bar{V}_t = \lambda v_0 + g \bar{V}_t, \quad \bar{V}_0 = 0$$

and the solution to this IVP is given by the right-hand-side of Eq 3, which was also noted by Prague *et al.* [19]. When  $\lambda$  is large, we can approximate  $1 - e^{-g/\lambda}$  with  $g/\lambda$ , and  $e^{-g/\lambda}$  with 1 in Eq 2. This implies that  $\tilde{V}_t \approx \bar{V}_t$  when recrudescence is fast.

## The first passage time of the limit of detection

Here, we derive a parametric probability distribution for the time to viral rebound after treatment interruption, which is our main tool for analyzing viral rebound data. Above, we have seen that the expectation of  $V_t$  is given by the first cumulant,  $\kappa_1(t) = \lambda v_0 \frac{1}{g} (e^{gt} - 1)$ , and that the variance equals  $\kappa_2(t) = \lambda v_0^2 \frac{1}{2g} (e^{2gt} - 1)$ . A naive way to derive an approximation for the time-to-rebound  $\tau$  is to approximate the distribution of  $V_t$  with  $\mathcal{N}(\kappa_1(t), \kappa_2(t))$ , a normal distribution with mean  $\kappa_1(t)$  and variance  $\kappa_2(t)$ , and this is essentially what we will do below. However, in the S1 Text, we will give a theoretical justification for this naive approach and use the Kramers-Moyal expansion to replace  $V_t$  with a transient Ornstein-Uhlenbeck (OU) process (see e.g. [44, 45]).

Armed with a Gaussian approximation of the distribution of  $V_t$ , we can derive an approximation of the distribution of the viral rebound time. Although numerical methods exist to compute the density of the true first passage time of the transient OU process  $V_t$  [46], here we make the assumption that the LoD  $\ell$  for the VL is much larger than the initial value  $v_0$ , such that we can reasonably approximate the survival function  $S(t) \equiv \mathbb{P}[\tau \geq t]$  with the cumulative density function (CDF)  $\mathbb{P}[V_t < \ell]$  (cf. [18]). This is a valid approximation, since  $V_t$  grows exponentially around the relatively large LoD  $\ell$ . In order to get a probability density function

for  $\tau$ , we simply differentiate the approximated survival function  $S(t)$  with respect to  $t$ :

$$f(t; \lambda, g, v_0, \ell) \propto -\frac{d}{dt} \mathbb{P}[V_t < \ell] = -\frac{d}{dt} \Phi\left(\frac{\ell - (\kappa_1(t) + V_0 e^{gt})}{\sqrt{\kappa_2(t)}}\right) \quad (5)$$

where  $\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{1}{2}x^2} dx$  is the CDF of the standard normal distribution  $\mathcal{N}(0, 1)$ . By expanding Eq 5, we get

$$f(t; \lambda, g, v_0, \ell) = \frac{1}{Z} \times \frac{1}{\sqrt{2\pi\kappa_2(t)}} \exp\left(-\frac{1}{2} \frac{(\ell - (\kappa_1(t) + V_0 e^{gt}))^2}{\kappa_2(t)}\right) \times \left[(\lambda v_0 + V_0 g) e^{gt} + \frac{\ell - (\kappa_1(t) + V_0 e^{gt})}{2\kappa_2(t)} \lambda v_0^2 e^{2gt}\right] \quad (6)$$

To prove that  $f$  is a proper probability distribution, we have to show that  $f$  is non-negative, and we have to find a normalizing constant  $Z$  for Eq 6. The reason that the right-hand-side of Eq 5 does not automatically define a proper probability density function (i.e.  $Z \neq 1$ ) is because the diffusion approximation of  $V_t$  can become negative, and declines exponentially towards  $-\infty$  with a non-zero probability. We have to condition that this non-biological event does not occur. The normalizing constant  $Z$  is equal to the probability of ever reaching the LoD  $\ell$ :

$$\begin{aligned} Z &= \lim_{t \rightarrow \infty} \int_{\ell}^{\infty} \frac{1}{\sqrt{2\pi\kappa_2(t)}} \exp\left(-\frac{1}{2} \frac{(\ell - (\kappa_1(t) + V_0 e^{gt}))^2}{\kappa_2(t)}\right) dx \\ &= 1 - \Phi\left(\lim_{t \rightarrow \infty} \frac{\ell - (\kappa_1(t) + V_0 e^{gt})}{\sqrt{\kappa_2(t)}}\right) = \Phi\left(\sqrt{\frac{2\lambda}{g}} + \frac{V_0}{v_0} \sqrt{\frac{2g}{\lambda}}\right) \end{aligned} \quad (7)$$

