Mathematical Biosciences and Engineering

https://www.aimspress.com/journal/mbe

MBE, 23(5): 1437–1460.
DOI: 10.3934/mbe.2026053
Received: 03 February 2026
Revised: 02 April 2026
Accepted: 08 April 2026
Published: 13 April 2026

Research article

A close look at the viral reduction rate in target cell limited models

Lifeng Han[1,*], Jiang Liu[2], Ye Yuan[2], Shiqian Zhang[1], Hao Zhu[2] and Ye Xiong[2,*]

[1] Department of Mathematics, Tulane University, 6823 St. Charles Avenue, New Orleans 70115, USA
[2] Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring 20993, USA

* **Correspondence:** Email: lhan1@tulane.edu, Ye.Xiong@fda.hhs.gov.

Abstract: The target cell limited (TCL) model is a cornerstone of within-host viral dynamics and is widely used to interpret viral load (VL) data and assess antiviral efficacy. We examine how TCL models predict VL reduction under drug treatments, thereby focusing on parameters that govern infection and viral production. Our simulations and analyses reveal a non-monotonic relationship: moderate drug effects can slow VL decline and prolong persistence above the limit of quantification. We provide a rigorous analytical characterization of the conditions under which this non-monotonic behavior arises. We derive an explicit formula for the critical parameter value that delineates monotonicity and show that this phenomenon broadly arises in generalized TCL models. Additionally, we propose a modified formulation with a sigmoid infection term, which reduces the extent of non-monotonicity while preserving the analytical tractability. These findings contribute to a deeper understanding of the TCL model behavior and provide a practical alternative formulation when a monotonic dose-response relationship is desired.

Keywords: target cell limited model; viral dynamics; antiviral efficacy; viral load reduction; non-monotonic drug response

1. Introduction

Understanding within-host viral dynamics is essential to elucidate the disease pathogenesis and design effective therapies. A time series of the viral load (VL), which is typically measured by quantitative polymerase chain reaction (PCR), in nasopharyngeal swabs or blood samples, provides valuable information on the replication kinetics of viruses such as SARS-CoV-2 [1]. Mathematical models are often employed to interpret these data, estimate key parameters, and guide clinical decision-making [2, 3].

Antiviral therapies, including monoclonal antibodies and small-molecule drugs such as nirmatrelvir/ritonavir, are now central components of the therapeutic arsenal against COVID-19. In clinical trials, VL reduction may be predictive of treatment efficacy [4]. The decline in log-transformed VL curves often

appears linear (see, for example, Figure 2 of [5]), which can be interpreted as the asymptotic decay rate of the virus, thus providing a link between the modeling analysis and the observed treatment outcomes.

Among mathematical models, the target cell limited (TCL) model is a widely used framework to describe viral dynamics after acute infection [6]. By capturing interactions among susceptible target cells, infected cells, and free virus, TCL models reproduce the typical rise and fall of VL observed in patients. Drug effects are commonly incorporated into these models by modifying one or more parameters in accordance with the drug's mechanism of action [7–9]. For example, neutralizing antibodies may reduce the rate of viral entry into target cells [10], while polymerase inhibitors may decrease viral production [11, 12]. Throughout this paper, we refer to drug effects in the context of the TCL model as a fold change in the relevant parameter, even though the drug dose is not explicitly modeled.

Despite their utility, a careful examination of the TCL model reveals an important feature: in some cases, stronger drug effects do not correspond to a faster VL reduction and the model predicts prolonged high levels of VL unless drug strength is raised beyond a certain threshold. This non-monotonic relationship complicates the interpretation of VL decline as a measure of treatment potency. The observation that there is a non-monotonic relationship between drug effectiveness and time to clear a virus was made previously by Best et al. [13] using a TCL model with an eclipse phase for Zika virus dynamics. Although prolonged viral shedding under intermediately strong drug effects predicted by the TCL model has received some attention [7, 14], the conditions under which this non-monotonic behavior arises have not been rigorously characterized. Some experimental evidence suggestive of non-monotonic dose-response relationships has been reported in nonhuman primate studies of favipiravir [15, 16], which we further discuss in Section 4. Clinically, flat exposure-response relationships for VL have been reported for monoclonal antibody therapies [17, 18], thereby suggesting that the drug effect might be saturated. When extrapolating to moderate drug doses, one often expects a monotonic relationship. In this study, we take a closer look at the viral reduction rate in TCL models, and determine the critical parameter values that give rise to the non-monotonic behavior. By generalizing the basic TCL models, we find a simple modeling choice that mitigates the non-monotonic behavior. Our goal is to characterize the conditions under which the non-monotonic behavior arises and provide an alternative model formulation when a monotonic dose-response is desired.

The remainder of this paper is organized as follows: in Section 2, we review the TCL model and present extensive simulations to identify the non-monotonic relationship between VL reduction and drug effects; Section 3 presents our analytical results that characterizes the relationship between drug effects and VL decline; and Section 4 discusses the implications and significance of our results.

2. Model description and simulation

2.1. TCL model

The TCL model [19] considered in this paper accounts for the target cell density, infected cell density, and free virus density at time t, denoted by $T(t)$, $I(t)$, and $V(t)$, respectively. Their dynamics are described by the following ordinary differential equations:

$$\frac{dT}{dt} = -\beta VT, \tag{2.1a}$$

$$\frac{dI}{dt} = \beta VT - \delta I, \tag{2.1b}$$

$$\frac{dV}{dt} = pI - cV, \tag{2.1c}$$

where β [mL RNA copies^{-1} day^{-1}] is the infection rate, p [RNA copies mL^{-1} day^{-1}] is the production rate of free virus per infected cell, and δ [day^{-1}] and c [day^{-1}] are the clearance rates for infected cells and free virus, respectively. Equation (2.1a) states that target cells are depleted through infection at a rate proportional to both virus and target cell densities (mass-action kinetics). Equation (2.1b) describes the dynamics of infected cells, which are produced by infection and cleared at rate δ. Equation (2.1c) describes the free virus dynamics: virus is produced by infected cells at rate p and cleared at rate c. A time-delayed version incorporates the eclipse phase of the infected cells as follows:

$$\frac{dT}{dt} = -\beta VT, \tag{2.2a}$$

$$\frac{dI_1}{dt} = \beta VT - kI_1, \tag{2.2b}$$

$$\frac{dI_2}{dt} = kI_1 - \delta I_2, \tag{2.2c}$$

$$\frac{dV}{dt} = pI_2 - cV, \tag{2.2d}$$

where I_1 denotes infected cells in the eclipse phase, and I_2 denotes infected cells capable of producing free virus. The parameter k controls the duration of the eclipse phase. Although there are several variants of TCL models, we focus on the basic forms that do not include target cell regeneration or a dynamic immune response.

