

BLIND UNEVEN PROLIFERATION OF CD4+ T CELLS DURING PRIMARY INFECTION GENERATES THE MAJORITY OF THE HIV RESERVOIR

FLORENCIA A. T. BOSHIER, DANIEL B. REEVES, ELIZABETH R. DUKE, DAVID A. SWAN,
MARTIN PRLIC, E. FABIAN CARDOZO-OJEDA AND JOSHUA T. SCHIFFER

1. MATHEMATICAL MODEL

1.1. **Competing Models.** We model the viral and T cell kinetics of primary infection by a system of non-linear differential equations. The full model, from which all competing models are derived, is shown in equation (1).

$$\begin{aligned}
 (1) \quad \frac{dE}{dt} &= r_E \left(1 - \frac{N + S + A + L + E}{K} \right) E - d_E E \\
 &\quad + w_1 h(A, E) \\
 \frac{dN}{dt} &= r_N \left(1 - \frac{N + S + A + L + E}{K} \right) N - d_N N - \alpha V N + \theta \\
 \frac{dS}{dt} &= r_S \left(1 - \frac{N + S + A + L + E}{K} \right) S - d_S S \\
 &\quad - \beta V S + g \alpha V N \\
 \frac{dL}{dt} &= r_L \left(1 - \frac{N + S + A + L + E}{K} \right) L - d_L L + f \beta V S - a L \\
 \frac{dA}{dt} &= (1 - f) \beta V S - d_A A - w_2 h(A, E) + a L \\
 \frac{dV}{dt} &= p A - c V
 \end{aligned}$$

with

$$(2) \quad h(A, E) = AE \quad \text{or} \quad \frac{E}{E + E_{50}} A.$$

Here $E(t)$ represents the concentration CD8+ T cells at time t . CD4+ T cells are represented by four compartments: non-susceptible, susceptible, latently infected and actively infected. Their concentrations at time t are denoted by $N(t)$, $S(t)$, $L(t)$ and $A(t)$ respectively. Finally, $V(t)$ represents the plasma viral load over time.

The model makes the following assumptions. The virus infects susceptible cells at a rate β . A fraction f of infections results in actively infected cells; and the remaining smaller fraction $(1 - f)$ are latently infected. The virus is cleared at a rate c . Actively infected cells produce virus at a rate p and die at a rate d_A . They are also killed by CD8 T cells at a rate w_2 , as described by the function $h(A, E)$. We consider two forms, either linear or concentration-dependent as given by equation (2). Latently infected cells activate at a rate a . These cells grow logistically at a maximum rate r_L and die at a rate d_L . Non-susceptible cells grow logistically with a maximum division rate of r_N and die at rate d_N . These cells can also undergo bystander killing due to active infection at a rate $(1 - g)\alpha$ and upregulate CCR5, becoming susceptible, at a rate $(g)\alpha$. CD8 cells also grow logistically with a maximum division rate of r_N and die at rate d_N . In one version of the model, they also undergo antigen stimulation at a rate w_1 as described by function $h(A, E)$. CD4+ and CD8+ T cells compete for resources. We modelled this assumption with a logistic equation that depends on the ratio of total CD4+ and CD8+T cells, $(N + S + L + A + E)$, to a carrying capacity K . Although infected cells contribute to total T cell counts, we assume that they do not proliferate.

We rescaled the model using the following substitutions $\hat{r}_E = (1 - d_E/r_E)$, $\hat{r}_N = (1 - d_N/r_N)$ and $\hat{r}_S = (1 - d_S/r_S)$. We assume that CD8+ and CD4+ T cells are in stable co-existence pre-infection. Hence we let $d_E/r_E = d_N/r_N = d_S/r_S = D$ and introduce a new parameter $\hat{K} = (1 - D)K$. We make the further assumption that the latent cells proliferate at the same rate as non-susceptible cells such that $r_L = r_N$ and $d_L = d_N$ hence $\hat{r}_L = \hat{r}_N$. This last assumption will be relaxed in our sensitivity analysis. Using these definitions, the full model takes the form:

$$\begin{aligned}
(3) \quad \frac{dE}{dt} &= \hat{r}_E \left(1 - \frac{N + S + A + L + E}{\hat{K}} \right) E + w_1 h(A, E) \\
\frac{dN}{dt} &= \hat{r}_N \left(1 - \frac{N + S + A + L + E}{\hat{K}} \right) N - \alpha V N + \theta \\
\frac{dS}{dt} &= \hat{r}_S \left(1 - \frac{N + S + A + L + E}{\hat{K}} \right) S \\
&\quad - \beta V S + g \alpha V N \\
\frac{dL}{dt} &= \hat{r}_L \left(1 - \frac{N + S + A + L + E}{\hat{K}} \right) L + f \beta V S \\
\frac{dA}{dt} &= (1 - f) \beta V S - d_A A - w_2 h(A, E) \\
\frac{dV}{dt} &= p A - c V
\end{aligned}$$

We constructed several competing models to explore which of the following mechanistic assumptions of the full model are required to achieve most parsimonious fit to the data:

- (1) Active infection does not directly effects non-susceptible cells ($\alpha = 0$).
- (2) CD8+ and CD4+ T cells do not compete for resources ($E(0) = 0$).
- (3) CD8+ T cells are not stimulated by infection ($w_1 = w_2 = 0$).
- (4) CD8+ T cells are stimulated by infection but do not have additional cytolytic effect ($w_2 = 0$).
- (5) CD8+ T cells are stimulated by infection and have additional CD8+ T cytolytic effect on actively infected cells.
- (6) Thymic source of non-susceptible cells $\theta > 0$.

We considered these individually and in combination. Overall, we constructed 12 competing models, the details of which are shown in table 1.

1.2. Model Fitting. To fit our system of non-linear ODE's we adopted a statistical model for the observed state variable vector v and the parameter set $\Psi = \{t_0, \hat{r}_E, \hat{r}_N, \hat{r}_S, \alpha, g, \beta\}$.

We used a non-linear mixed effects model for the state variable vector v , such that for individual j at observation time t_{ij} :

$$(4) \quad v_{ij} = f_v(t_{ij}, \Psi_j) + \varepsilon_v.$$

Where f_v is the numerical solution of the system of non-linear ODEs under consideration (presented in table 2) at time t_{ij} with parameters Ψ_j . The variable ε_v represents the measurement noise which we assume to be normally distributed with zero mean and state-variable dependent variance σ_v^2 (i.e. $\varepsilon_v \sim \mathcal{N}(0, \sigma_v^2)$).

For each estimated parameter we also used a mixed-effects model. We assumed that for each individual j , each parameter $\psi_j \in \Psi_j$ is drawn from a probability distribution across the population. The distribution includes fixed effects $\bar{\psi}$ representing the mode over the population, and the random effects η_j representing its variability in the population which we assume to be normally distributed with zero mean and parameter-dependent variance σ_ψ^2 (i.e. $\eta_j \sim \mathcal{N}(0, \sigma_\psi^2)$).

For each of the system of ODE's considered, we fit each statistical model to all data points from all individuals simultaneously using a maximum likelihood approach. We assumed individual observations of each state variable v_{ij} for each individual j at time point t_{ij} are independent. We assumed the following functional forms for the parameters:

- $\hat{r}_*, \beta, \alpha, g, t_0$ were modelled as $\psi_j = \bar{\psi} e^{\eta_j}$
- K was modelled as $10^{\psi_j} = 10^{\bar{\psi} + \eta_j}$.

We fixed the standard deviation of the measurement errors for the observations σ_v , and estimated each parameter's fixed effects $\bar{\psi}$ and standard deviation of the random effects σ_ψ using the Stochastic Approximation of the Expectation Maximization (SAEM) algorithm embedded in the Monolix software [?].