The fact that  $f$  is non-negative follows from a simple calculation, where we have to make the reasonable assumption the viral load at time  $t = 0$  is below the limit of detection ( $V_0 < \ell$ ):

$$\begin{aligned} f(t; \lambda, g, v_0, \ell) \geq 0 &\Leftrightarrow (\lambda v_0 + V_0 g) \kappa_2(t) + (\ell - (\kappa_1(t) + V_0 e^{gt})) \frac{1}{2} \lambda v_0^2 e^{gt} \geq 0 \\ &\Leftrightarrow (\lambda v_0 + V_0 g)(e^{2gt} - 1) + (\ell g - \lambda v_0(e^{gt} - 1) - V_0 g e^{gt}) e^{gt} \geq 0 \\ &\stackrel{V_0 < \ell}{\Leftrightarrow} (\lambda v_0 + V_0 g)(e^{2gt} - 1 - e^{gt}(e^{gt} - 1)) \geq 0 \\ &\Leftrightarrow e^{gt} \geq 1 \end{aligned}$$

which is true for all non-negative  $t$ .

The expression for the rebound-time distribution  $f$ , given  $\ell$ , allows for estimation of the parameters  $\lambda$ ,  $v_0$  and  $g$  by maximization of the likelihood or other inference methods. Notice that Eqs 6 and 7 somewhat simplify when we take the initial condition to be  $V_0 = 0$ .

However, to justify that we can replace  $V_t$  with a recurrent OU process, and hence approximate its distribution with a Gaussian, we have to assume that  $v_0$  is relatively small compared to  $V_t$  (see S1 Text). This means that taking the initial condition  $V_0 = 0$  might be problematic. In Fig 2 we compare simulated rebound times with the approximated rebound-time distribution  $f(t; \lambda, g, v_0, \ell)$  where we have taken  $V_0 = 0$ . For large  $\lambda$  the approximation and simulations are in good correspondence, but when  $\lambda$  is small we find a discrepancy. Below we solve this by taking an initial value  $V_0 > 0$ .

## A mixed effects model for treatment-interruption data

Above, we derived our main tool for analyzing viral rebound data: a probability distribution for the time to viral rebound. However, in order to apply this to our SIV rebound data, we need additional statistical methodology, which we develop here. As the VL can only be observed periodically, in any treatment-interruption study the time of viral rebound  $\tau$  is doomed to be interval-censored or right-censored. The viral dynamics after an interval-censored rebound event can be used to narrow the window in which this event occurred [19]. As the VL reaches its peak, the growth rate slows down. Therefore, using a model of pure exponential growth could easily underestimate the initial growth rate. To avoid this we use a logistic growth model with carrying capacity  $K$  to infer the exponential growth rate  $g$  and the time-to-rebound  $\tau$  from the VL time series. Hence, at  $t$  days after treatment interruption, the model predicts a VL equal to

$$\hat{V}(t) = \frac{\ell}{\ell/K + (1 - \ell/K)e^{-g(t-\tau)}} \quad (8)$$

such that  $\hat{V}(\tau) = \ell$ . To model a proportional measurement error [47], we assume that the observed VL has a log-normal distribution around the predicted value:

$\log V(t) \sim \mathcal{N}(\log \hat{V}(t), \sigma^2)$ . The likelihood of a left-censored observation (i.e. the VL is below the LoD) is replaced by the cumulative density of the normal distribution.

To account for the limited number of observations, we use random and mixed effects for the parameters  $K$ ,  $g$  and  $\lambda$ . Since we know that the time of treatment initiation ( $t_{\text{ART}}$ ) is a predictor for both  $\lambda$  and  $g$ , we define

$$\log g = \alpha_g t_{\text{ART}}^* + \epsilon_g, \quad \log \lambda = \alpha_\lambda t_{\text{ART}}^* + \epsilon_\lambda \quad (9)$$

where  $\epsilon_g$  and  $\epsilon_\lambda$  are normally-distributed random effects (a standard assumption), the variable  $t_{\text{ART}}^*$  is the standardized treatment initiation time, and  $\alpha_g$  and  $\alpha_\lambda$  are fixed effects.