The effect of an antiviral drug is often modeled as a fold change in β or p, expressed as $\beta = \beta_0(1-u)$ or $p = p_0(1-u)$, where β_0 and p_0 are the baseline values, and u represents the fraction of decrease due to the drug effect. We simply refer to u as the *drug effect*. In a pharmacodynamic setting, u is a function of the drug concentration $C(t)$, which depends on time. For example,

$$u(t) = \frac{C(t)}{EC_{50} + C(t)} \tag{2.3}$$

where EC_{50} is the half-maximal effective concentration. In this paper, we solely focus on the drug effect, without explicitly modeling drug concentration, and assume that the drug effect indefinitely persists for simplicity.

2.2. Numerical experiments

In this section, we examine the TCL model with the baseline parameter values reported in [20] using numerical simulations. The parameter values are summarized in Table 1. In the original study, this model was used to study SARS-CoV-2-infected individuals treated with the oral antiviral Paxlovid.

Viral reduction after treatment is often included as a predictive biomarker in clinical trials [4]. The case $u = 0$ (no reduction in β or p) serves as the control, thus providing a baseline value of viral reduction to which treated patients can be compared. We simulate antiviral treatment by reducing the parameter values to various degrees at day 2.7 (treatment before the VL peak) and day 5.5 (treatment after the VL peak). As seen in Figure 1, the extent of viral reduction over a given time period is largely determined by the slope of the VL curve, which corresponds to the asymptotic decay rate of the free

virus (V). For treatment initiated before the VL peak, only large reductions in β or p, which lead to an immediate reduction of VL, appear effective in reducing the VL. The model predicts that moderate reductions in β or p ($u = 0.9$) result in a much slower viral decline and a notably prolonged persistence above the LOQ, limit of quantification, in some cases exceeding 60 days. For treatment initiated after the VL peak, the differences among cases are not discernible when the drug effect is on β. Additionally, we observe that reductions in p can induce a transient reduction in the VL but still fail to alter the asymptotic rate of the VL decline if the treatment begins after the VL peak. A comprehensive view is provided in Figure 2, where the asymptotic VL reduction rate is plotted against the drug effect and treatment time. The result suggests that antiviral drugs have little to no effects on viral reduction when administered approximately after day 4, as indicated by the homogeneous color pattern beyond this point. This highlights the importance of initiating an antiviral treatment for COVID-19 before the VL peak, which is consistent with the findings reported in [21]. Of particular interest is the non-monotonic relationship between the VL reduction rate and the drug effect for treatments applied between days 2 and 3. The analytically derived threshold (3.8) (see Section 3.1) that delineates monotonicity is overlaid on the heat map and agrees well with the simulation results.

Table 1. Model parameters for Patient 1 ([20]) in the TCL model with eclipse stage (2.2). The initial conditions are $T(0) = 8 \times 10^7, I_1(0) = 1, I_2(0) = 0$ and $V(0) = 0$.

Parameter	Value	Units	Description
β	2.32×10^{-9}	[mL RNA copies^{-1} day^{-1}]	Infection rate
p	1.77×10^3	[RNA copies mL^{-1} day^{-1}]	Viral production rate
k	4	[day^{-1}]	Rate of transition from eclipse phase to productively infected cells.
δ	1.7	[day^{-1}]	Death rate of productively infected cells.
c	10	[day^{-1}]	Clearance rate of free virions.

