

Article

# Time Intervals in Sequence Sampling, Not Data Modifications, Have a Major Impact on Estimates of HIV Escape Rates

Vitaly V. Ganusov <sup>1,2</sup> 

<sup>1</sup> Department of Microbiology, University of Tennessee, Knoxville, TN 37996, USA; vitaly.ganusov@gmail.com

<sup>2</sup> Department of Mathematics, University of Tennessee, Knoxville, TN 37996, USA

Received: 9 December 2017; Accepted: 22 February 2018; Published: 27 February 2018

**Abstract:** The ability of human immunodeficiency virus (HIV) to avoid recognition by humoral and cellular immunity (viral escape) is well-documented, but the strength of the immune response needed to cause such a viral escape remains poorly quantified. Several previous studies observed a more rapid escape of HIV from CD8 T cell responses in the acute phase of infection compared to chronic infection. The rate of HIV escape was estimated with the help of simple mathematical models, and results were interpreted to suggest that CD8 T cell responses causing escape in acute HIV infection may be more efficient at killing virus-infected cells than responses that cause escape in chronic infection, or alternatively, that early escapes occur in epitopes mutations in which there is minimal fitness cost to the virus. However, these conclusions were challenged on several grounds, including linkage and interference of multiple escape mutations due to a low population size and because of potential issues associated with modifying the data to estimate escape rates. Here we use a sampling method which does not require data modification to show that previous results on the decline of the viral escape rate with time since infection remain unchanged. However, using this method we also show that estimates of the escape rate are highly sensitive to the time interval between measurements, with longer intervals biasing estimates of the escape rate downwards. Our results thus suggest that data modifications for early and late escapes were not the primary reason for the observed decline in the escape rate with time since infection. However, longer sampling periods for escapes in chronic infection strongly influence estimates of the escape rate. More frequent sampling of viral sequences in chronic infection may improve our understanding of factors influencing the rate of HIV escape from CD8 T cell responses.

**Keywords:** human immunodeficiency virus (HIV); cytotoxic T lymphocyte (CTL) response; viral escape; mathematical models

## 1. Introduction

Due to the absence of a proof-reading capability, human immunodeficiency virus (HIV)-1 (HIV hereafter) has a relatively high average mutation rate of  $2\text{--}4 \times 10^{-5}$  per bp per replication cycle [1–5]; however, even higher mutation rates have been suggested [5,6]. Given a relatively short genome ( $\sim 10^4$  bp), mutant viruses arise only in 20–40% of replication cycles. Even with such perhaps surprising fidelity given a short virus replication cycle time ( $\sim 1\text{--}2$  days, [7]) and large population size ( $\sim 10^7\text{--}10^8$  of virus-infected cells [8,9]), HIV is able to generate multiple variants during the course of infection. Some of these variants will be unrecognized by the host's immune responses, resulting in escape, which may prevent viral control. Indeed, the ability of HIV to escape from cellular (CD8 T cell) and humoral (antibody) responses is often defined as the major obstacle preventing the development of efficient vaccines [10–13]. One of the early mathematical models of why HIV-infected individuals progress to acquired immunodeficiency syndrome (AIDS) also involves viral escape from immunity [14,15].

While HIV clearly escapes from both cellular and humoral immune responses, frequent sampling of the escape dynamics suggested that escape from virus-specific CD8 T cell responses occurs earlier than from antibody-mediated responses [16–19]. Escape from CD8 T cell responses has been accurately mapped for several HIV-infected individuals from the earliest stages of infection [16,19,20]. Specifically, these studies tracked changes in the HIV genome from the virus sequence founding the infection (transmitted/founder virus) and mapped CD8 T cell response specific to the whole viral proteome. Genetic changes in regions recognized by CD8 T cells was treated as evidence of T cell recognition of the virus, and thus the speed at which the founder viral sequence was lost was indicative of T cell response strength. Multiple studies developed mathematical models to estimate the rate at which founding wild-type virus was lost from the population (and consequently, the escape variants accumulated in the population); this escape rate was then interpreted as the rate at which epitope-specific CD8 T cells eliminate HIV-infected cells [21–27]. Estimation of HIV escape rates was deemed complicated due to the sparse nature of the sequence data coming from the relatively infrequent sampling of virus populations and a low number of viral sequences obtained at a given time point in most studies. In particular, many of the data on escape measuring frequency of the wild-type (or escape) variant include measurements of only either 100% or 0% in two sequential time points (or in some cases, with one intermediate frequency of the wild-type virus between the two extreme values of 100% and 0%). To estimate HIV escape rates in such data, the data were modified to include some presence of the escape variant or the wild-type virus [21,22]. For example, if out of 10 sequences of a specific epitope region 10 sequences were of the wild-type virus, it was proposed to add 1 extra sequence with a mutated epitope, resulting in the frequency of the wild-type  $f_w = 10/11 = 91\%$  instead of the observed 100%. Similarly, if 5 out of 5 sequences at the subsequent time point were mutated, adding 1 wild-type sequence resulted in the wild-type variant frequency of  $f_w = 1/6 = 17\%$  instead of 0%. Such modifications were proposed to provide minimal escape rates [21,25]. It remained unknown whether such data modifications were important for making general observations regarding HIV escape rates; previous studies utilizing such data modifications suggested that the rate of HIV escape declines with time since infection, implying a potential weakening of the immune responses or a simple consequence of the virus escaping in the least costly positions [25,28]. More recent studies have highlighted that the proposed data modification may have had a strong impact on estimated escape rates, thus questioning the validity of previous conclusions [29,30]. Here we extend previous analyses by applying a novel method of estimating rates of viral escape which does not involve modification of the data, using previously published data. The method uses sampling of the sequence data using the beta distribution, which serves as a continuous approximation of the binomial distribution. With this sampling method, we show that at least for three HIV-infected patients, previous data modifications were not responsible for the predicted decline of the escape rate with time since infection. However, the sampling method remains highly sensitive to the time frequency of data sampling—the method estimates slower escapes for less-frequently-sampled data. This analysis suggests that better understanding of the mechanisms behind HIV escape from CD8 T cell responses is not likely to come from improved methods of sequence data analysis, but from better data with improved time resolution and depth of sequencing.