All we have to do now is describe a model for the parameter  $\tau$ —the rebound time. For this we consider three different scenarios.

**The multiple-reactivation model.** In order to split the effect of the first reactivation event from subsequent events, we explicitly model the first reactivation time  $T_1 \sim \text{Exp}(\lambda)$ . The likelihood of the difference  $\tau - T_1$  is then given by Eq 6, with initial condition  $V_0 = v_0$ . The parameter  $v_0$  is modeled as a fixed effect, and we chose a prior distribution around the estimates for macaques reported previously [26]. The prior distributions and hyper-parameters for all the model's parameters are listed in Table 2. We chose broad prior distributions for all the (hyper) parameters; notice that the prior distributions are defined on a logarithmic scale.

**The single-reactivation model.** Eqs 8 and 9 remain valid for the single-reactivation model and the reactivation time  $T_1$  is again assumed to be exponentially distributed with rate  $\lambda$ . However, the difference  $\tau - T_1$  now has a Dirac-delta distribution, as it is completely determined by  $g$ ,  $v_0$  and  $\ell$ :

$$\tau - T_1 = \frac{1}{g} \log \left( \frac{\ell}{v_0} \right) \quad (10)$$

To account for this, the rebound time  $\tau$  is no longer a free parameter in the single-reactivation model, but instead defined by Eq 10.

**The conditionally-deterministic multiple-reactivation model.** The approximation for the multiple-reactivation model that was developed by Pinkevych *et al.* [21] is deterministic after the first recrudescence event. The time between this first event and rebound can be



derived using Eq 2 from solving  $\tau - t_0$  from  $\tilde{V}_\tau = \ell$ , which leads to

$$\tau - t_0 = \frac{1}{g} \log \left[ \left( 1 - e^{-g/\lambda} \right) \frac{\ell}{v_0} + e^{-g/\lambda} \right]$$

As in the case of the single-reactivation model, we let  $t_0 = T_1 \sim \text{Exp}(\lambda)$ .

**Model comparison.** In order to statistically compare the three different models, we calculated the Watanabe–Akaike information criterion (WAIC; [27]) as

$$\text{WAIC} = -2 \sum_i (\log \langle \mathcal{L}(D_i | p_s) \rangle_s - \text{Var}_s [\log \mathcal{L}(D_i | p_s)]) \quad (11)$$

where the index  $i$  runs through the observations (i.e. VL measurements), and  $s$  runs through the Monte-Carlo samples from the posterior distribution. The function  $\mathcal{L}(D_i | p_s)$  denotes the likelihood of observation  $D_i$  given parameters  $p_s$ . Moreover, we write  $\langle x_s \rangle_s$  for the sample mean of  $x$  and  $\text{Var}_s[x_s]$  for the sample variance. The results of the model comparisons are listed in S2 Table.

The mixed-effects model is implemented in the probabilistic programming language Stan [48]. For each model, we ran 4 independent chains of length 5000 and 1 : 20 thinning, resulting in a 1000 samples from the posterior distribution. The Gelman-Rubin statistic  $\hat{R}$  was close to 1 for all parameters, indicating good convergence of the chains. The scripts and data used for the analyses can be downloaded from <https://github.com/lanl/multiple-reactivation-model>.

## Supporting information

**S1 Fig. Model fits, used data points and posterior predicted rebound time distributions for all macaques.** The panels (DPI: days post infection) show the VL data (black dots connected by black lines, with red dots for left-censored observations; the grey dots are ignored) taken from all 25 macaques for whom rebound was observed, and the stochastic multiple-reactivation model prediction (blue lines: posterior mean; dark blue band: 50% credible interval (CrI), light blue band: 50% posterior predictive interval). The estimated time-to-rebound ( $\tau$ ) is given by the vertical black line. The density plots in the background indicate the posterior predictive distribution of  $\tau$ . The green distributions are conditioned on the estimated time of the initial recrudescence event, the purple distributions are unconditional.  
(PDF)

**S2 Fig. Marginal posterior densities of the first recrudescence times.** Marginal densities of  $T_1$  (blue) and the extrapolated  $\hat{T}_1$  (red) for each macaque are estimated with our multiple-reactivation model. The numbers on top indicate the time of ART initiation.  
(PDF)