1.3. Model Selection. For the best fit of each model, we computed the log-likelihood ($\log \mathcal{L}$) and the Akaike Information Criteria (AIC):

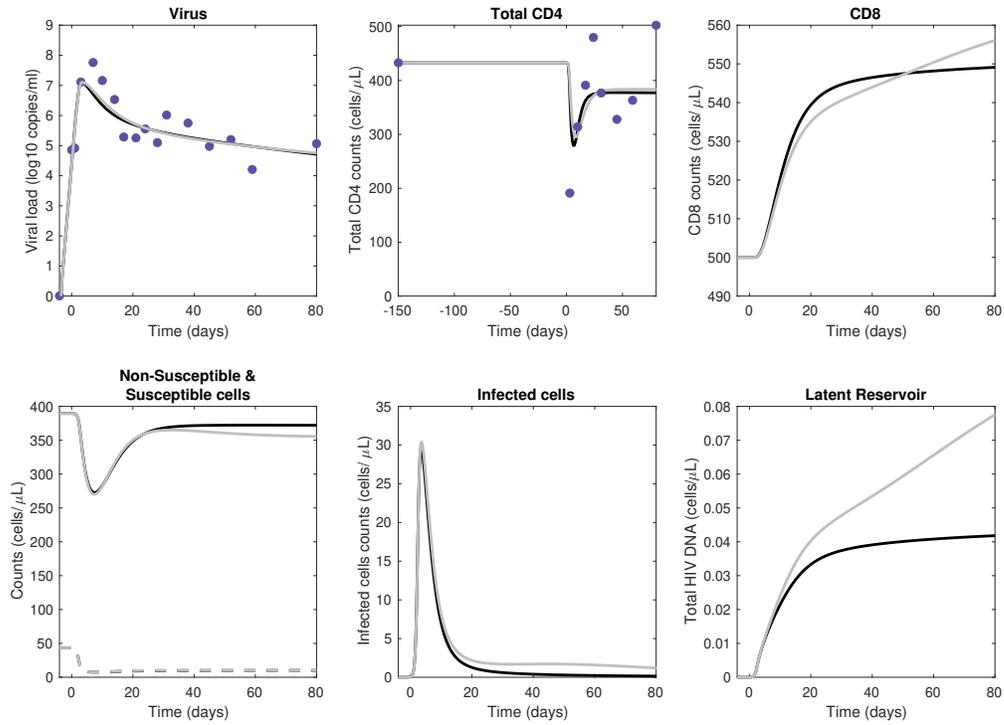
$$(5) \quad AIC = -2 \log \mathcal{L} + 2l,$$

where l is the number of parameters estimated. We assume models are equivalently supported by the data if the difference between their AIC values is less than two. The corresponding AIC scores for the 12 models considered is shown in table 2. We see that the most parsimonious models contain blind homeostatic proliferation of CD4 and CD8 T cells as well as bystander killing and up-regulation of CCR5 but do not require the assumptions of infection-induced proliferation of CD8+ T cells or additional CD8+ T cell lytic effect on actively infected cells. A diagram of the final model is shown in figure 1.

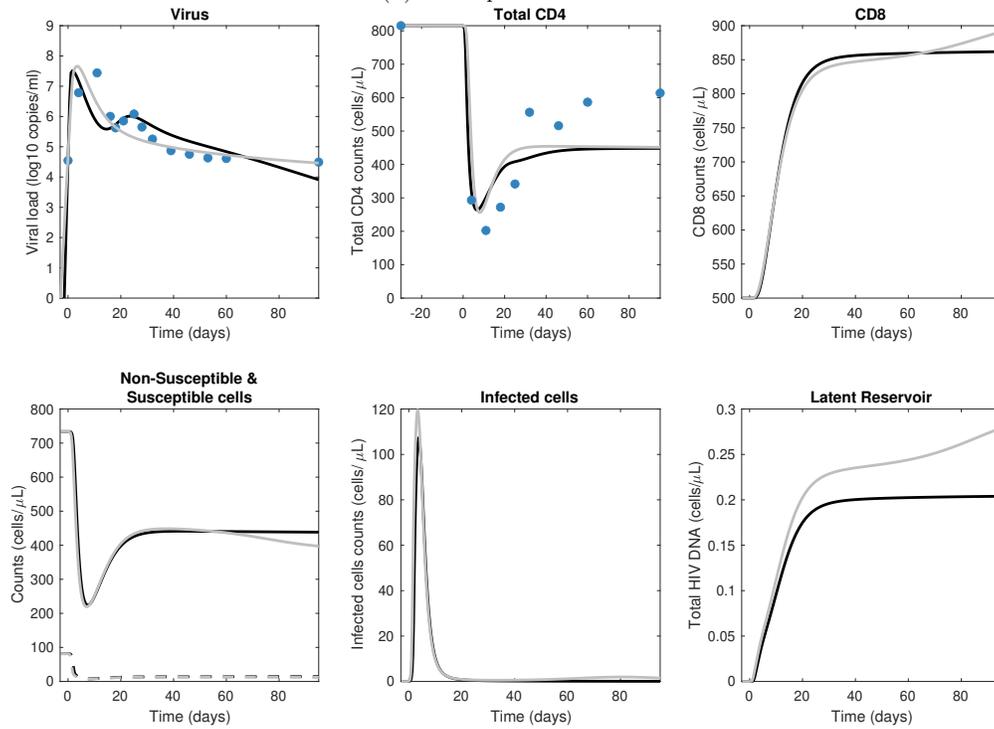
TABLE 1. Summary of AIC scores for all tested models. The most parsimonious model by AIC is highlighted in grey.

Model Number	Removed Mechanisms	AIC
1	$E(0) = w_1 = w_2 = \alpha = 0$	2234.54
2	$E(0) = 0,$	1026.05
3	$\alpha = w_1 = w_2 = 0$	2054.30
4	$w_1 = w_2 = 0$	653.05
5	$w_2 = 0, h$ linear	657.31
6	h linear	829.47
7	$w_2 = 0, h$ concentration dependent	657.33
8	h concentration dependent	677.49
9	$\alpha = 0, h$ linear	2042.10
10	$[\alpha = 0, h$ hill-function	2068.32
11	$\alpha = w_1 = w_2 = 0$ thymus term, $\theta > 0$	657.68
12	$E(0) = w_1 = w_2 = \alpha = 0$ thymus term, $\theta > 0$	2230.68

1.4. Individual fits. In figure 1, we show the results of deterministic models 4 and 5 for each of the 12 participants from the Females Rising through Education, Support and Health cohort (FRESH) [?, ?]. Model 4 assumes that CD8+ T cells expand only due to lymphopaenia in a blind fashion. Model 5 assumes both blind homeostatic proliferation as well as antigen-driven proliferation. For CD4+ T cells and viral load, we superimpose the digitally extracted empirical data with a colour code that matches figure 3 in the main text. For measurements of cellular concentrations, we scaled simulation results, which are run in a liter of blood, to a μL . For viral concentrations, we scale to a mL . Simulation results from our model closely recapitulate longitudinal measures of viral load and total CD4+ counts, which is the sum of $N(t)$, $S(t)$, $L(t)$ and $I(t)$. The behaviour of the CD8+ cell trajectories falls within the range observed during primary infection in [?].