## 2. Materials and Methods

**Data.** Experimental data used in this paper are from previous publications [16,25]. In short, individuals with recent HIV-1 infection were recruited into the study. Patients donated blood at regular time intervals, and viral RNA sequences were obtained using single genome amplification (SGA) techniques. Three patients from the Center for HIV/AIDS Vaccine Immunology (CHAVI) were analyzed: CH40, CH77, and CH58. These patients were infected with a single transmitted/founder virus, and changes in the viral genome were mapped to the HIV-specific cytotoxic T lymphocyte (CTL) responses. In viral sequences, there were also changes that had signatures of viral escape from CTL responses, but no CTL responses specific to proteins in these specific regions have been detected [16]. We re-analyzed viral sequence data for all detected escapes and restricted some analyses to only

escapes from detected CTL responses. Subjects are controlled in accordance with the tenets of the Declaration of Helsinki.

**Model.** We used a previously suggested mathematical model of viral escape from a single CTL response [21–23,25]. The model was fit to experimental data using likelihood method based on binomial distribution (Equation (2) [31]).

**Statistics.** When fitting mathematical models to experimental data, it is important to estimate confidence intervals for the model parameters. One widely-used approach is bootstrap, in which either experimental data or residuals (difference between model prediction and the data) are sampled with replacement [32]. Two types of bootstraps exist: non-parametric and parametric bootstrap. In nonparametric bootstrap, data are sampled with replacement into a bootstrap sample and then the model is fitted to the sample dataset. For example, one could resample with replacement residuals from the model fit of the data, add resampled residuals to the model predictions to generate a sampled dataset, and fit the model to these sampled data. In contrast, in parameteric bootstrap, one assumes a particular distribution underlying the data and generates samples using this distribution. For example, residuals resulting from a given model fit may follow a normal distribution with zero mean and variance  $\sigma^2$ . Therefore, using normal distribution  $N(0, \sigma^2)$ , one could generate another bootstrap sample by adding randomly generated residuals to predictions of the model best fit [32]. In our analysis, we use a version of the parametric bootstrap whereby we generate samples of data assuming that escape variant frequency follows a beta distribution with  $m$  mutant and  $w$  wild-type sequences. A beta distribution is a continuous approximation of the binomial distribution, and the beta distribution has been used to calculate confidence intervals for proportions [33].

**Programming.** Model fits of the data were done in Mathematica 11 using a routine FindMinimum. All codes are available from the author upon request. An example of a Mathematica notebook illustrating the strategy of the sampling method to estimate escape rates is available as a supplement to this paper.

### 3. Results

#### 3.1. A Sampling Method to Estimate Escape Rates without Data Modification

Several previous studies have shown that the kinetics of viral escape from a single CD8 T cell response can be described by the logistic equation

$$f(t) = \frac{f_0}{f_0 + (1 - f_0)e^{-\epsilon t}}, \quad (1)$$

where  $f_0$  and  $f(t)$  are the frequency of the escape variant at time  $t = 0$  and time  $t$ , respectively, and  $\epsilon$  is the escape rate [21–23,25]. When the escape variant is generated by mutation from the wild type, the predicted dynamics of the frequency of the escape variant is approximately given by the logistic equation [31]. Because the generation of escape variants is a stochastic process, this simple model can be interpreted as describing virus dynamics when an escape variant reached some sufficient frequency when the importance of genetic drift on virus dynamics becomes negligible. The time at which the wild-type or mutant variant reaches 50% is called  $t_{50} = \frac{1}{\epsilon} \log \left( \frac{1-f_0}{f_0} \right)$  [25]. The use of a deterministic model that ignores the stochasticity of virus dynamics could in part be justified by the large population size of HIV during acute and early HIV infection, reaching  $10^7 - 10^9$  infected CD4 T cells [8,9] and high virus mutation rate on the order of  $\sim 10^{-5}$  [1–6].