**S3 Fig. Comparison between simulated rebound times and an alternative approximation for the time-to-rebound distribution.** In this case, the law of  $V_t$  is approximated with a Gamma distribution with mean  $\kappa_1(t)$  and variance  $\kappa_2(t)$ . The simulated empirical distributions are shown in color, and our approximation is shown in black. The predicted PDF (A) is calculated with numerical differentiation. (B) The survival function (i.e. the fraction of subjects  $S(t)$  that do not have a detectable VL at time  $t$ ) is defined by Eq S5 in S1 Text. For the top, middle, and bottom panels different values of  $\lambda$  are used ( $\lambda = 5 \text{ d}^{-1}$ ,  $1 \text{ d}^{-1}$ , and  $0.2 \text{ d}^{-1}$  respectively). Notice the different time scale on the horizontal axes. For the remaining parameters, we used the values:  $g = 0.5 \text{ d}^{-1}$ ,  $v_0 = 0.1 \text{ copies mL}^{-1}$ , LoD  $\ell = 50 \text{ copies mL}^{-1}$ .  
(PDF)

**S4 Fig. Comparison between simulated rebound times and an alternative approximation for the time-to-rebound distribution.** In this case, the master equation is approximated using the WKB *ansatz*. The simulated empirical distributions are shown in color, and our approximation is shown in black. (A) The probability density function (PDF; defined by Eq S11 and Eq S10 in [S1 Text](#)). (B) The survival function (i.e. the fraction of subjects  $S(t)$  that do not have a detectable VL at time  $t$ ) is calculated with numerical integration. For the top, middle, and bottom panels different values of  $\lambda$  are used ( $\lambda = 5 \text{ d}^{-1}$ ,  $1 \text{ d}^{-1}$ , and  $0.2 \text{ d}^{-1}$  respectively). Notice the different time scale on the horizontal axes. For the remaining parameters, we used the values:  $g = 0.5 \text{ d}^{-1}$ ,  $v_0 = 0.1 \text{ copies mL}^{-1}$ ,  $\text{LoD } \ell = 50 \text{ copies mL}^{-1}$ . (PDF)

**S5 Fig. Example realizations (in blue) of the generalized viral load process  $V_t$  with clone-specific growth rates given by Eq S12 in [S1 Text](#).** The black curve shows the expected value  $\mathbb{E}[V_t] = \kappa_1$  (Eq S15). The inset shows the probability density function of the random growth rate  $G_t$ . The used parameter values are  $g = 0.5 \text{ d}^{-1}$ ,  $\sigma_G = 0.05 \text{ d}^{-1}$  (corresponding to  $u \approx 0.175$ ),  $v_0 = 0.1 \text{ copies mL}^{-1}$ , and  $\lambda = 1 \text{ d}^{-1}$ . (PDF)

**S6 Fig. Comparison between simulated rebound times and an approximation for the time-to-rebound distribution.** This model allows for variation in the exponential growth rate. The law of  $V_t$  is approximated with a Gamma distribution with mean  $\kappa_1$  (Eq S15 in [S1 Text](#)) and variance  $\kappa_2$  (Eq S16). The simulated empirical distributions are shown in color, and our approximation is shown in black. The predicted PDF (A) is calculated with numerical differentiation. (B) The survival function (i.e. the fraction of subjects  $S(t)$  that do not have a detectable VL at time  $t$ ) is defined as  $S(t) = \gamma(k, \ell/\eta)$  with  $\gamma$  the regularized incomplete Gamma function with parameters  $\eta = \kappa_2/\kappa_1$  and  $k = \kappa_1^2/\kappa_2$ . For the top, middle, and bottom panels different values of  $\lambda$  are used ( $\lambda = 5 \text{ d}^{-1}$ ,  $1 \text{ d}^{-1}$ , and  $0.2 \text{ d}^{-1}$  respectively). Notice the different time scale on the horizontal axes. For the remaining parameters, we used the values:  $g = 0.5 \text{ d}^{-1}$ ,  $\sigma_G = 0.05 \text{ d}^{-1}$  (corresponding to  $u \approx 0.175$ ),  $v_0 = 0.1 \text{ copies mL}^{-1}$ ,  $\text{LoD } \ell = 50 \text{ copies mL}^{-1}$ . The gray curves correspond to the approximate rebound time distribution with a constant growth rate ( $G \equiv g$ ) and are identical to the black curves in [S3 Fig](#). (PDF)