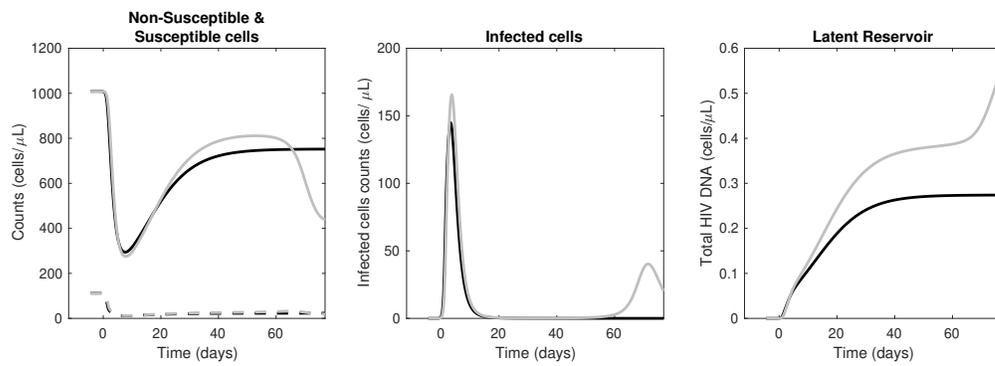
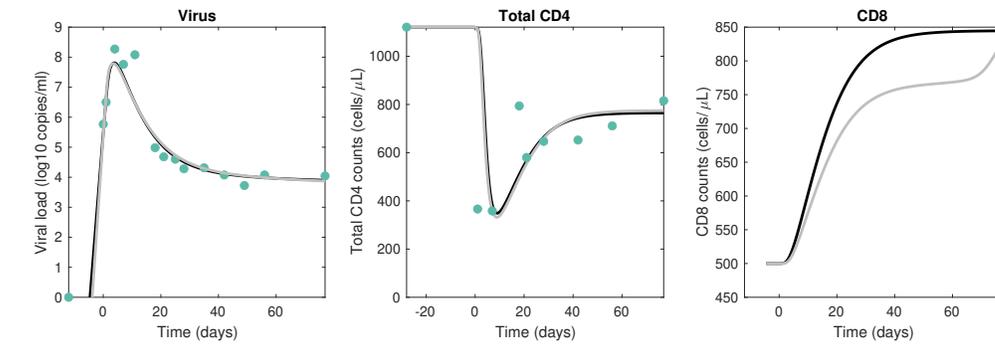


(A) Participant 1

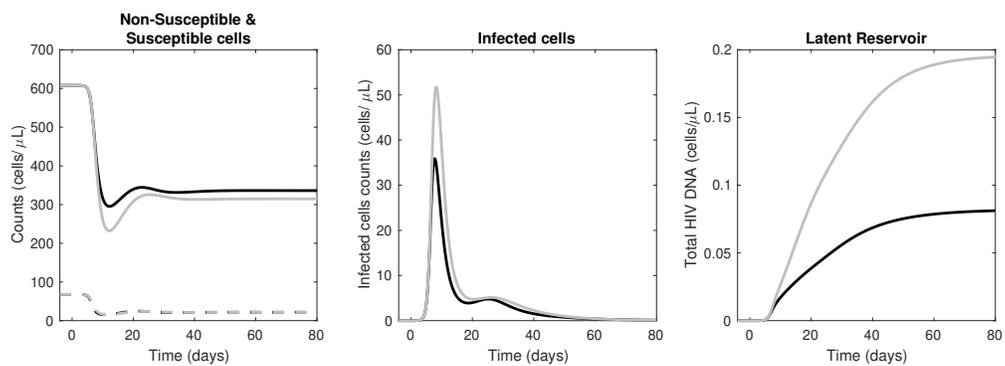
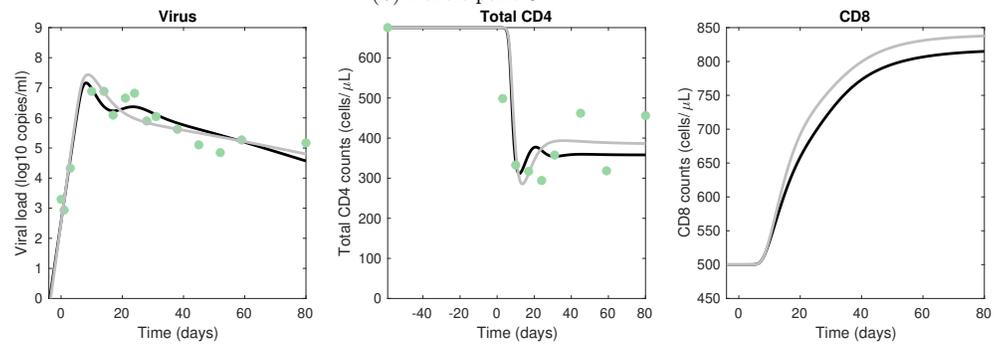


(B) Participant 2

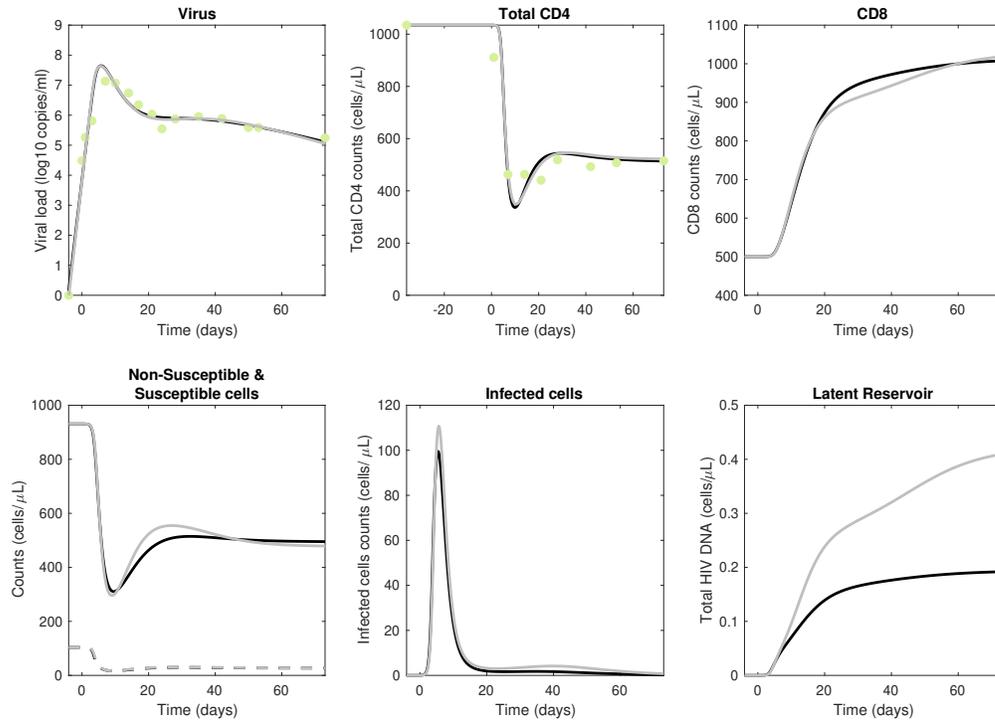
BLIND UNEVEN PROLIFERATION OF CD4+ T CELLS DURING PRIMARY INFECTION GENERATES THE MAJORITY OF THE HIV RESERVOIR