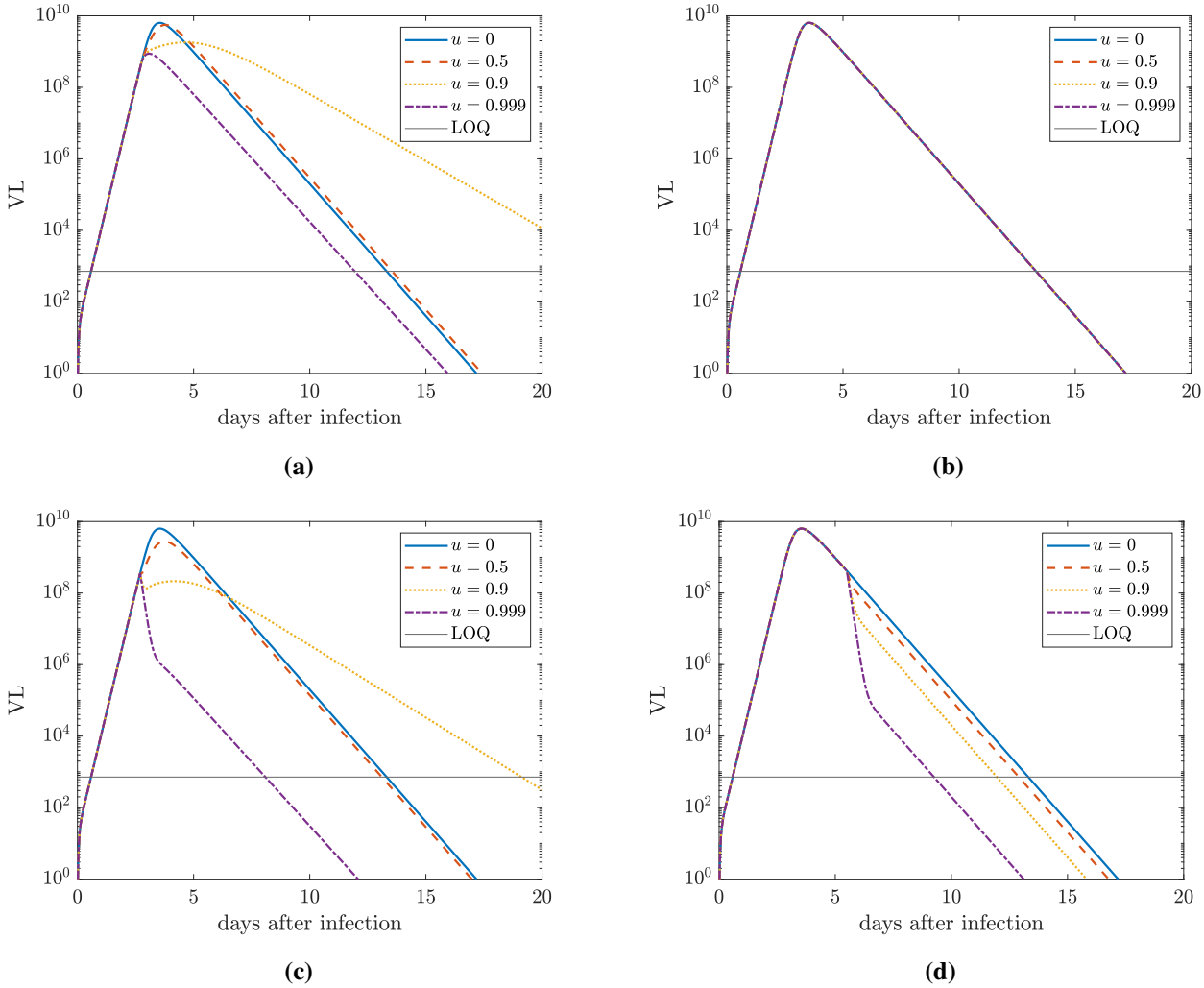


Figure 1. VL dynamics under different drug effect (u) starting before and after the VL peak described by the TCL model with an eclipse stage (2.2). Drug effect (u) causes a fold change of the parameter from its baseline values in the form $\beta = (1-u)\beta_0$ and $p = (1-u)p_0$, where β_0 and p_0 are baseline values. (a) Drug effects on β started at 2.7 days after infection. (b) Drug effects on β started at 5.5 days after infection. (c) Drug effects on p started at day 2.7 days after infection. (d) Drug effects on p started at 5.5 days after infection. For all cases, the parameter values and initial conditions are taken from Patient 1 in [20] and summarized in Table 1. All curves are obtained by numerical integration of (2.2) using MATLAB's `ode45` solver.

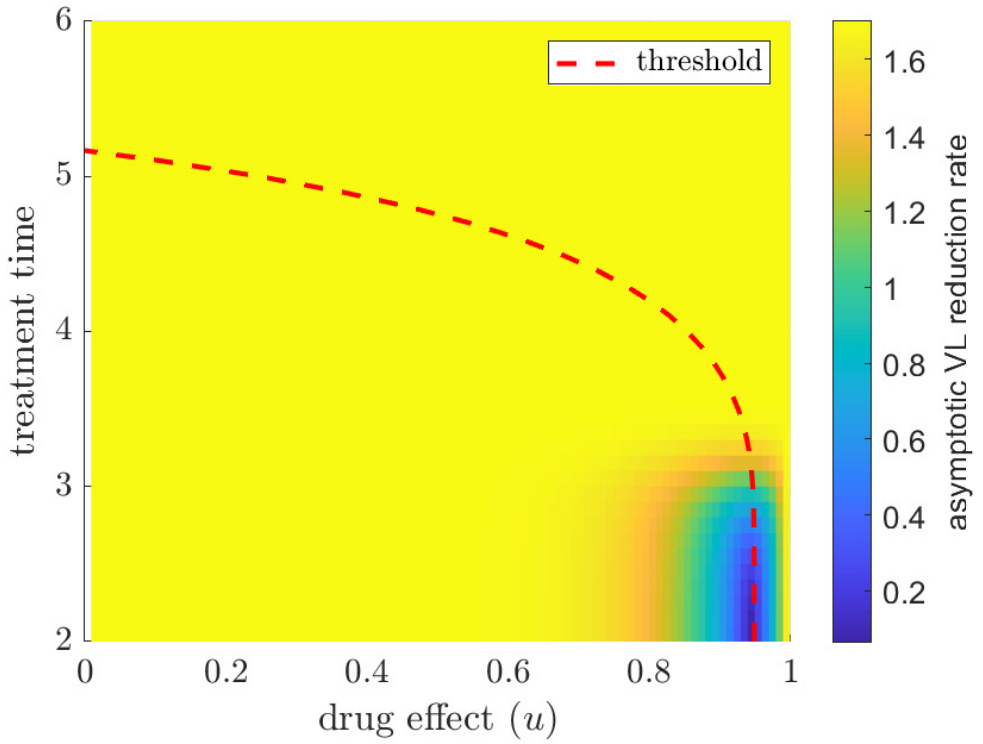


Figure 2. Dependence of the asymptotic VL reduction rate on the treatment time (days after infection) and drug effect (μ) in the TCL model with an eclipse stage (2.2). The color indicates numerical estimated asymptotic VL reduction rates as a function of the drug effect (μ) and the day of treatment initiation after infection. The dashed line represents the threshold drug effect below which the asymptotic VL reduction rate monotonically increases with the drug effect. The threshold drug effect is based on the critical infection rate $\beta^\dagger$ given in (3.8). Baseline parameter values and initial conditions are taken from Patient 1 in [20] and summarized in Table 1.

3. Analysis of viral reduction rate

We conduct a mathematical analysis with a focus on the asymptotic viral reduction rate to confirm our findings from the numerical simulations in the previous section. Section 3.1 explores the dependence of the viral reduction rate on parameters in an n-stage TCL model. In Section 3.2, we generalize the infection term in the TCL model and propose a modeling choice that mitigates the non-monotonic behavior revealed by the numerical experiments in Section 2.2.

3.1. N-stage model

The basic versions of the TCL model (2.1 and 2.2) have been extensively analyzed in [22–24]. Model (2.2) adds an eclipse stage to (2.1). Multi-stage extensions of the TCL model incorporating eclipse phases have been introduced previously [25, 26]. Inspired by these works, we consider a general n-stage model as follows:

$$\frac{dT}{dt} = -\beta VT, \tag{3.1a}$$

$$\frac{dI_1}{dt} = \beta VT - k_1 I_1, \tag{3.1b}$$

$$\frac{dI_2}{dt} = k_1 I_1 - k_2 I_2, \tag{3.1c}$$

$$\vdots$$

$$\frac{dI_{n-1}}{dt} = k_{n-2} I_{n-2} - k_{n-1} I_{n-1}, \tag{3.1d}$$

$$\frac{dI_n}{dt} = k_{n-1} I_{n-1} - \delta I_n, \tag{3.1e}$$

$$\frac{dV}{dt} = pI_n - cV. \tag{3.1f}$$

The positivity and steady-state analysis in [22–24] is easily extended to the n-stage model. Here, we state the relevant results about steady states for completeness. There is only one equilibrium, which is the following set:

$$\{(T, I_1, \cdots, I_n, V) : V = 0 \text{ and } I_j = 0 \text{ for } j = 1, \cdots, n\}.$$

Furthermore, the minimum global attractor is as follows:

$$\{(T, I_1, \cdots, I_n, V) : T < T_c \text{ and } V = 0 \text{ and } I_j = 0 \text{ for } j = 1, \cdots, n\}, \tag{3.2}$$

where $T_c = \frac{c\delta}{p\beta}$.

By differentiating with respect to time (t), it can be shown that the following quantity is conserved:

$$L_0 = T - \frac{c\delta}{\beta p} \log T + \sum_{j=1}^{n} I_j + \frac{\delta}{p} V,$$

where L_0 is a constant determined by the initial conditions. For a general initial condition $T_0, I_{1,0}, \cdots, I_{n,0}, V_0$, we have $L_0 = T_0 - \frac{c\delta}{\beta p} \log T_0 + \sum_{j=1}^{n} I_{j,0} + \frac{\delta}{p} V_0$.