Data on viral escape are generally given as the percent of escape variant in the viral population at different time points with a number of viral sequences obtained. For example, at time 10 days post symptoms, 15 viral sequences were obtained, and 3 of these contained mutated epitopes, making the frequency of the escape variant 3/15 or 20%. To estimate escape rate it, is advised to use a maximum

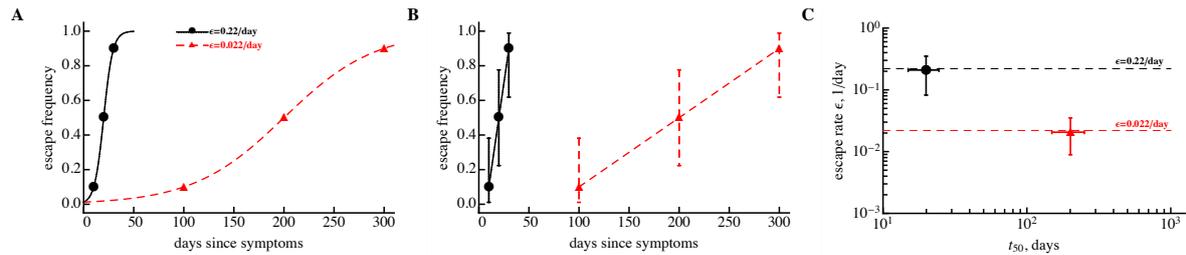
likelihood approach [31] in which the binomial distribution-based likelihood of the model given experimental data is given by

$$L \sim \prod_{i=1}^n f(t_i)^{m_i} (1 - f(t_i))^{w_i}, \quad (2)$$

where  $m_i$  and  $w_i$  are the number of escape mutant and wild-type sequences in the  $i$ th sample taken at time  $t_i$  and there are  $n$  time points. Note that we ignored constant terms for the binomial distribution since these do not matter during likelihood maximization. Parameters of the model including the escape rate  $\varepsilon$  and initial variant frequency  $f_0$  are then found by maximizing likelihood [31]. As discussed in the introduction, the maximization of the likelihood function Equation (2) fails for finite values of the escape rate if the data do not have measurements in which both wild-type and mutant sequences are observed. In some cases when there is only one time point at which both wild-type and mutant viruses are detected and for all other time points only one variant is observed, it is possible to estimate the escape rate without data modification [34].

To obtain an estimate of the escape rate for the data when there are only wild-type or only mutant variants detected at a given time point, we propose to use a parameteric bootstrap approach. In this approach, we do not fit the basic mathematical model (Equation (1)) to the actual data. Instead, we generate pseudorandom samples of the data which are found using a beta distribution for every sample in every time point [33]. Specifically, for the  $N_i = w_i + m_i$  sampled sequences at time  $t_i$ , where  $m_i$  is the number of escape variant sequences and  $w_i$  is the number of wild-type sequences, the resampled frequency of the escape variant  $x_i$  is given by the beta distribution  $Beta(m_i + 1/2, w_i + 1/2)$  [33]. This procedure is repeated for all  $n$  time points, resulting in a bootstrap sample. In our analyses, we used the routine `Random` in Mathematica. Due to randomness, the resulting bootstrapped sample will have both wild-type and escape variants present at multiple time points. The model can then be fitted to these bootstrapped data using the likelihood approach according to Equation (2). Repeating this sampling procedure a given number of times (e.g.,  $10^3$  simulations), a distribution of escape rates, initial mutant frequency, and the time to 50% of the escape variants can be generated. The mean (or median) of these distributions and 95% interval of the distribution can be then used to characterize the escape rate  $\varepsilon$  and the time to 50% escape variant  $t_{50}$ .

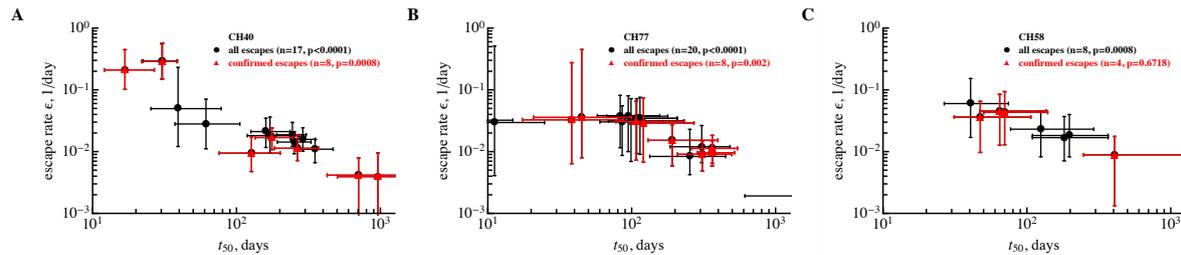
While the sampling method was specifically designed to estimate escape rates from the data which cannot be fit by simple mathematical models (e.g., as in Equation (1)), it was important to confirm that for well-sampled data the method was able to recover the escape rate that would be found by fitting the models directly to such data. As a proof of principle, we simulated viral escape using the logistic model (Equation (1)) and assumed that it was possible to detect both wild-type and escape variant viruses at three sequential time points (Figure 1). Obviously, the logistic model can well fit these data and recover a correct escape rate value (Figure 1A) without the need for additional methods. Importantly, however, sampling these data using the beta distribution and estimating the escape rate for these bootstrap samples allowed for a relatively accurate estimate of the fast and slow escape rates (Figure 1C). Additional simulations with more randomly chosen escape rates and samplings confirmed that the method of sampling recovered the same escape as the model fitted directly to the escape data. The results of these analyses were similar if we used data in which both wild-type and escape variants were present at two sequential time points (results not shown). In our simulations, sampling using beta distribution never yielded a 0% or 100% for the wild-type virus, and thus always allowed us to find a finite value for the average escape rate.