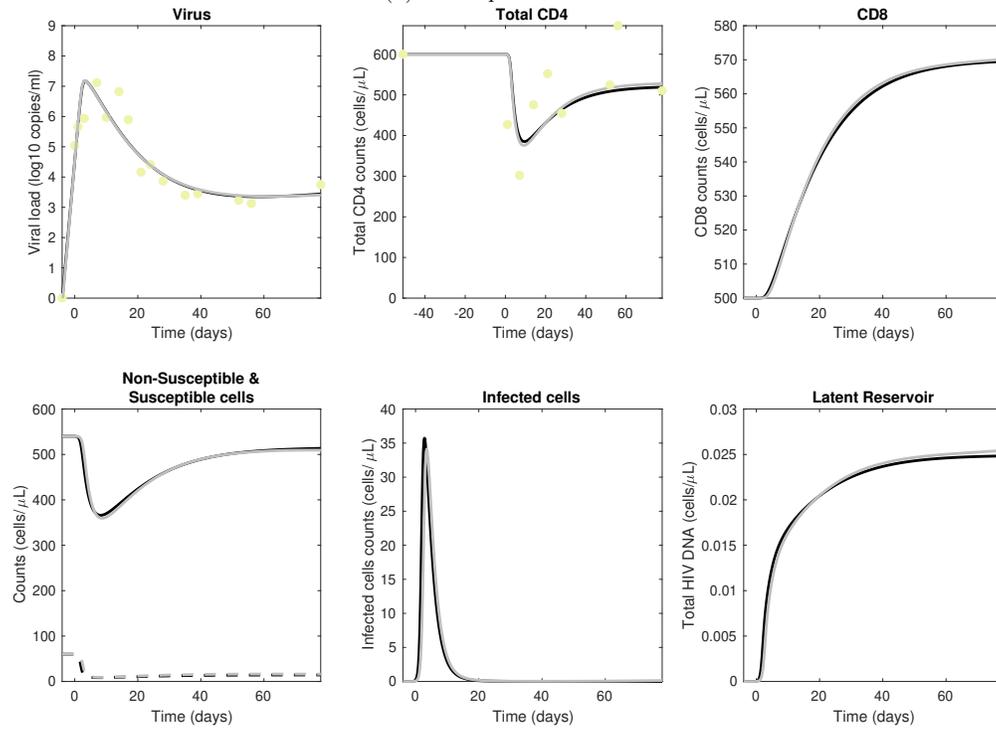
(c) Participant 3



(d) Participant 4

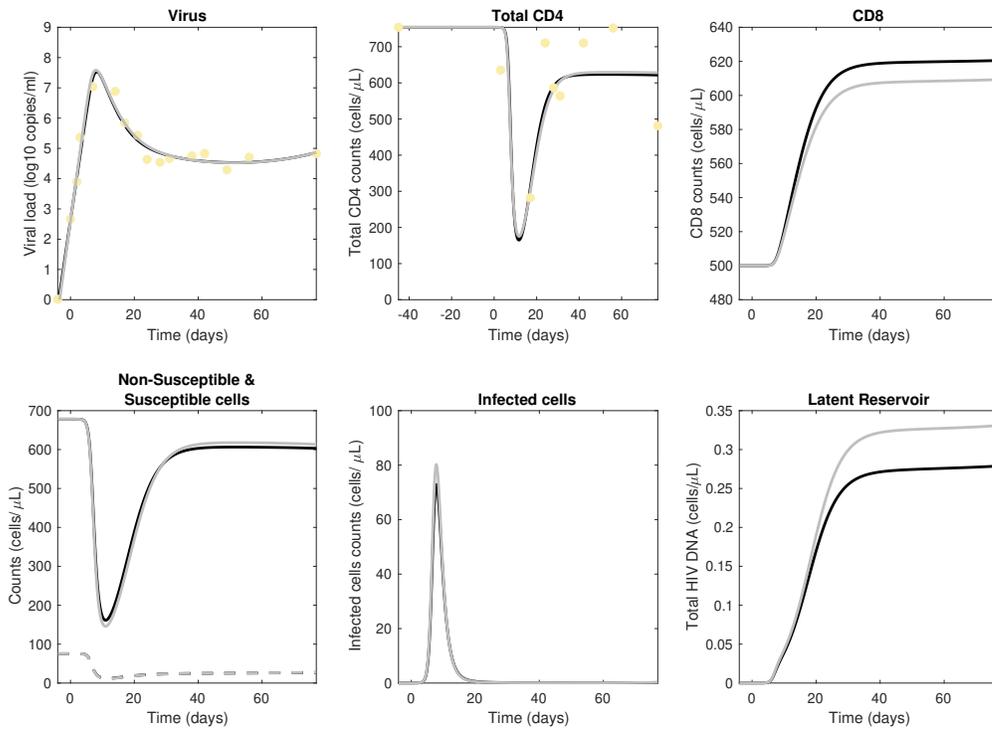


(E) Participant 5

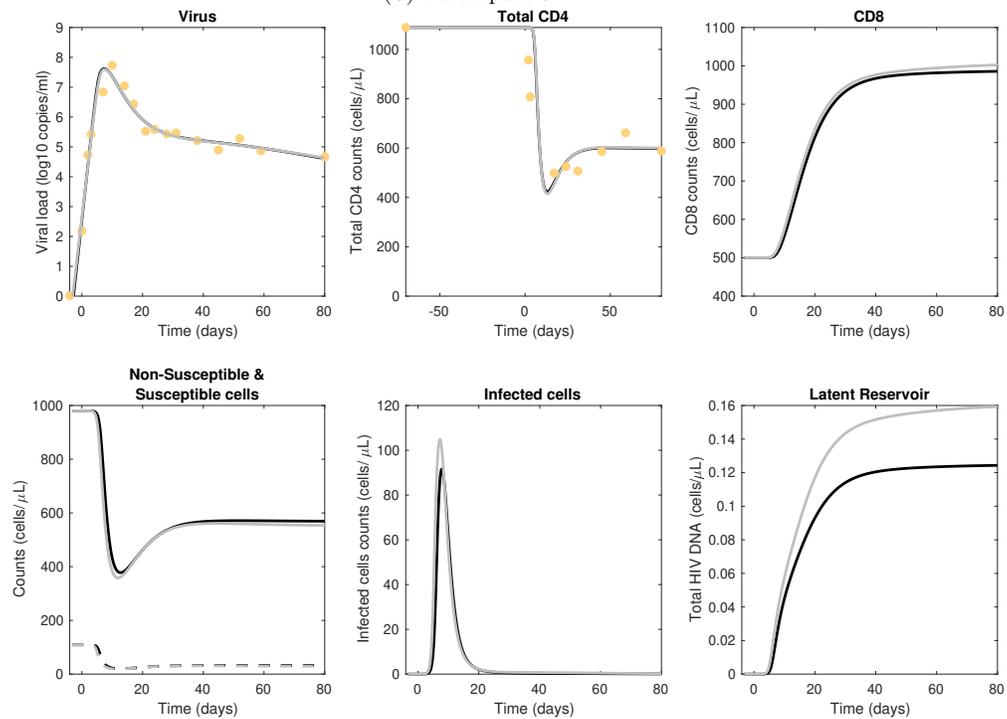


(F) Participant 6

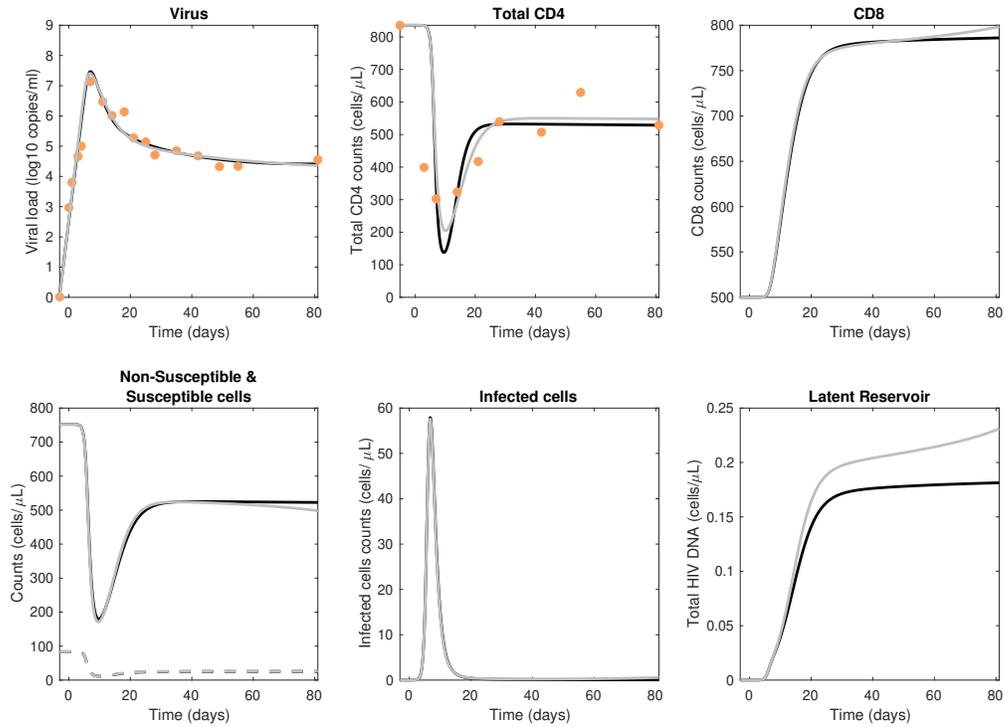
BLIND UNEVEN PROLIFERATION OF CD4+ T CELLS DURING PRIMARY INFECTION GENERATES THE MAJORITY OF THE HIV RESERVOIR



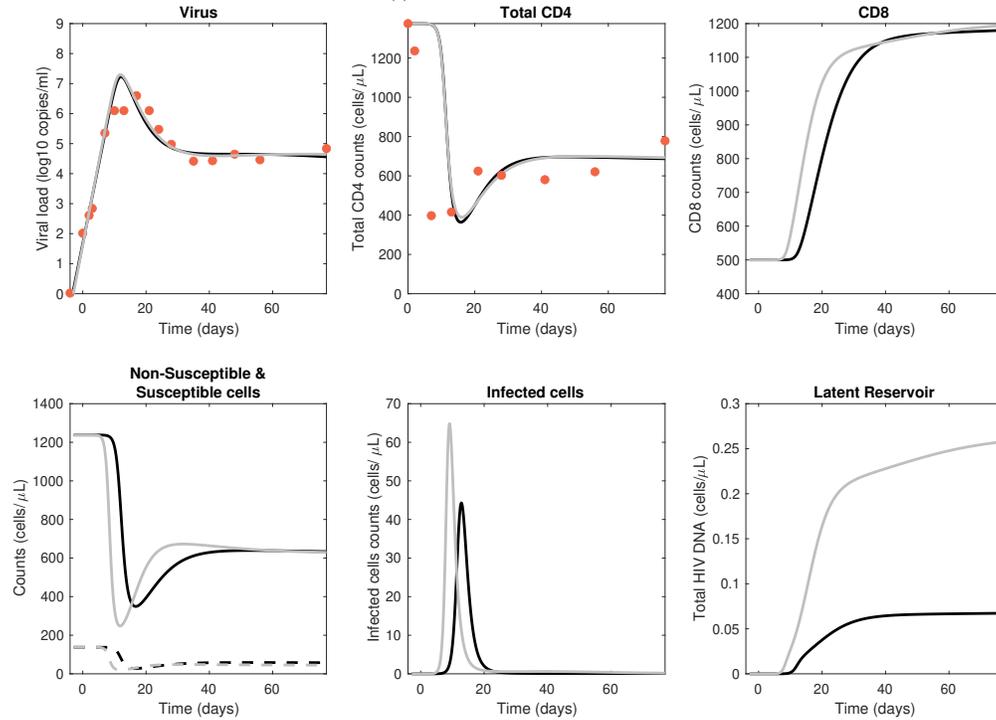
(G) Participant 7