It follows the same argument as in [22–24] that the system approaches the equilibrium

$$(T, I_1, \cdots, I_n, V) \to (T_\infty, 0, \cdots, 0),$$

where T_∞ is a solution of the equation

$$T_\infty - \frac{c\delta}{\beta p} \log T_\infty = L_0 \tag{3.3}$$

and must satisfy $T_\infty < \frac{c\delta}{\beta p}$, since the system approaches the minimum global attractor (3.2). We note that T_∞ does not depend on k_i. In preparation for the later analysis, we can implicitly differentiate (3.3) with respect to β to get the following:

$$\frac{\partial T_\infty}{\partial \beta} = -\frac{\frac{c\delta}{\beta^2 p} (\log T_\infty - \log T_0)}{1 - \frac{c\delta}{\beta p T_\infty}} < 0. \tag{3.4}$$

Similarly, we get the following:

$$\frac{\partial T_\infty}{\partial p} = -\frac{\frac{c\delta}{\beta p^2}\log\left(\frac{T_\infty}{T_0}\right) + \frac{\delta}{p^2}V_0}{1 - \frac{c\delta}{\beta p T_\infty}} < 0.$$

T_∞ is also known as the final size in the literature on mathematical epidemics [27] and dates back to [28].

Note that the equilibrium set is a normally hyperbolic invariant manifold [29]. To find the asymptotic decay rate of V, we linearize System (3.1) at the equilibrium

$$(T_\infty, 0, \cdots, 0),$$

and focus on the normal direction by discarding the T-direction, which has a zero eigenvalue. As a result, we get the characteristic equation, whose roots determine the exponential rates of decay (or growth) of small perturbations near the equilibrium

$$\left(\prod_{i=1}^{n-1}(k_i + \lambda)\right)(\delta + \lambda)(c + \lambda) = \beta p\, T_\infty \left(\prod_{i=1}^{n-1} k_i\right). \tag{3.5}$$

All roots of (3.5) have negative real parts, which we now verify. Let $P(\lambda) = \left(\prod_{i=1}^{n-1}(k_i + \lambda)\right)(\delta + \lambda)(c + \lambda)$ denote the left-hand side and $K = \beta p\, T_\infty \prod_{i=1}^{n-1} k_i$ the right-hand side of (3.5). Since $T_\infty < \frac{c\delta}{\beta p}$, we have $K < \delta c \prod_{i=1}^{n-1} k_i = P(0)$. Note that $P(\lambda)$ is a real polynomial of degree $n+1$ whose roots are $-k_1, \ldots, -k_{n-1}, -\delta, -c$, which are all negative. For any real $\lambda \geq 0$, each factor in $P(\lambda)$ is positive and at least as large as the corresponding factor at $\lambda = 0$, so $P(\lambda) \geq P(0) > K$. More precisely, on the real axis, $P(\lambda)$ is strictly increasing for $\lambda > -\min\{k_1, \ldots, k_{n-1}, \delta, c\}$, and $P(0) > K > 0 = P(-m)$, where $m = \min\{k_1, \ldots, k_{n-1}, \delta, c\}$, so there is exactly one real root $\lambda_1 \in (-m, 0)$. For complex roots, note that for any root $\lambda = a + bi$ of $P(\lambda) = K > 0$, we have $|P(\lambda)| = \prod |k_i + \lambda| \cdot |\delta + \lambda| \cdot |c + \lambda|$. If $a \geq 0$, then $|k_i + \lambda| \geq k_i$, $|\delta + \lambda| \geq \delta$, and $|c + \lambda| \geq c$, giving $|P(\lambda)| \geq P(0) > K$, which is a contradiction. Hence, all roots have negative real parts. Define the following:

$$r = -\lambda_1, \tag{3.6}$$

which is the asymptotic decay rate of VL, that is, $V \sim O(\exp(-rt))$. Furthermore,

$$0 < r < \min\{k_1, \ldots, k_{n-1}, \delta, c\},$$

since $\lambda_1 \in (-m, 0)$ as shown above. It is of interest to determine the dependence of r on the parameters β and p. Clinically, this dependence indicates how a drug-regulated parameter would affect VL reduction.

A geometric argument, noting that the left-hand side of (3.5) is a polynomial of λ and the right-hand side is a constant controlled by βT_∞, shows that $\frac{\partial r}{\partial \beta}$ has the opposite sign as $\frac{\partial}{\partial \beta}(\beta T_\infty)$, where we treat T_∞ as a function of β. Using (3.4), we can calculate that

$$\frac{\partial}{\partial \beta}(\beta T_\infty) = \frac{T_0 + \sum_{i=1}^{n} I_{i,0} + \frac{\delta}{p}V_0 - \frac{c\delta}{\beta p}}{\gamma - \frac{c\delta}{\beta p T_\infty}}. \tag{3.7}$$

Therefore, there exists a threshold β value

$$\beta^\dagger = \frac{c\delta}{pT_0 + p\sum_{i=1}^{n} I_{i,0} + \delta V_0} \tag{3.8}$$

such that $\frac{\partial r}{\partial \beta} < 0$ if $\beta < \beta^\dagger$ and $\frac{\partial r}{\partial \beta} > 0$ if $\beta > \beta^\dagger$. Similarly, we have a threshold p value

$$p^\dagger = \frac{c\delta}{\beta(T_0 + \sum_{i=1}^{n} I_{i,0})}$$

such that $\frac{dr}{dp} < 0$ if $p < p^\dagger$ and $\frac{\partial r}{\partial p} > 0$ if $p > p^\dagger$.

Since β and p are parameters typically assumed to be reduced by antiviral drugs, the above results show that there can be a non-monotonic relationship between the drug effect and the VL reduction rate in the TCL models. In particular, if the drug effect does not move the parameter values below a certain threshold, then reducing the infection rate or the virus proliferation rate may make the VL reduction rate slower. We will propose a modification to the TCL model and show that it mitigates this non-monotonic behavior in the next section.

Although it is not a main focus of this analysis, we can also deduce the sign of $\frac{\partial r}{\partial k_i}$. The parameters k_i are the rates at which infected cells transition through intermediate stages. A typical parameter range is $k_i > \max\{\delta, c\}$ for $i = 1, 2, \cdots, n-1$. Within this parameter range, we can show that $\frac{\partial r}{\partial k_i} > 0$, which shows that a faster transition through intermediate infection stages leads to a faster decay of the VL (see Appendix A).