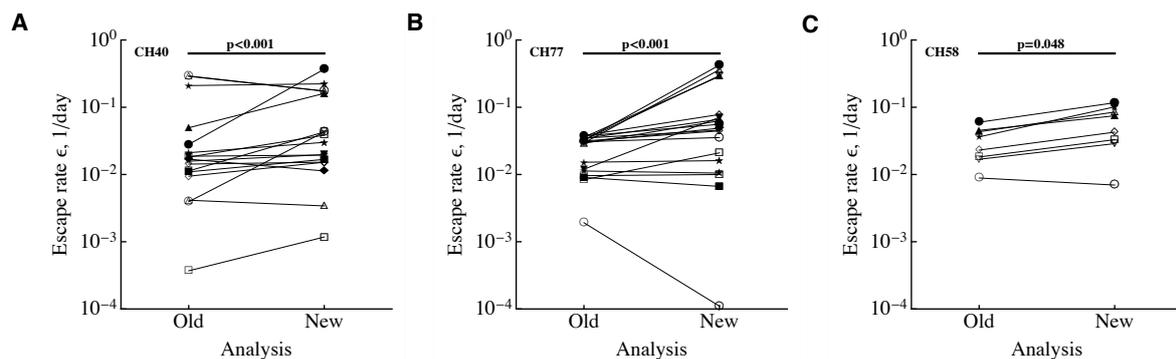
**Figure 1.** A sampling method accurately recovers virus escape rate for data when both wild-type and escape variant are detected at multiple sequential time points. We simulated viral escape using a logistic equation with two escape rates:  $\epsilon = 0.22/\text{day}$  (bullets) and  $\epsilon = 0.022/\text{day}$  (triangles). Three data points at which the escape variant was present at 10%, 50%, and 90% out of  $N = 10$  sequences were generated (panel A). These simulated data were fitted by the simple logistic model (Equation (1)) using a likelihood approach, leading to expected escape rates (lines in panel A). These data were then resampled using the beta distribution (panel B, black solid line is for  $\epsilon = 0.22/\text{day}$  and red dashed line is for  $\epsilon = 0.022/\text{day}$ ) and fitted by the same model using the likelihood method. Resulting averages and their 95% confidence intervals match well with the numbers estimated using actual data (panel C). Dashed lines in panel (C) show estimates of the escape rates found in panel (A), and 95% confidence intervals in panel (B) are predicted using Jeffrey's intervals [33].

### 3.2. The Rate of HIV Escape Declines with Time since Infection

Previous work from several groups suggested that during HIV infection the rate of HIV escapes from T cell responses declines over time [25,28]. However, this conclusion was challenged on multiple levels, including the need for modification of the data to provide minimal estimates of the escape rate [29]. Indeed, it was possible that for many late escapes, previous studies provided minimal estimates of the escape rate and thus the observed pattern of the escape rate decline with time since infection was an artifact of data modification [26,29,30]. To address this specific point, we re-analyzed previously published data on HIV escape [16,25]. Specifically, we used data for all documented escapes in three patients from the Center for HIV/AIDS Vaccine Immunology (CHAVI) as described in previous publications, and limited these data to the first 800 days post-infection. This was done because of the bias towards lower estimates the sampling method introduces in datasets with longer sampling times (see Discussion). We then resampled data for all escapes, fitted the simple logistic model (Equation (1)) to these data, and calculated the average escape rate ( $\epsilon$ ) and time of escape ( $t_{50}$ ). This reanalysis revealed a consistent and statistically significant decline in the estimated rate of HIV escape from CD8 T cell responses with the time since infection (Figure 2). This was true for all documented escapes (i.e., escapes which had the pattern of viral escape from T cell responses but may or may not have had measurable T cell response against the wild-type epitope) in two out of three patients with escapes with detected T cell responses against the wild-type epitope. The poor correlation in the latter case in patient CH58 is likely due to a small number of confirmed escapes ( $n = 4$ ). Interestingly, there was a statistically significant correlation between mean escape rates obtained from parametric sampling of the data and previous estimates found with data modifications; however, the novel sampling method consistently provided higher estimates of the escape rate than previously reported (Figure 3) [25]. This is consistent with a recent report that used deep sequencing in patients in acute HIV infection and found higher estimates of the HIV escape rates from CD8 T cell responses [19]. Taken together, our analysis suggests that data modification was not the reason for the detected decline in the escape rate with time since infection [25].