**S1 Text. Approximations of the process  $V_t$ .** Here we derive the diffusion approximation of  $V_t$  by applying the Kramers-Moyal expansion to the master equation of the stochastic process  $V_t$ . Further, we explore two other approximations of  $V_t$ . First, we substitute the Gaussian distribution of  $V_t$  at time  $t$  with a Gamma distribution. Second, we replace the Kramers-Moyal expansion with the Wentzel-Kramers-Brillouin *ansatz*. Finally, we consider a generalization of the stochastic multiple-reactivation model that takes into account within-host variation in the exponential growth rate. We first derive the CGF, and then use the Gamma-distribution method to again derive an approximate rebound-time distribution for this generalized model. (PDF)

**S1 Table. Parameter estimates and credible intervals.** Parameter estimates from the fully stochastic multi-reactivation model (“rebound” columns). and the acute infection (“acute” columns). The point estimates correspond to the mode of the marginal posterior distributions. (PDF)

**S2 Table. Watanabe-Akaike information criterion for the three models.** The Watanabe-Akaike information criterion (WAIC) is averaged over 10 MCMC runs to account for Monte-Carlo error, which is indicated by the standard error of the mean (SEM). The variance

$\hat{k} \equiv \text{Var}_s[\log \mathcal{L}(D_i|p_s)]$  in [Eq 11](#) can be interpreted as the effective number of parameters. (PDF)

## Acknowledgments

We gratefully acknowledge Garrett T. Nieddu for his technical support.

## Author Contributions

**Conceptualization:** Christiaan H. van Dorp, Jessica M. Conway.

**Data curation:** Dan H. Barouch, James B. Whitney.

**Formal analysis:** Christiaan H. van Dorp.

**Funding acquisition:** James B. Whitney, Alan S. Perelson.

**Investigation:** Christiaan H. van Dorp, Jessica M. Conway, Alan S. Perelson.

**Methodology:** Christiaan H. van Dorp, Jessica M. Conway, Alan S. Perelson.

**Project administration:** Alan S. Perelson.

**Resources:** Dan H. Barouch, James B. Whitney.

**Software:** Christiaan H. van Dorp.

**Supervision:** Jessica M. Conway, Alan S. Perelson.

**Validation:** Christiaan H. van Dorp, Jessica M. Conway, Alan S. Perelson.

**Visualization:** Christiaan H. van Dorp.

**Writing – original draft:** Christiaan H. van Dorp, Jessica M. Conway.

**Writing – review & editing:** Christiaan H. van Dorp, Jessica M. Conway, Alan S. Perelson.

## References

1. Murray AJ, Kwon KJ, Farber DL, Siliciano RF. The latent reservoir for HIV-1: How immunologic memory and clonal expansion contribute to HIV-1 persistence. *J Immunol.* 2016; 197(2):407–417. <https://doi.org/10.4049/jimmunol.1600343> PMID: 27382129
2. Whitney JB, Hill AL, Sanisetty S, Penaloza-MacMaster P, Liu J, Shetty M, et al. Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. *Nature.* 2014; 512(7512):74–77. <https://doi.org/10.1038/nature13594> PMID: 25042999
3. Okoye AA, Hansen SG, Vaidya M, Fukazawa Y, Park H, Duell DM, et al. Early antiretroviral therapy limits SIV reservoir establishment to delay or prevent post-treatment viral rebound. *Nat Med.* 2018; 24(9):1430–1440. <https://doi.org/10.1038/s41591-018-0130-7> PMID: 30082858
4. Colby DJ, Trautmann L, Pinyakorn S, Leyre L, Pagliuzza A, Kroon E, et al. Rapid HIV RNA rebound after antiretroviral treatment interruption in persons durably suppressed in Fiebig I acute HIV infection. *Nat Med.* 2018; 24(7):923–926. <https://doi.org/10.1038/s41591-018-0026-6> PMID: 29892063
5. Whitney JB, Lim SY, Osuna CE, Kublin JL, Chen E, Yoon G, et al. Prevention of SIVmac251 reservoir seeding in rhesus monkeys by early antiretroviral therapy. *Nat Commun.* 2018; 9(1):5429. <https://doi.org/10.1038/s41467-018-07881-9> PMID: 30575753
6. Byraredy SN, Arthos J, Cicala C, Villinger F, Ortiz KT, Little D, et al. Sustained virologic control in SIV+ macaques after antiretroviral and  $\alpha_4\beta_7$  antibody therapy. *Science.* 2016; 354(6309):197–202. <https://doi.org/10.1126/science.aag1276> PMID: 27738167
7. Borducchi EN, Cabral C, Stephenson KE, Liu J, Abbink P, Ng'ang'a D, et al. Ad26/MVA therapeutic vaccination with TLR7 stimulation in SIV-infected rhesus monkeys. *Nature.* 2016; 540(7632):284–287. <https://doi.org/10.1038/nature20583> PMID: 27841870