3.2. Modified model

In order to mitigate the aforementioned non-monotonic behavior, we replace the mass-action term (TV) in (3.1a) and (3.1b) with a more general form $f(T)V$. Specifically, (3.1a) and (3.1b) now become the following, respectively:

$$\frac{dT}{dt} = -\beta f(T)V, \tag{3.9a}$$

$$\frac{dI_1}{dt} = \beta f(T)V - \delta I_1, \tag{3.9b}$$

respectively, while keeping the other equations in (3.1) the same. We start with general conditions on $f(T)$:

(C1) $f(T)$ is non-negative, non-decreasing;

(C2) $f(0) = 0$ and $f(T_{max}) = T_{max}$;

(C3) $f(T) \to 0$ as $T \to 0$ at least linearly.

Since $f(T)$ represents the availability of susceptible cells, Condition (C1) is needed for $f(T)$ to be biologically meaningful. In Condition (C2), T_{max} is the maximum amount of target cells, which is typically the amount of target cells at the beginning of the infection. Condition (C2) ensures that $f(T)$ is equivalent to the mass action term at least at the two extremes. Thus, the modified model maintains certain connections with the original model. Condition (C3) ensures that the asymptotic decay rate still depends on β and p, which we will explain below.

The steady-state analysis is easily extended to the modified model. Specifically, the modified model has the same equilibrium set. For the modified model, there also exists the following conserved quantity:

$$T + \sum_i^n I_i + \frac{\delta}{p}V - \frac{c\delta}{\beta p}\int_{T_*}^{T}\frac{du}{f(u)} \;=\; L_0.$$

where T_* and L_0 are constants. For a general initial condition $T_0, I_{1,0}, \cdots, I_{n,0}, V_0$, we have $T_* = T_0$ and $L_0 = T_0 + \sum_i^n I_{i,0} + \frac{\delta}{p}V_0$ and the system approaches $(T_\infty, 0, \cdots, 0)$. Condition (C3) implies that $\int_{0^+} f(T)dT = \infty$, so that T_∞ cannot reach zero (i.e., $T_\infty > 0$). Therefore the dependence of the asymptotic decay rate $r = -\lambda$ on β and p can be established similarly as before by analyzing the characteristic equation as follows:

$$\left(\prod_{i=1}^{n-1}(k_i + \lambda)\right)(\delta + \lambda)(c + \lambda) = \beta p\, f_\infty\left(\prod_{i=1}^{n-1} k_i\right). \tag{3.10}$$

where we use a short-hand notation $f_\infty = f(T_\infty)$.

We note that without Condition (C3), T_∞ may reach zero and make the right hand side of (3.10) zero. In this case, $r = \min\{\delta, c, k_1, \cdots, k_{n-1}\}$.

From here, we could proceed with the same calculation as in Section 3.1. However, the notation would become more cluttered, making it harder to gain further insight. Thus, we proceed with nondimensionalization and work with an approximate model to study an alternative choice of $f(T)$ that mitigates the non-monotonic behavior. In addition, using the dimensionless form shows that the effects of parameters β and p on the dynamics are essentially the same when ignoring fast transients.

3.3. Approximate modified model

Assuming that the virus dynamics are on a much faster time scale allows dimension reduction, which results in the so-called quasi-steady-state approximation [30]. To illustrate the separation of time scales, we conduct a nondimensionalization of System (2.1) by introducing the dimensionless variables $\hat{t} = \frac{t}{t_c}$, $\hat{T} = \frac{T}{T_c}$, $\hat{I} = \frac{I}{I_c}$ and $\hat{V} = \frac{V}{V_c}$, where the characteristic scales are as follows:

$$t_c = \frac{1}{\delta},\; V_c = \frac{p}{c}T(0),\; T_c = I_c = T_{max}. \tag{3.11}$$

This gives the following dimensionless system:

$$\begin{aligned}
\frac{d\hat{T}}{d\hat{t}} &= -\hat{\beta}f(\hat{T})\hat{V},\\
\frac{d\hat{I}}{d\hat{t}} &= \hat{\beta}f(\hat{T})\hat{V} - \hat{I},\\
\frac{d\hat{V}}{d\hat{t}} &= \frac{1}{\epsilon}(\hat{I} - \hat{V}),
\end{aligned} \tag{3.12a}$$

where $\epsilon = \delta/c$ and

$$\hat{\beta} = \frac{\beta p T_{max}}{\delta c},$$

which shows that β and p have the same effect on the dynamics.

We will drop the hats (ˆ) for the dimensionless variables for notational convenience. Assuming that the clearance of free virus is much faster than the clearance of infected cells, we have $\epsilon \ll 1$. Consider the asymptotic expansion $x = x_0 + \epsilon x_1 + \epsilon^2 x_2^2 + \cdots$, where x can be either T, I, or V. To the leading order, $I_0 = V_0$, and hence

$$\frac{dT}{dt} = -\beta f(T) I, \tag{3.13a}$$

$$\frac{dI}{dt} = \beta f(T) I - I, \tag{3.13b}$$

where we have dropped the subscript 0 for notational convenience since we focus on the leading-order solution. We call (3.13) the approximate modified model.

Again, we find a conserved quantity for (3.13) as follows:

$$I + T - \frac{1}{\beta}\int_{T_0}^{T} \frac{1}{f(u)} du = T_0 + I_0.$$

Because of our scaling, the initial condition for the target cell is $T(0) = 1$. At the beginning of an infection, one may assume that the target cell density is at its maximum and an infected cell is introduced. For an infection trajectory to occur (i.e., the trajectory starting at (1,0) goes into the positive quadrant), it requires $\frac{dI}{dT}\Big|_{T=1} < 0$, and hence gives $\beta > 1$. We note that $\frac{d^2 I}{dT^2} < 0$; thus, on the T-I phase plane, the trajectory is concave down. A biological relevant solution that represents an infection trajectory is a heteroclinic orbit that starts at $(1, 0)$ and ends at $(T_\infty, 0)$, where T_∞ satisfies $T_\infty - \frac{1}{\beta}\int_{T_0}^{T_\infty} \frac{1}{f(u)} du = 1$. For a general initial condition T_0 and I_0, the trajectory follows

$$T_\infty - \frac{1}{\beta}\int_{T_0}^{T_\infty} \frac{1}{f(u)} du = T_0 + I_0. \tag{3.14}$$

For the approximate modified model (3.13), the asymptotic decay rate is $r = 1 - \beta f(T_\infty)$. Thus,

$$\begin{aligned} \frac{dr}{d\beta} &= -\frac{d(\beta f(T_\infty))}{d\beta} \\ &= \frac{-\frac{1}{\beta} + f_\infty + f'_\infty \cdot (T_0 + I_0 - T_\infty)}{\frac{1}{\beta f_\infty} - 1}. \end{aligned}$$