**Figure 2.** A sampling method which does not involve data modification still predicts a decline in the estimated escape rate with the time since infection. We used a novel approach of data sampling and estimated the rate of human immunodeficiency virus (HIV) escape from CD8 T cell responses for three Center for HIV/AIDS Vaccine Immunology (CHAVI) patients using previously published data [16,25]. Rates ( $\epsilon$ ) and time to 50% of the escape variant ( $t_{50}$ ) and associated errors are shown for three different patients in different panels (panel A: CH40, panel B: CH77, panel C: CH58); bullets indicate all escapes (in black) and triangles are for confirmed escapes (in red) [16]. Confirmed escapes were defined as fixed changes in HIV genome with a clear signature of viral escape with detected epitope-specific T cell response. Other escapes had a well-defined signature of escape but no epitope-specific T cell response was detected [16]. Resulting  $p$  values from the Spearman rank correlation test are indicated on individual panels. Panel (B) does not show an estimate for one escape due to very low escape rate  $\epsilon$  and long  $t_{50}$ .

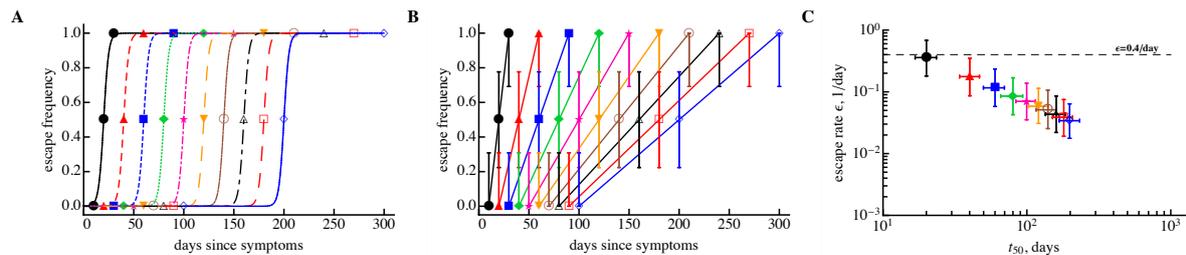


**Figure 3.** A sampling method provides higher estimates of the escape rates. We plot estimates of HIV escape rates in three patients (panel A: CH40, panel B: CH77, panel C: CH58) as was found in a previous study [25] employing data modifications (“Old” analysis) and estimates found using sampling method (“New” analysis). Different markers on plots indicate escape rate estimates for different escapes in our previous analysis [25]. Comparison of the escape rates in two analyses was done using Wilcoxon signed rank test and  $p$  values from these tests are indicated on individual panels.

### 3.3. Time Frequency of Sampling Biases Estimates of the Escape Rate

Our results so far suggest that data modifications were not likely responsible for the observed decline in the estimated escape rate with time since infection. Therefore, it is important to understand whether other features in the data could be responsible for this declining trend. Previous studies have highlighted that the frequency of sampling in experimental data have changed over time, starting from weekly sampling and changing to monthly/bi-monthly sampling [16]. Such a pattern may naturally lead to underestimates of escape rates in part due to clonal interference between different escape variants [26,29,30]. Therefore, we sought to determine if the sampling method was capable of accurately recovering the escape rate independently of the sampling time period. We simulated viral escape with the same escape rate but assumed different initial frequency for the virus and different time intervals between measurements. This allowed wild-type and escape variant to be present in the data only at one time point (Figure 4A). The resulting “data” looked somewhat similar to those observed during HIV infection (Figure 4B). By fitting the logistic model to bootstrapped

samples, we found that the sampling period had a dramatic impact on the estimated escape rate (Figure 4C). In an extreme case, the method underestimated the escape rate 10-fold! Thus, sampling frequency (i.e., weekly, monthly, bi-monthly, etc.) has a dramatic impact on the estimate of the escape rate even in the absence of data modification and when one uses a sampling method. This result has two consequences. First, in order to understand the kinetics of HIV escape from T cell responses, more frequent sampling during chronic HIV infection will be needed. Second, previous results on the independence of the escape rate on the magnitude of the T cell response [25] and the decline of the escape rate with the time since infection need to be re-evaluated using additional, perhaps better sampled data.