(H) Participant 8



(i) Participant 9



(j) Participant 10

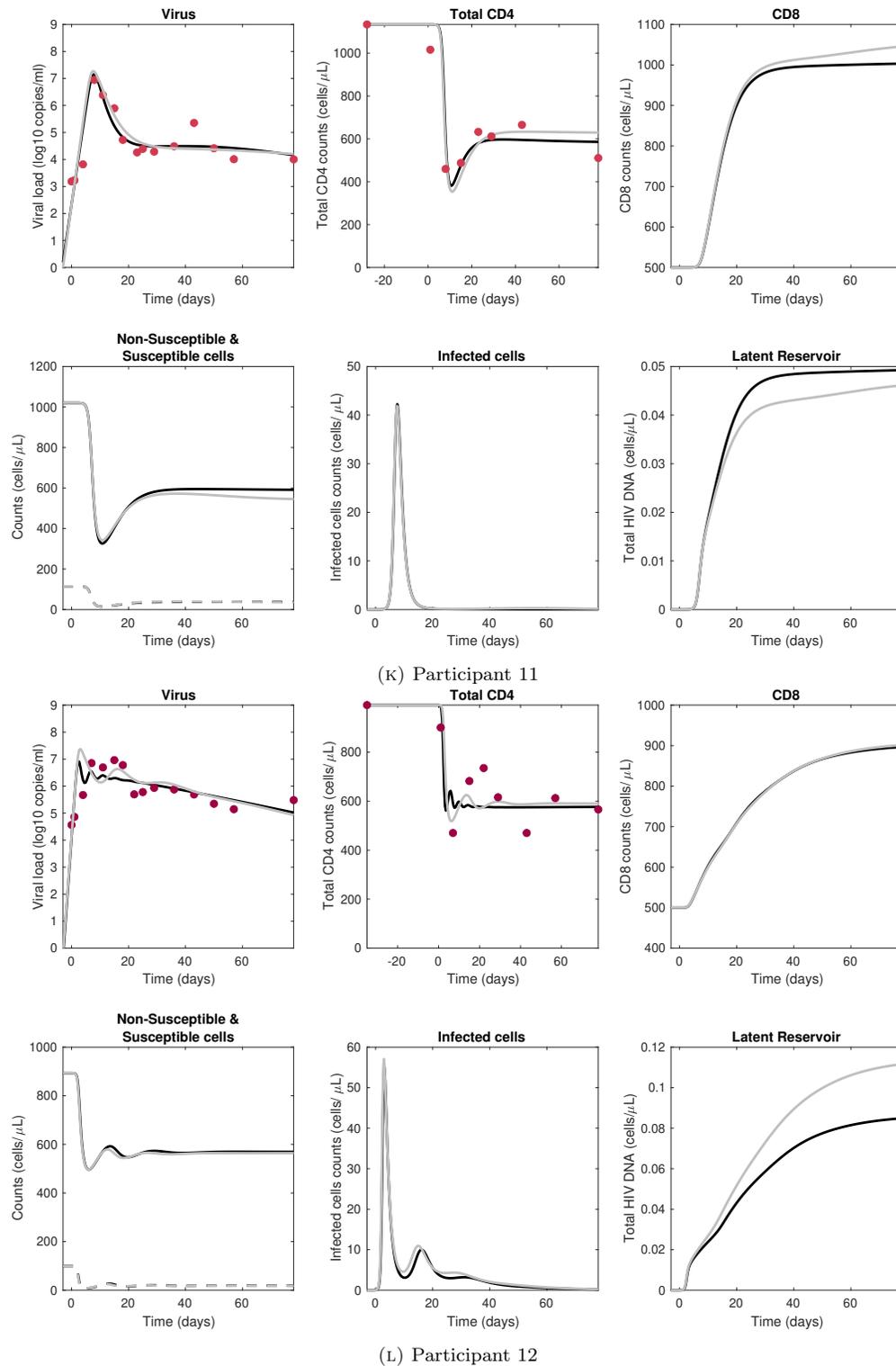


FIGURE 1. Simulations of the primary infection models for all participants table 2, superimposed onto the digitized data where available. Colour correspond to participants as in figure 3 of the main text. Black lines correspond to our chosen model 4 and grey lines correspond to model 5. Dashed lines represent susceptible cells.