The denominator is positive. The sign of $\frac{dr}{d\beta}$ only depends on the numerator. For a very large β, the numerator is positive. For a very small β, the numerator is negative because f is bounded and, assuming that f' is also bounded. Therefore, there must exist a $\beta^\dagger$ such that $\frac{dr}{d\beta}\Big|_{\beta=\beta^\dagger} = 0$. This shows that the non-monotonic relationship between the viral reduction rate (r) and the dimensionless parameter $\hat{\beta}$ (and hence the original parameter β and p) exists in a broad situation. Furthermore, $\beta^\dagger$ must satisfy the following equation:

$$-\frac{1}{\beta} + f_\infty + f'_\infty \cdot (T_0 + I_0 - T_\infty) = 0. \tag{3.15}$$

For $f(T) = T$ (the original model), solving (3.15) gives $\beta^\dagger = \frac{1}{T_0+I_0}$. For a general $f(T)$, an explicit solution may not be available. In an attempt to get a numerical solution, we can design an iterative algorithm based on (3.15). Let $\alpha = 1/\beta$ and define

$$h(T_\infty) = f(T_\infty) + f'(T_\infty)(T_0 + I_0 - T_\infty) \tag{3.16}$$

and

$$T_\infty = g(\alpha), \tag{3.17}$$

where $g(\alpha)$ is implicitly defined by (3.14). The iterative solver for $\alpha^\dagger$ is described below.

Algorithm 1 Iterative Solver for $\alpha^\dagger$

1: **Input:** Initial guess α_0, tolerance ε, maximum iterations $N_{\max}$
2: **Output:** Approximate solution $\alpha^\dagger$
3: Initialize $j \leftarrow 0$
4: **repeat**
5: Compute $T_{\infty,j+1} \leftarrow g(\alpha_j)$
6: Compute $\alpha_{j+1} \leftarrow h(T_{\infty,j+1})$
7: $k \leftarrow k + 1$
8: **until** $|\alpha_j - \alpha_{j-1}| < \varepsilon$ **or** $k \geq N_{\max}$
9: Set $\alpha^\dagger \leftarrow \alpha_j$

We note that the above algorithm is essentially an iterative mapping

$$\alpha_{j+1} = f_\infty(\alpha_j) + f'_\infty(\alpha_j)(T_0 + I_0 - T_\infty(\alpha_j)). \tag{3.18}$$

or written in terms of T_∞ as

$$T_{\infty,j} = g(h(T_{\infty,j-1})), \tag{3.19}$$

which gives rise to the following theorem.

Theorem 1 (Non-monotonicity of viral reduction rate)**.** *Consider the approximate modified model* (3.13) *with f satisfying conditions (C1)–(C3), and suppose $f \in C^2$. Let $r = 1 - \beta f(T_\infty)$ be the asymptotic viral reduction rate, where T_∞ is determined by* (3.14)*. Define g and h as in* (3.17) *and* (3.16)*, respectively. If*

$$|g'(h(x))h'(x)| < 1 \quad \textit{for all } x \in [0, T_0], \tag{3.20}$$

then there exists a unique critical value $\beta^\dagger > 0$ such that $\frac{dr}{d\beta}\Big|_{\beta=\beta^\dagger} = 0$. Moreover,

(i) $\frac{dr}{d\beta} < 0$ for $\beta < \beta^\dagger$ (reducing β increases the viral reduction rate);

(ii) $\frac{dr}{d\beta} > 0$ for $\beta > \beta^\dagger$ (reducing β decreases the viral reduction rate).

Proof. The critical value $\alpha^\dagger = 1/\beta^\dagger$ is a fixed point of the mapping (3.18). Since $T_\infty \in [0, T_0]$ and condition (3.20) holds, the composite map $g \circ h$ is a contraction on $[0, T_0]$. By the Contraction Mapping Theorem, the fixed point exists and is unique. The sign of $\frac{dr}{d\beta}$ follows from the analysis of the numerator in the expression for $\frac{dr}{d\beta}$, which changes sign exactly once at $\beta = \beta^\dagger$. □

A direct calculation can write (3.20) in terms of $f(x)$ as follows:

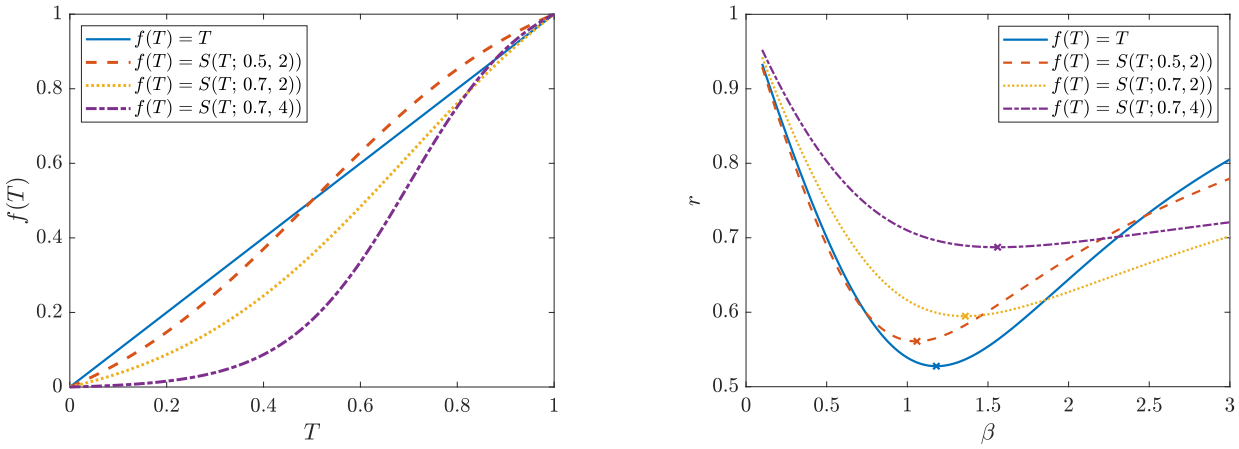
$$\left| \frac{f(x)f''(x)\left(\int_{T_0}^{x} \frac{1}{(T_0+I_0-x)f(u)}du\right)^2}{(T_0+I_0-x)f(x)\int_{T_0}^{x}\frac{1}{f(u)}du - 1} \right| < 1 \text{ for all } x \in (0, T_0). \tag{3.21}$$

Moreover, (3.18) looks like a linear approximation of $f(T)$ at $T = T_\infty(\alpha_j)$. An algorithmically inspired and biological plausible choice of $f(\cdot)$ is the following sigmoid-shaped function:

$$S(x; x_0, \kappa) = \frac{\tanh(\kappa(x - x_0)) + \tanh(\kappa x_0)}{\tanh(\kappa(1 - x_0)) + \tanh(\kappa x_0)}, \tag{3.22}$$