**Figure 4.** Time interval between measurements has a strong bias on the estimate of the escape rate and leads to a correlation between the escape rate ( $\epsilon$ ) and time of escape ( $t_{50}$ ). We generated artificial data in which viral sequences were measured in three time points, resulting in 0%, 50%, and 100% of the escape variant with  $N = 10$  sequences per time point (panel A). In all cases, escape can be described by the logistic equation (Equation (1)) assuming a delayed immune response and escape rate  $\epsilon = 0.4$ /day. We resampled these simulated “data” using a beta distribution (panel B, different lines indicate different escapes as in panel A) and estimated the average escape rate (panel C). We show the prediction of the logistic curve obtained using  $\epsilon = 0.4$ /day (panel A), sampled data (panel B), and estimates of the escape rate obtained using parametric bootstrap (panel C). The horizontal dashed line in (panel C) denotes the theoretical escape rate of 0.4/day. The 95% confidence intervals in (panel B) are predicted using Jeffrey’s intervals [33], and in (panel C) CIs were calculated from bootstrapped samples.

#### 4. Discussion

Several lines of evidence suggest that CD8 T cell responses play an important role in control of HIV replication [16,35,36]. Escape of HIV from T cell responses is one type of indirect evidence in support of this hypothesis. Accurate estimates of the rate at which CD8 T cells are eliminating the virus-infected cells (or more generally, suppressing virus replication) and comparison of these rates with other kinetic parameters of HIV replication in vivo may be useful in testing this hypothesis further [22]. The rate at which HIV escapes from CD8 T cell responses may be one metric indicating the strength of the HIV-specific CD8 T cell response. However, multiple studies have now shown that estimating the rate of HIV escape using current data is complicated.

We have proposed a new method of sampling sequence data to estimate escape rates in situations when direct fitting of simple deterministic mathematical models to the data fails. An alternative approach to estimating escape rates from sparsely sampled data would be to use more mechanistic stochastic models of HIV dynamics and evolution; for example, models that take into account virus effective population size, as well as the rate of HIV mutation and recombination [26]. However, it is unclear how well such complex models with multiple parameters represent HIV dynamics in vivo, and as such, biases in parameter estimates associated with such models are not known. Combining the use of such stochastic models with the method of data sampling using the beta distribution may allow quantification of potential biases in parameter estimates in the models and to further understand mechanisms behind HIV escape from CD8 T cell responses.

Using the sampling method, we have shown that the observed phenomenon of a decline of the rate of HIV escape with time since infection is not the result of modification of data to accommodate

fits of simple models of viral escape. However, the method suffers from a great sensitivity to the time frequency of sampling in the data, since increasing the time intervals between measurements naturally decreases the estimated escape rate. Furthermore, the inclusion of more data at later time points in which wild-type virus is nearly lost will also bias average estimates found by sampling downwards, which further limits the applicability of this method for the estimation of escape rates. Specifically, if for the first escape in Figure 1A we added more measurements at later time points (e.g., at day 100 or 300, when the escape variant is likely to be at 100% frequency), sampling would have provided much lower estimates of the escape rate than the  $\varepsilon = 0.22/\text{day}$  found by the model fitting to the original, not resampled data (results not shown). The sampling method also requires more computational power, as at least  $10^3$  samples are generally required to accurately estimate confidence intervals [32]. While fitting a simple logistic model to escape data is generally fast, fitting more complicated models assuming, for example, sequential or concurrent escape, can be time-consuming [34]. Finally, the method relies on the approximation of the binomial distribution by the beta distribution, which may fail in some circumstances [33].

Overall, our results so far suggest that the current state of data on HIV escape does not allow us to make solid inferences on the escape rate without making strong assumptions either about the data or about the underlying dynamics of the virus. It appears that the way forward is to generate novel data in which patient sampling occurs at regular intervals over the course of a few months of infection, perhaps with the use of novel techniques such as deep sequencing. Consistent temporal sampling rather than deep sampling is likely to be more beneficial for determining if the rate of HIV escape from T cell responses declines with time since infection [31].

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/1999-4915/10/3/99/s1>.

**Acknowledgments:** This was primarily supported by the AHA grant (13SDG16960053) to Vitaly V. Ganusov and in part by the NIH grant (R01 GM118553) to Vitaly V. Ganusov. This paper was improved by advice and comments from Yiding Yang, Elizabeth Johnson, and Viktor Zenkov.

**Author Contributions:** Vitaly V. Ganusov conceived and designed the analyses; Vitaly V. Ganusov performed the simulations and analyses; Vitaly V. Ganusov wrote the paper.