8. Borducchi EN, Liu J, Nkolola JP, Cadena AM, Yu WH, Fischinger S, et al. Antibody and TLR7 agonist delay viral rebound in SHIV-infected monkeys. *Nature*. 2018; 563(7731):360–364. <https://doi.org/10.1038/s41586-018-0600-6> PMID: 30283138
9. Hill AL, Rosenbloom DI, Goldstein E, Hanhauser E, Kuritzkes DR, Siliciano RF, et al. Real-time predictions of reservoir size and rebound time during antiretroviral therapy interruption trials for HIV. *PLoS Pathog*. 2016; 12(4):e1005535. <https://doi.org/10.1371/journal.ppat.1005535> PMID: 27119536
10. Pinkevych M, Cromer D, Tolstrup M, Grimm AJ, Cooper DA, Lewin SR, et al. HIV reactivation from latency after treatment interruption occurs on average every 5–8 days—implications for HIV remission. *PLoS Pathog*. 2015; 11(7):e1005000. <https://doi.org/10.1371/journal.ppat.1005000> PMID: 26133551
11. Conway JM, Perelson AS, Li JZ. Predictions of time to HIV viral rebound following ART suspension that incorporate personal biomarkers. *PLoS Comput Biol*. 2019; 15(7):1–26. <https://doi.org/10.1371/journal.pcbi.1007229> PMID: 31339888
12. Archin NM, Kirchherr JL, Sung JAM, Clutton G, Sholtis K, Xu Y, et al. Interval dosing with the HDAC inhibitor vorinostat effectively reverses HIV latency. *J Clin Invest*. 2017; 127(8):3126–3135. <https://doi.org/10.1172/JCI92684> PMID: 28714868
13. Sogaard OS, Graversen ME, Leth S, Olesen R, Brinkmann CR, Nissen SK, et al. The depsipeptide romidepsin reverses HIV-1 latency in vivo. *PLoS Pathog*. 2015; 11(9):1–22. <https://doi.org/10.1371/journal.ppat.1005142> PMID: 26379282
14. Peterson CW, Wang J, Deleage C, Reddy S, Kaur J, Polacino P, et al. Differential impact of transplantation on peripheral and tissue-associated viral reservoirs: Implications for HIV gene therapy. *PLoS Pathog*. 2018; 14(4):1–22. <https://doi.org/10.1371/journal.ppat.1006956> PMID: 29672640
15. Vansant G, Bruggemans A, Janssens J, Debyser Z. Block-and-lock strategies to cure HIV infection. *Viruses*. 2020; 12(1). <https://doi.org/10.3390/v12010084> PMID: 31936859
16. Reeves DB, Duke ER, Hughes SM, Pric M, Hladik F, Schiffer JT. Anti-proliferative therapy for HIV cure: a compound interest approach. *Sci Rep*. 2017; 7(1):4011. <https://doi.org/10.1038/s41598-017-04160-3> PMID: 28638104
17. Reeves DB, Duke ER, Wagner TA, Palmer SE, Spivak AM, Schiffer JT. A majority of HIV persistence during antiretroviral therapy is due to infected cell proliferation. *Nat Commun*. 2018; 9(1):4811. <https://doi.org/10.1038/s41467-018-06843-5> PMID: 30446650