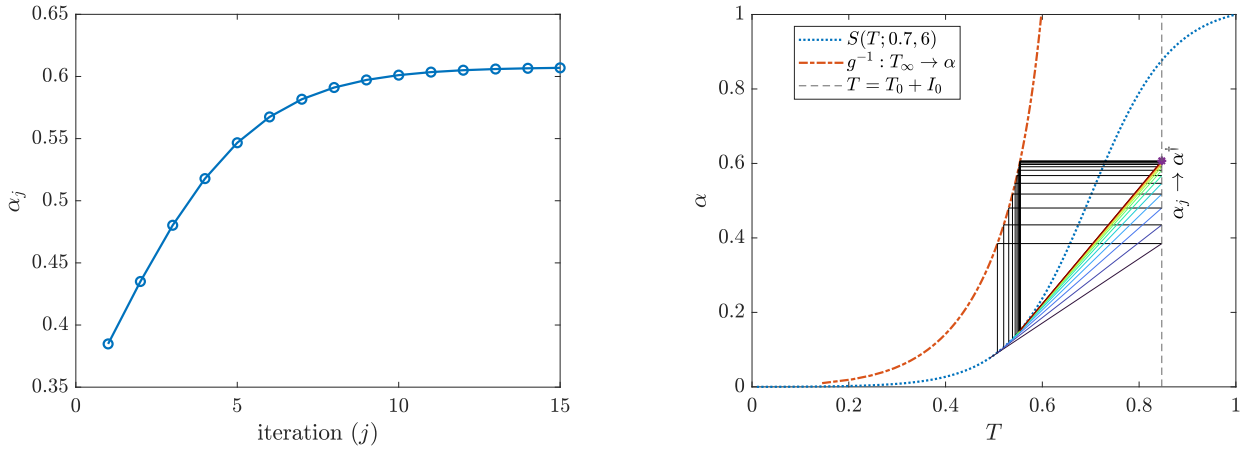
where x_0 is the inflection point, and κ controls the steepness of the sigmoid (Figure 3a). If we choose $x_0 > T_0$ and κ large enough, it is possible to achieve an arbitrarily small $\alpha^\dagger$ (hence arbitrarily large $\beta^\dagger$) so that the non-monotonicity does not appear in a typical range of parameter values (Figure 3b).

However, although the sigmoid function $S(x; x_0, \kappa)$ does not facilitate a direct calculation to verify (3.20) or (3.21), a numerical verification shows (3.21) is satisfied in the cases we consider. Moreover, a convergence was often observed in a few iterations in our numerical experiments (Figure 4).



(a) **(b)**

Figure 3. (a) The sigmoid infection function $S(T; x_0, \kappa)$ for different choices of (x_0, κ), compared with the linear $f(T) = T$ (mass-action infection term). (b) Asymptotic VL reduction rate (r) as a function of β for different choices of $f(T)$ for the approximate modified model (3.13). The dimensionless approximate modified model (3.13) is used here with $T(0) = 0.6781$ and $I(0) = 0.1688$. The cross ($\times$) indicates the lowest point in each curve.



(a) **(b)**

Figure 4. (a) α_j values plotted for each iteration showing that $\alpha_j \to \alpha^\dagger$ in a few iterations. (b) Visualization of the iteration where g^{-1} is the inverse function of $g(x)$ defined in (3.17).

4. Discussion

In this paper, we conducted a mathematical and simulation-based analysis of the TCL model, thus contributing to a thorough understanding of this widely used model. The focus is on the parameters β (infection rate) and p (viral production rate), which are often considered the main targets of antiviral drug effects. Our simulations reveal that in a typical infection, the TCL models may predict a non-monotonic relationship between the VL reduction rate and the strength of drug effect. More specifically, there is a parameter region (which often includes typical parameter values used in the literature) in which reducing the infection rate or the viral production rate (antiviral drug effects) leads to a prediction of a slower VL reduction. This finding has important implications: if the TCL models are correct about the non-monotonic exposure-response curve, then the standard practice of estimating antiviral potency using a single parameter such as the half-maximal effective concentration (EC_{50}) as in (2.3) is at best misleading; indeed, it has been observed that the EC_{50} alone is not sufficient to characterize antiviral potency, as the slope of the dose-response curve imposes class-specific limits on the inhibitory potential of drugs [31], and the in vivo EC_{50} often differs significantly from the in vitro measurements [32]; on the other hand, if the TCL models are wrong about the non-monotonic exposure-response curve, then fitting them to data to estimate the EC_{50} will yield unreliable results.

Indeed, fact sheets published by the U.S. Food and Drug Administration (FDA) for monoclonal antibody drugs bamlanivimab and REGEN-COV explicitly report that a flat exposure-response relationship for the VL was identified [17, 18], which suggests that the EC_{50} was not identified. Although we cannot confirm whether TCL models were used in these analyses, the reports show that moderate drug effects are not covered by existing data, thus raising caution about extrapolating the TCL model predictions to regions without data support.

Moreover, for certain intermediate values of β, the model predicts notably long sojourn times of the VL above the LOQ, extending as long as 60 days. Although the TCL model that ignores immune responses is not intended to simulate long-term infections, the short-term slow decline of the VL

observed under an intermediate fold change of β from its baseline values (Figure 1) warrants careful examination. Such outcomes highlight the need for further experimental investigation to confirm or refute these predictions and to evaluate whether additional mechanisms, such as immune-mediated clearance, must be incorporated for improved realism.

We acknowledge that prolonged viral shedding is clinically documented in immunocompromised or immunosuppressed individuals. For example, Aydillo et al. [33] reported SARS-CoV-2 RNA detection up to 78 days after symptom onset in patients on immunosuppressive therapy, and Scherer et al. [34] documented infections lasting 42 to 302 days in patients with B-cell deficiencies. Since the TCL model does not include an immune response, prolonged viral persistence per se is not inconsistent with the model's assumptions. The specific prediction that intermediate drug effects lead to a *slower* VL decline than no treatment at all is a notable feature of TCL models that merits careful scrutiny. Target cell preservation has been suggested as an explanation: drug effects slow the depletion of target cells, which leads to more cells available for subsequent rounds of infection [35]. However, as shown in our analysis, T_∞ increases monotonically with the drug effect (Eq (3.4)), so very high drug effects also preserve target cells, yet the VL declines rapidly. The non-monotonicity of the VL reduction rate arises because $\frac{\partial r}{\partial \beta}$ (or $\frac{\partial r}{\partial p}$) has the same sign as $\frac{\partial(\beta T_\infty)}{\partial \beta}$ (or $\frac{\partial(pT_\infty)}{\partial p}$), as seen from the characteristic equation (3.5), which reflects a trade-off between a decreasing β (or p) and an increasing T_∞ — not from target cell preservation alone. Nonetheless, the resulting non-monotonic dose-response complicates the use of VL reduction as a biomarker of drug efficacy.