TABLE 2. Individual Parameter Values.

Participant	t_0	r_N	r_S	r_E	g	α	β
Participant 1	-4.02E+00	4.63E-01	2.26E-01	4.77E-02	5.54E-01	1.08E-05	-3.65E+00
Participant 2	-3.12E+00	3.59E-01	2.29E-01	1.29E-01	5.89E-01	8.90E-06	-3.80E+00
Participant 3	-4.20E+00	1.76E-01	1.12E-01	6.36E-02	6.52E-01	5.76E-06	-4.01E+00
Participant 4	-4.01E+00	2.91E-01	8.07E-01	1.05E-01	2.36E-01	1.52E-05	-4.01E+00
Participant 5	-4.40E+00	2.80E-01	3.60E-01	1.38E-01	4.49E-01	8.21E-06	-4.10E+00
Participant 6	-4.17E+00	1.12E-01	1.79E-01	2.94E-02	4.33E-01	8.06E-06	-3.79E+00
Participant 7	-3.98E+00	4.12E-01	1.76E-01	3.82E-02	3.38E-01	1.73E-05	-4.04E+00
Participant 8	-3.56E+00	1.92E-01	2.43E-01	1.33E-01	4.82E-01	6.66E-06	-4.16E+00
Participant 9	-3.55E+00	5.05E-01	3.50E-01	1.01E-01	1.65E-01	2.57E-05	-4.05E+00
Participant 10	-3.77E+00	2.09E-01	2.58E-01	1.41E-01	5.64E-02	2.09E-05	-4.40E+00
Participant 11	-3.60E+00	2.73E-01	4.35E-01	1.54E-01	3.64E-02	2.56E-05	-4.22E+00
Participant 12	-3.22E+00	2.37E-01	1.39E+00	1.02E-01	8.73E-02	1.32E-05	-3.94E+00

2. SENSITIVITY ANALYSIS

We performed a sensitivity analysis to examine the effects of the mean and variance of the turnover rate of latently infected reservoir cells, and fraction of latency (m , v , and f , respectively) on the size of the latent reservoir throughout infection and its clonal structure at 50 days post-infection.

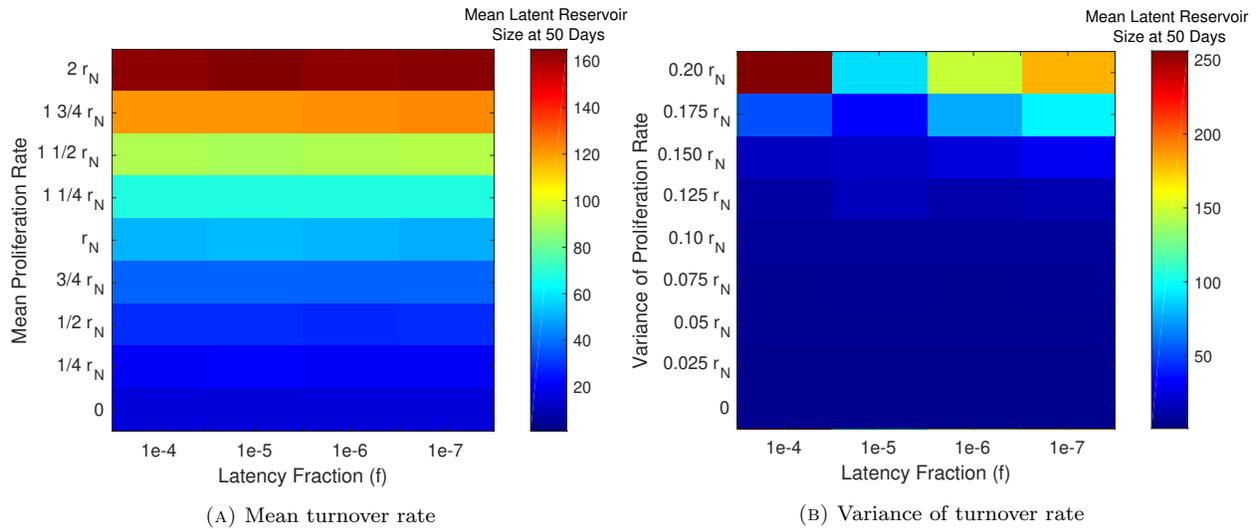


FIGURE 2. Heatmap showing total reservoir size at 50 days post infection for Participant 11 a) according to the latency fraction f and the mean turnover rate μ of latently infected cells L . b) according to the latency fraction f and the turnover rate variance σ of latently infected cells.