**Conflicts of Interest:** The author declares no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of results; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Mansky, L.M.; Temin, H.M. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J. Virol.* **1995**, *69*, 5087–5094.
2. Mansky, L.M. Forward mutation rate of human immunodeficiency virus type 1 in a T lymphoid cell line. *AIDS Res. Hum. Retroviruses* **1996**, *12*, 307–314.
3. Mansky, L.M. Retrovirus mutation rates and their role in genetic variation. *J. Gen. Virol.* **1998**, *79 Pt 6*, 1337–1345.
4. Sanjuán, R.; Nebot, M.R.; Chirico, N.; Mansky, L.M.; Belshaw, R. Viral mutation rates. *J. Virol.* **2010**, *84*, 9733–9748, doi:10.1128/JVI.00694-10.
5. Geller, R.; Domingo-Calap, P.; Cuevas, J.M.; Rossolillo, P.; Negroni, M.; Sanjuán, R. The external domains of the hiv-1 envelope are a mutational cold spot. *Nat. Commun.* **2015**, *6*, 8571, doi:10.1038/ncomms9571.
6. Cuevas, J.M.; Geller, R.; Garijo, R.; López-Aldeguer, J.; Sanjuán, R. Extremely high mutation rate of HIV-1 in vivo. *PLoS Biol.* **2015**, *13*, e1002251, doi:10.1371/journal.pbio.1002251.
7. Perelson, A.S. Modelling viral and immune system dynamics. *Nat. Rev. Immunol.* **2002**, *2*, 28–36.
8. Haase, A.T. Population biology of HIV-1 infection: Viral and CD4+ T cell demographics and dynamics in lymphatic tissues. *Annu. Rev. Immunol.* **1999**, *17*, 625–656.
9. Estes, J.D.; Kityo, C.; Ssali, F.; Swainson, L.; Makamdop, K.N.; Del Prete, G.Q.; Deeks, S.G.; Luciw, P.A.; Chipman, J.G.; Beilman, G.J.; et al. Defining total-body aids-virus burden with implications for curative strategies. *Nat. Med.* **2017**, *23*, 1271–1276, doi:10.1038/nm.4411.

10. Walker, B.D.; Korber, B.T. Immune control of HIV: the obstacles of HLA and viral diversity. *Nat. Immunol.* **2001**, *2*, 473–475, doi:10.1038/88656.
11. Barouch, D.; Kunstman, J.; Kuroda, M.; Schmitz, J.; Santra, S.; Peyerl, F.; Krivulka, G.; Beaudry, K.; Lifton, M.; Gorgone, D.A.; et al. Eventual AIDS vaccine failure in a rhesus monkey by viral escape from cytotoxic T lymphocytes. *Nature* **2002**, *415*, 335–339.
12. Mullins, J.I.; Rolland, M.; Allen, T.M. Viral evolution and escape during primary human immunodeficiency virus-1 infection: Implications for vaccine design. *Curr. Opin. HIV AIDS* **2008**, *3*, 60–66, doi:10.1097/COH.0b013e3282f233d9.
13. Rolland, M.; Tovanabutra, S.; Decamp, A.C.; Frahm, N.; Gilbert, P.B.; Sanders-Buell, E.; Heath, L.; Magaret, C.A.; Bose, M.; Bradfield, A.; et al. Genetic impact of vaccination on breakthrough HIV-1 sequences from the STEP trial. *Nat. Med.* **2011**, *17*, 366–371.
14. Nowak, M.A.; Anderson, R.M.; McLean, A.R.; Wolfs, T.F.; Goudsmit, J.; May, R.M. Antigenic diversity thresholds and the development of AIDS. *Science* **1991**, *254*, 963–969.
15. Alizon, S.; Magnus, C. Modelling the course of an hiv infection: insights from ecology and evolution. *Viruses* **2012**, *4*, 1984–2013, doi:10.3390/v4101984.
16. Goonetilleke, N.; Liu, M.K.; Salazar-Gonzalez, J.F.; Ferrari, G.; Giorgi, E.; Ganusov, V.V.; Keele, B.F.; Learn, G.H.; Turnbull, E.L.; Salazar, M.G.; et al. The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. *J. Exp. Med.* **2009**, *206*, 1253–1272.
17. McMichael, A.J.; Borrow, P.; Tomaras, G.D.; Goonetilleke, N.; Haynes, B.F. The immune response during acute HIV-1 infection: Clues for vaccine development. *Nat. Rev. Immunol.* **2010**, *10*, 11–23.
18. Bar, K.J.; Tsao, C.-Y.; Iyer, S.S.; Decker, J.M.; Yang, Y.; Bonsignori, M.; Chen, X.; Hwang, K.-K.; Montefiori, D.C.; Liao, H.-X.; et al. Early Low-Titer Neutralizing Antibodies Impede HIV-1 Replication and Select for Virus Escape. *PLoS Pathog.* **2012**, *8*, e1002721, doi:10.1371/journal.ppat.1002721.
19. Kijak, G.H.; Sanders-Buell, E.; Chenine, A.-L.; Eller, M.A.; Goonetilleke, N.; Thomas, R.; Leviyang, S.; Harbolick, E.A.; Bose, M.; Pham, P.; et al. Rare hiv-1 transmitted/founder lineages identified by deep viral sequencing contribute to rapid shifts in dominant quasispecies during acute and early infection. *PLoS Pathog.* **2017**, *13*, e1006510, doi:10.1371/journal.ppat.1006510.
20. Liu, M.K.P.; Hawkins, N.; Ritchie, A.J.; Ganusov, V.V.; Whale, V.; Brackenridge, S.; Li, H.; Pavlicek, J.W.; Cai, F.; Rose-Abrahams, M.; et al. Vertical T cell immunodominance and epitope entropy determine HIV-1 escape. *J. Clin. Investig.* **2013**, *123*, 380–393, doi:10.1172/JCI65330.
21. Fernandez, C.; Stratov, I.; De Rose, R.; Walsh, K.; Dale, C.; Smith, M.; Agy, M.; Hu, S.; Krebs, K.; Watkins, D.I.; et al. Rapid viral escape at an immunodominant simian-human immunodeficiency virus cytotoxic T-lymphocyte epitope exacts a dramatic fitness cost. *J. Virol.* **2005**, *79*, 5721–31.
22. Asquith, B.; Edwards, C.; Lipsitch, M.; McLean, A. Inefficient cytotoxic T lymphocyte-mediated killing of HIV-1-infected cells in vivo. *PLoS Biol.* **2006**, *4*, e90.
23. Ganusov, V.V.; De Boer, R.J. Estimating costs and benefits of CTL escape mutations in SIV/HIV Infection. *PLoS Comput. Biol.* **2006**, *2*, e24.
24. Fischer, W.; Ganusov, V.V.; Giorgi, E.E.; Hraber, P.T.; Keele, B.F.; Leitner, T.; Han, C.S.; Gleasner, C.D.; Green, L.; Lo, C.C.; et al. Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. *PLoS ONE* **2010**, *5*, e12303.
25. Ganusov, V.V.; Goonetilleke, N.; Liu, M.K.P.; Ferrari, G.; Shaw, G.M.; McMichael, A.J.; Borrow, P.; Korber, B.T.; Perelson, A.S. Fitness Costs and Diversity of the Cytotoxic T Lymphocyte (CTL) Response Determine the Rate of CTL Escape during Acute and Chronic Phases of HIV Infection. *J. Virol.* **2011**, *85*, 10518–10528, doi:10.1128/JVI.00655-11.
26. Kessinger, T.A.; Perelson, A.S.; Neher, R.A. Inferring HIV Escape Rates from Multi-Locus Genotype Data. *Front. Immunol.* **2013**, *4*, 252. doi:10.3389/fimmu.2013.00252.
27. Leviyang, S. Constructing lower-bounds for ctl escape rates in early siv infection. *J. Theor. Biol.* **2014**, *352*, 82–91, doi:10.1016/j.jtbi.2014.02.020.
28. Asquith, B.; McLean, A. In vivo CD8+ T cell control of immunodeficiency virus infection in humans and macaques. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 6365–6370.
29. Leviyang, S.; Ganusov, V.V. Broad CTL Response in Early HIV Infection Drives Multiple Concurrent CTL Escapes. *PLoS Comput. Biol.* **2015**, *11*, e1004492, doi:10.1371/journal.pcbi.1004492.