18. Hill AL, Rosenbloom DI, Fu F, Nowak MA, Siliciano RF. Predicting the outcomes of treatment to eradicate the latent reservoir for HIV-1. *Proc Natl Acad Sci USA*. 2014; 111(37):13475–13480. <https://doi.org/10.1073/pnas.1406663111> PMID: 25097264
19. Prague M, Gerold JM, Balelli I, Pasin C, Li JZ, Barouch DH, et al. Viral rebound kinetics following single and combination immunotherapy for HIV/SIV. *bioRxiv*. 2019;.
20. Hill AL, Rosenbloom DI, Siliciano JD, Siliciano RF. Insufficient evidence for rare activation of latent HIV in the absence of reservoir-reducing interventions. *PLoS Pathog*. 2016; 12(8):e1005679. <https://doi.org/10.1371/journal.ppat.1005679> PMID: 27560936
21. Pinkevych M, Kent SJ, Tolstrup M, Lewin SR, Cooper DA, Sogaard OS, et al. Modeling of experimental data supports HIV reactivation from latency after treatment interruption on average once every 5–8 days. *PLoS Pathog*. 2016; 12(8):e1005740. <https://doi.org/10.1371/journal.ppat.1005740> PMID: 27560972
22. Pearson JE, Krapivsky P, Perelson AS. Stochastic theory of early viral infection: continuous versus burst production of virions. *PLoS Comput Biol*. 2011; 7(2):1–17. <https://doi.org/10.1371/journal.pcbi.1001058> PMID: 21304934
23. Conway JM, Konrad BP, Coombs D. Stochastic analysis of pre- and postexposure prophylaxis against HIV infection. *SIAM J Appl Math*. 2013; 73(2):904–928. <https://doi.org/10.1137/120876800>
24. De Scheerder MA, Vrancken B, Dellicour S, Schlub T, Lee E, Shao W, et al. HIV rebound is predominantly fueled by genetically identical viral expansions from diverse reservoirs. *Cell Host Microbe*. 2019; 26(3):347–358. <https://doi.org/10.1016/j.chom.2019.08.003> PMID: 31471273
25. Fennessey CM, Pinkevych M, Immonen TT, Reynaldi A, Venturi V, Nadella P, et al. Genetically-bar-coded SIV facilitates enumeration of rebound variants and estimation of reactivation rates in nonhuman primates following interruption of suppressive antiretroviral therapy. *PLoS Pathog*. 2017; 13(5):e1006359. <https://doi.org/10.1371/journal.ppat.1006359> PMID: 28472156
26. Pinkevych M, Fennessey CM, Cromer D, Tolstrup M, Sogaard OS, Rasmussen TA, et al. Estimating initial viral levels during Simian Immunodeficiency Virus/Human Immunodeficiency Virus reactivation from latency. *J Virol*. 2018; 92(2):e01667–17. <https://doi.org/10.1128/JVI.01667-17> PMID: 29118123
27. Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J Mach Learn Res*. 2010; 11:3571–3594.

28. Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc.* 1995; 90(430):773–795. <https://doi.org/10.1080/01621459.1995.10476572>
29. Ramratnam B, Bonhoeffer S, Binley J, Hurley A, Zhang L, Mittler JE, et al. Rapid production and clearance of HIV-1 and hepatitis C virus assessed by large volume plasma apheresis. *The Lancet.* 1999; 354(9192):1782–1785. [https://doi.org/10.1016/S0140-6736\(99\)02035-8](https://doi.org/10.1016/S0140-6736(99)02035-8) PMID: 10577640
30. Jones BR, Kinloch NN, Horacek J, Ganase B, Harris M, Harrigan PR, et al. Phylogenetic approach to recover integration dates of latent HIV sequences within-host. *Proc Natl Acad Sci USA.* 2018; 115(38): E8958–E8967. <https://doi.org/10.1073/pnas.1802028115> PMID: 30185556
31. Queen SE, Mears BM, Kelly KM, Dorsey JL, Liao Z, Dinoso JB, et al. Replication-competent simian immunodeficiency virus (SIV) Gag escape mutations archived in latent reservoirs during antiretroviral treatment of SIV-infected macaques. *J Virol.* 2011; 85(17):9167–9175. <https://doi.org/10.1128/JVI.00366-11> PMID: 21715484
32. Li JZ, Etemad B, Ahmed H, Aga E, Bosch RJ, Mellors JW, et al. The size of the expressed HIV reservoir predicts timing of viral rebound after treatment interruption. *AIDS.* 2016; 30(3):343–353. <https://doi.org/10.1097/QAD.0000000000000953> PMID: 26588174
33. Hataye JM, Casazza JP, Best K, Liang CJ, Immonen TT, Ambrozak DR, et al. Principles governing establishment versus collapse of HIV-1 cellular spread. *Cell Host Microbe.* 2019; 26(6):748–763. <https://doi.org/10.1016/j.chom.2019.10.006> PMID: 31761718
34. Mendoza P, Gruell H, Nogueira L, Pai JA, Butler AL, Millard K, et al. Combination therapy with anti-HIV-1 antibodies maintains viral suppression. *Nature.* 2018; 561(7724):479–484. <https://doi.org/10.1038/s41586-018-0531-2> PMID: 30258136
35. Crowell TA, Colby DJ, Pinyakorn S, Sacdalan C, Pagliuzza A, Intasan J, et al. Safety and efficacy of VRC01 broadly neutralising antibodies in adults with acutely treated HIV (RV397): a phase 2, randomised, double-blind, placebo-controlled trial. *Lancet HIV.* 2019; 6(5):e297–e306. [https://doi.org/10.1016/S2352-3018\(19\)30053-0](https://doi.org/10.1016/S2352-3018(19)30053-0) PMID: 31000477
36. Lim SY, Osuna CE, Hraber PT, Hesselgesser J, Gerold JM, Barnes TL, et al. TLR7 agonists induce transient viremia and reduce the viral reservoir in SIV-infected rhesus macaques on antiretroviral therapy. *Sci Transl Med.* 2018; 10(439). <https://doi.org/10.1126/scitranslmed.aao4521> PMID: 29720451
37. Pinkevych M, Fennessey CM, Cromer D, Reid C, Trubey CM, Lifson JD, et al. Predictors of SIV recrudescence following antiretroviral treatment interruption. *eLife.* 2019; 8:e49022. <https://doi.org/10.7554/eLife.49022> PMID: 31650954
38. de Souza MS, Pinyakorn S, Akapirat S, Pattanachaiwit S, Fletcher JL, Chomchey N, et al. Initiation of antiretroviral therapy during acute HIV-1 infection leads to a high rate of nonreactive HIV serology. *Clin Infect Dis.* 2016; 63(4):555–561. <https://doi.org/10.1093/cid/ciw365> PMID: 27317797
39. Manak MM, Jagodzinski LL, Shutt A, Malia JA, Leos M, Ouellette J, et al. Decreased seroreactivity in individuals initiating antiretroviral therapy during acute HIV infection. *J Clin Microbiol.* 2019; 57(10). <https://doi.org/10.1128/JCM.00757-19> PMID: 31217270
40. Ndhlovu ZM, Kazer SW, Nkosi T, Ogunshola F, Muema DM, Anmole G, et al. Augmentation of HIV-specific T cell function by immediate treatment of hyperacute HIV-1 infection. *Sci Transl Med.* 2019; 11(493). <https://doi.org/10.1126/scitranslmed.aau0528> PMID: 31118290
41. Cao Y, Cartwright EK, Silvestri G, Perelson AS. CD8+ lymphocyte control of SIV infection during antiretroviral therapy. *PLoS Pathog.* 2018; 14(10):e1007350. <https://doi.org/10.1371/journal.ppat.1007350> PMID: 30308068
42. Norris JR. Continuous-time Markov chains I. In: *Markov Chains.* Cambridge: Cambridge University Press; 1997. p. 60–107.
43. Privault N. *Understanding Markov chains.* 2nd ed. Singapore: Springer; 2018.
44. van Kampen NG. *Stochastic processes in physics and chemistry.* 3rd ed. Amsterdam: Elsevier; 2007.
45. Steele JM. *Stochastic calculus and financial applications.* New York: Springer; 2001.
46. Alili L, Patie P, Pedersen JL. Representations of the first hitting time density of an Ornstein-Uhlenbeck process. *Stochastic Models.* 2005; 21(4):967–980. <https://doi.org/10.1080/15326340500294702>
47. Lima VD, Wang L, Brumme C, Wu L, Montaner JS, Harrigan PR. Estimation of measurement error in plasma HIV-1 RNA assays near their limit of quantification. *PLoS One.* 2017; 12(2):e0171155. <https://doi.org/10.1371/journal.pone.0171155> PMID: 28152073
48. Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, et al. Stan: A probabilistic programming language. *J Stat Softw.* 2017; 76(1):1–32. <https://doi.org/10.18637/jss.v076.i01>