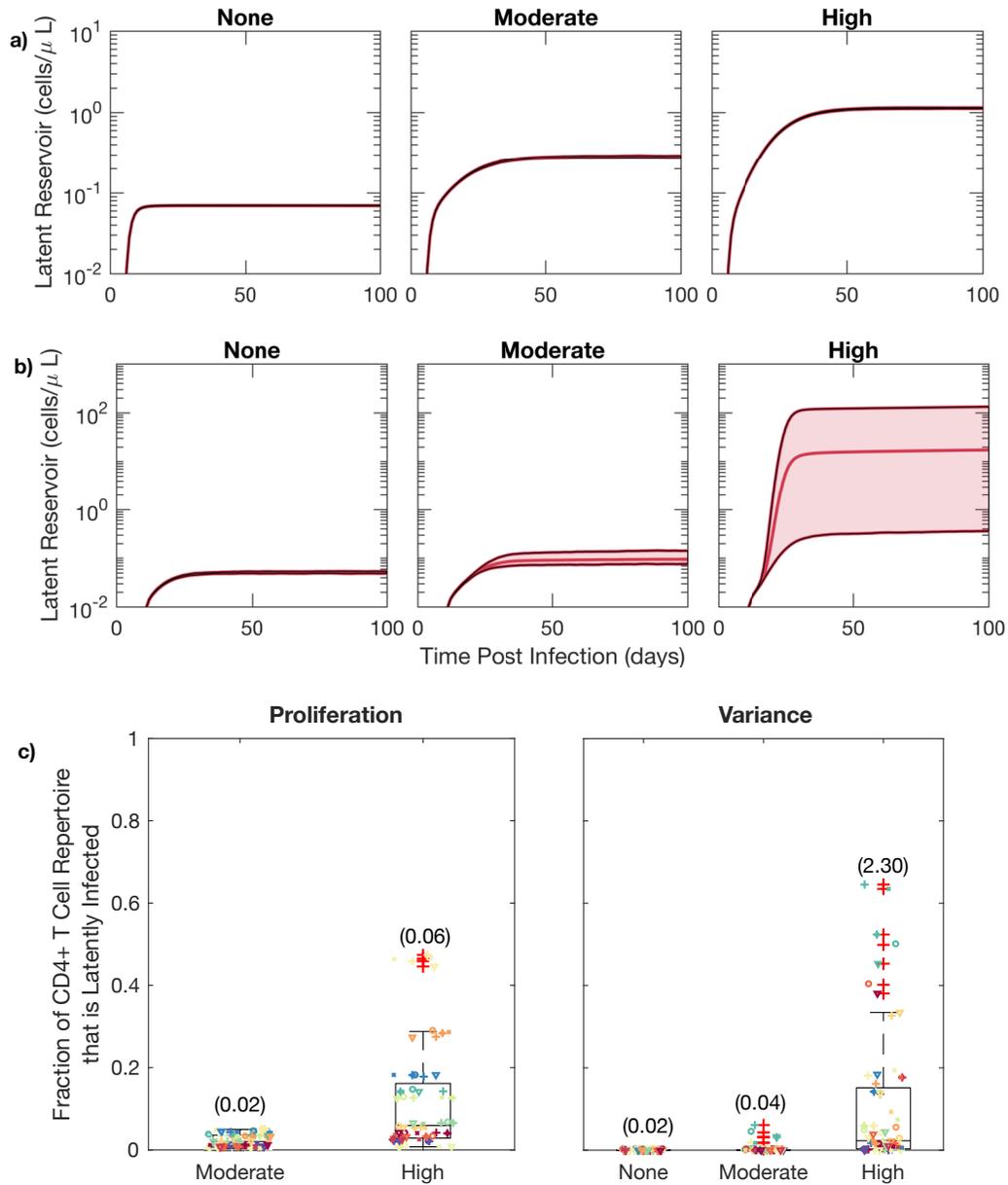


FIGURE 3. The mean and variance of lymphocyte proliferation rates impact reservoir size. Latent reservoir size over time post-infection a) no ($m = 0$), moderate ($m = 0.5r_N$) and high ($m = 1.5r_N$) mean proliferation rates with moderate variance ($v = 0.075r_N$) b) no ($v = 0$), moderate ($v = 0.075r_N$) and high variance ($v = 0.175r_N$) of proliferation rates c) Fraction of the CD4+ T cell repertoire that is latently infected.

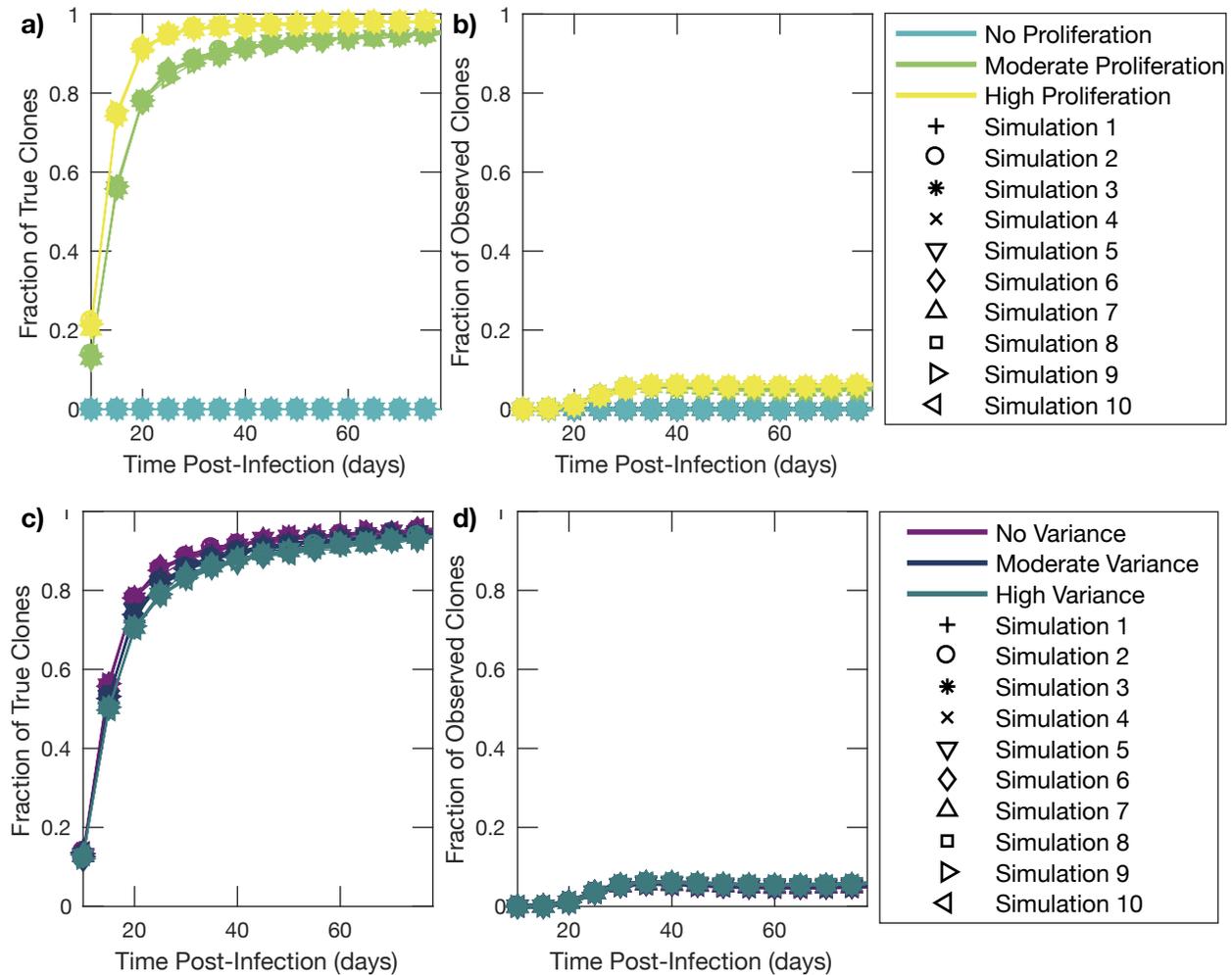


FIGURE 4. The mean and variance of lymphocyte proliferation rates only modestly impact proportion of clones in the reservoir. Fraction of true clones from 10 stochastic simulations of one individual. a-b) The range of mean proliferation rates does not effect the fraction of true clones in the reservoir. a) Fraction of true clones in the latent reservoir, b) Fraction of observed true clones in presence of active infection for an individual considered. (no proliferation, $m = 0$, moderate $m = r_N$ and high $m = 2r_N$.) c-d) The range of variance of these does not effect the fraction of true clones in the reservoir. c) Fraction of true clones in the latent reservoir, d) Fraction of observed true clones in presence of active infection for an individual considered. (no variance, $v = 0$, moderate $v = 0.075r_N$ and high $v = 0.175r_N$).