30. Garcia, V.; Feldman, M.W.; Regoes, R.R. Investigating the Consequences of Interference between Multiple CD8+ T Cell Escape Mutations in Early HIV Infection. *PLoS Comput. Biol.* **2016**, *12*, e1004721, doi:10.1371/journal.pcbi.1004721.
31. Ganusov, V.V.; Neher, R.A.; Perelson, A.S. Modeling HIV escape from cytotoxic T lymphocyte responses. *J. Stat. Mech.* **2013**, *2013*, P01010.
32. Efron, B.; Tibshirani, R. *An Introduction to the Bootstrap*; Chapman & Hall: New York, NY, USA, 1993.
33. Brown, L.; Cai, T.; DasGupta, A.; Agresti, A.; Coull, B.; Casella, G.; Corcoran, C.; Mehta, C.; Ghosh, M. Interval estimation for a binomial proportion. *Stat. Sci.* **2001**, *16*, 101–133.
34. Yang, Y.; Ganusov, V.V. Kinetics of HIV-specific CTL responses plays a minimal role in determining HIV escape dynamics. *Front. Immunol.* **2018**, *9*, 1–15, doi:10.3389/fimmu.2018.00140.
35. Geldmacher, C.; Currier, J.R.; Herrmann, E.; Haule, A.; Kuta, E.; McCutchan, F.; Njovu, L.; Geis, S.; Hoffmann, O.; Maboko, L.; et al. CD8 T-cell recognition of multiple epitopes within specific Gag regions is associated with maintenance of a low steady-state viremia in human immunodeficiency virus type 1-seropositive patients. *J. Virol.* **2007**, *81*, 2440–2448, doi:10.1128/JVI.01847-06.
36. Streeck, H.; Jolin, J.S.; Qi, Y.; Yassine-Diab, B.; Johnson, R.C.; Kwon, D.S.; Addo, M.M.; Brumme, C.; Routy, J.-P.; Little, S.; et al. Human immunodeficiency virus type 1-specific cd8+ t-cell responses during primary infection are major determinants of the viral set point and loss of cd4+ t cells. *J. Virol.* **2009**, *83*, 7641–7648, doi:10.1128/JVI.00182-09.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).