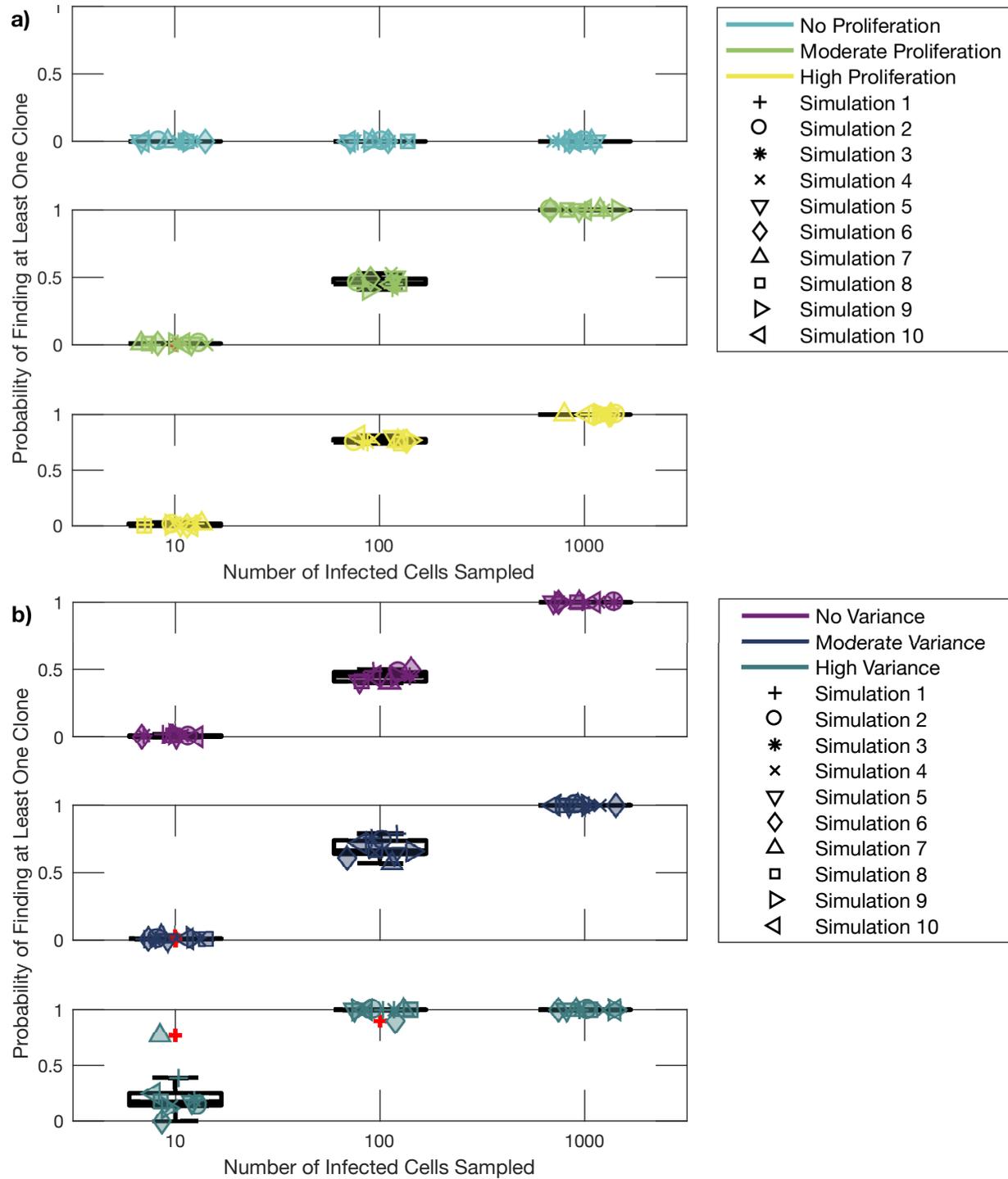


FIGURE 5. Probability of clone detection is influenced by sequencing depth as well as the mean and variance of lymphocyte proliferation rates. Probability of finding at least one clone from in silico sampling of infected cells at 50 days post infection of 10 stochastic simulations for one individual. a) Each panel represents three scenarios: no proliferation, $(m, v) = (0, 0)$, moderate $(m, v) = (r_N, 0)$ and high $(m, v) = (2r_N, 0)$. b) Each panel represents three scenarios: no variance $((m, v) = (r_N, 0))$, second has moderate turnover rates $((m, v) = (r_N, 0.075r_N))$, third has high turnover rates $((m, v) = (r_N, 0.175r_N))$.

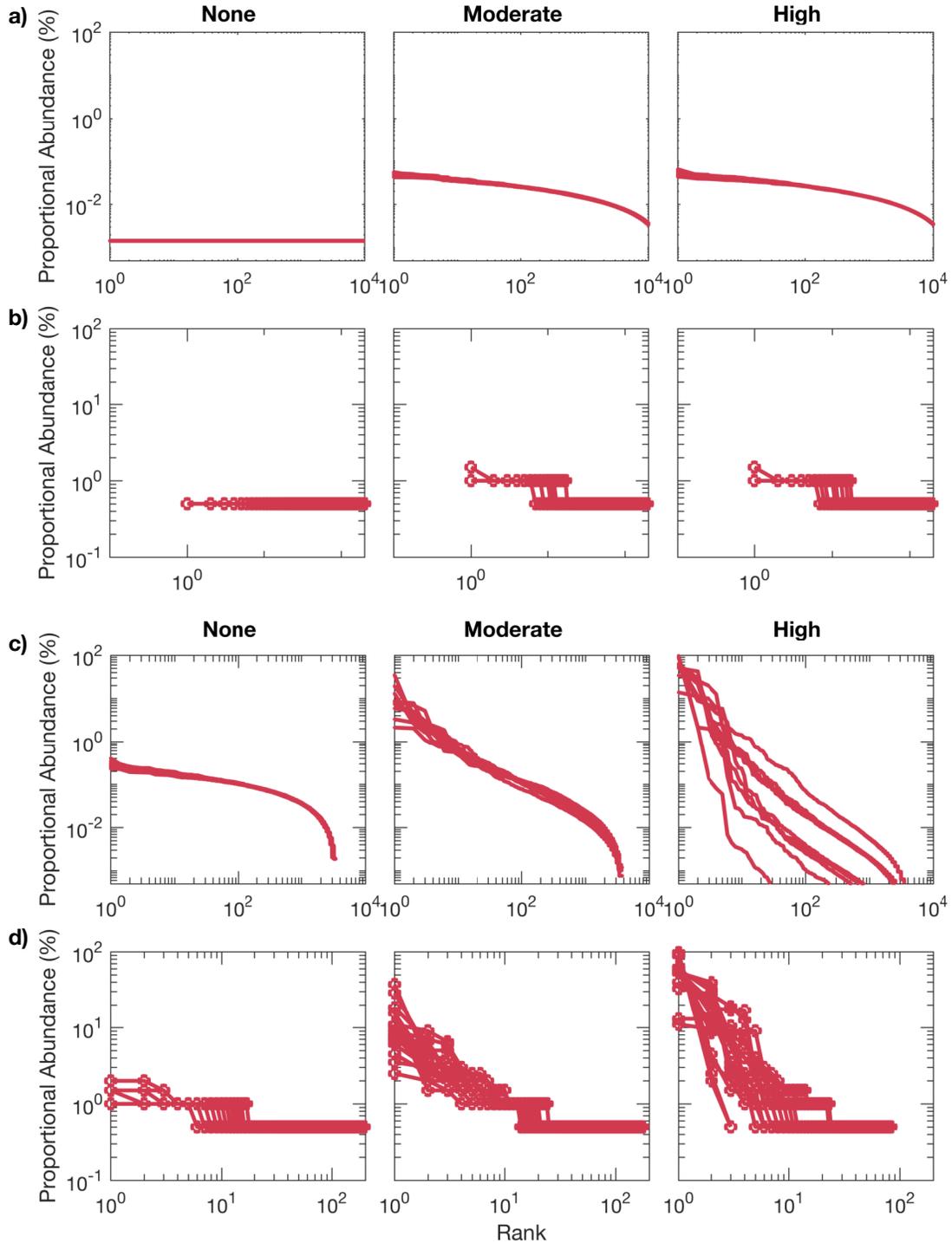


FIGURE 6. The mean and variance of lymphocyte proliferation rates impact reservoir clonal structure. a-b) Results from 10 stochastic simulations with varying turnover rates and no variance $v = 0$. a) Population rank-abundance curves at 50 days post infection. b) Rank-abundance curve of a sample of 200 latent reservoir cells at 50 days post infection. Each column represents one of three scenarios: the first has no proliferation $((m, v) = (0, 0))$, second has moderate turnover rates $((m, v) = (r_N, 0))$, third has high turnover rates $((m, v) = (2r_N, 0))$. With no proliferation the reservoir is composed entirely of singlets. Increased turnover results in the larger clones. c-d) Results from 10 stochastic simulations with varying variance of turnover rates, but fixed mean turnover rate $m = r_N$. c) Population rank-abundance curves at 50 days post infection. c) Rank-abundance curve of a sample of 200 latent reservoir cells at 50 days post-infection. Each column