Some experimental evidence suggestive of non-monotonic dose-VL relationships has been reported in nonhuman primate studies of favipiravir against SARS-CoV-2 [15] and the Ebola virus [16], though the small sample sizes limit the statistical significance and the underlying mechanisms — drug toxicity and survival bias, respectively — may differ from the target cell preservation effect predicted by the TCL models.

In a SARS-CoV-2 human challenge study that involved healthy volunteers [36], the weak antiviral remdesivir did not noticeably prolong the viral clearance, and Iyaniwura et al. [37] showed that the simple TCL model provided the worst fit among seven candidate models for these data, with the best model incorporating both innate and adaptive immune responses. On the other hand, Sachak-Patwa et al. [38] demonstrated that a simple TCL model can reproduce influenza viral load dynamics generated by a more complex immune-response model, with the TCL model parameters effectively absorbing the effects of differing immune responses. This suggests that TCL models retain practical value to interpret the VL data even when immune dynamics are the underlying driver. Additionally, we note that Paxlovid-associated viral rebound — where VL temporarily becomes undetectable before flaring back up [35] — is a distinct phenomenon from what we study here, namely the prolonged monotonic decline of the VL under intermediate drug effects.

On the logarithmic scale, clinical VL data often exhibit an approximately linear decline [5]. In our simulations, the VL reduction follows an exponential trajectory shortly after reaching the peak VL. These observations motivated us to study the asymptotic reduction rate of the VL (denoted as r in this paper), which is a quantity more amenable to mathematical analysis. Few studies have addressed the asymptotic reduction rate of VL, and those that exist are neither focused on the parameter dependence [39] nor intended to understand the drug effects on viral dynamics [40]. Our analysis confirmed that the non-monotonic relationship between the drug effects and the VL reduction rates exists in the TCL models, arising from the competing effects of β (or p) and T_∞ in the characteristic

equation (3.5) as discussed above.

Furthermore, we found that the non-monotonic relationship exists in a broad family of TCL models (3.9). We identified the existence and uniqueness of a critical parameter threshold ($\beta^\dagger$) that delineates monotonicity. The generalization of the standard TCL model facilitated a modification of the model formulation to mitigate the non-monotonic behavior. We found that a sigmoid function (3.22) can reduce the extent of non-monotonicity in the dose-response relationship. A sigmoid function is also biologically plausible, given the nonlinear susceptibility of target cells: a sigmoidal $f(T)$ captures the threshold, cooperativity, and saturation effects not represented by the standard mass-action formulation. This simple modification provides a practical pathway to improve model-based inferences in drug evaluation studies without increasing the model complexity.

Our analysis of the modified model used a dimensionless form based on the quasi-steady-state approximation, but the results are easily extended to the full model. The dimensionless form shows that the infection rate β and the viral reproduction rate p have the same effects on the asymptotic VL reduction. Additionally, it allows us to see the transient effect of p as it appears in the scaling (3.11), which indicates a fast transient change in V when p is changed. This explains the more prominent transient effects of changing p compared to changing β, as seen in Figure 1. Although the viral clearance rate c is sometimes a parameter that may be increased by a drug effect, it is not a focus of this paper. In the typical parameter range we consider, c is already large and a further increase in c would not make a noticeable difference.

The steady-state approximation is justified by the much faster viral clearance rate compared to the death rate of infected cells and transition rates through intermediate infection stages. Several studies [19, 41] have reported that including a delay in viral production (eclipse phase) improved data fitting or yielded more reasonable parameter estimates. For the n-stage model (3.1), our preliminary analysis (see Appendix) showed that the time delay decreases the VL reduction. The n-stage model can be written in the form of integro-differential equations with an Erlang-type distributed time delay via the so-called linear chain trick (for example, see [42] Chapter 3.6). There is renewed interest in generalizing the linear chain trick [43]. Whether our results hold for more general time delays is left to a future study.

Although our work focuses on TCL models, which describe within-host viral dynamics, the same or similar equations are also used to model population-level epidemics, known as the susceptible–infected–recovered (SIR) model without birth and death [44]. The connection is even more prominent when VL is measured in wastewater and modeled with SIR-type models [45]. In particular, wastewater-based epidemiological studies that integrate viral shedding dynamics with surveillance data [46] or account for post-recovery viral shedding [47] rely on an accurate characterization of viral decline rates, to which our findings on non-monotonic parameter dependence are directly relevant. In the field of mathematical epidemiology, the focus is often on the final size of the remaining susceptible individuals at the end of an epidemic (corresponding to T_∞ in our notation). However, to our knowledge, the parameter dependence of the asymptotic convergence rate has not received attention in existing studies that model epidemics using equations the same as or similar to the TCL models we consider here.

A natural question is whether the proposed modified model with a sigmoid infection term can fit clinical data as well as the standard TCL model. Rather than fitting clinical data directly, we demonstrate that the modified model can closely replicate the VL dynamics produced by the standard TCL model through parameter re-estimation. Specifically, for several choices of sigmoid parameters (x_0, κ), the

kinetic parameters (β, k, δ, c) were re-estimated to minimize the mean squared error in $\log_{10}$ VL against the standard model output. The results (Appendix B, Figure B.1) show that the modified model can reproduce the standard TCL model's VL trajectory with high fidelity, thus suggesting that the modified model has sufficient flexibility to approximate the standard model's dynamics, while providing the additional benefit of reduced non-monotonicity in the dose-response relationship.

5. Conclusions

Overall, our work highlights the importance of careful analysis to uncover and understand subtle but consequential features of mathematical models that would otherwise remain unnoticed. While the TCL model has been invaluable in virology and pharmacology, our results underscore the necessity of critical evaluation when it is applied to antiviral efficacy studies. In addition to the simple improved model formulation we proposed, future work should extend these analyses by incorporating immune dynamics and alternative viral clearance mechanisms, which may resolve some of the issues identified here while preserving the analytical tractability of the model.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This project was supported by an appointment to the regulatory science research program at the US Food and Drug Administration. LH's work was supported in part by an appointment to the Research Participation Program at CBER, US Food and Drug Administration, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the US Department of Energy and FDA.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. Y. M. Bar-On, A. Flamholz, R. Phillips, R. Milo, SARS-CoV-2 (COVID-19) by the numbers, *eLife*, **9** (2020), e57309. https://doi.org/10.7554/eLife.57309
2. A. Marc, J. T. Schiffer, F. Mentré, A. S. Perelson, J. Guedj, Viral dynamic models during COVID-19: Are we ready for the next pandemic?, *CPT: Pharmacomet. Syst. Pharmacol.*, **14** (2025), 1289–1297. https://doi.org/10.1002/psp4.70055
3. O. Nave, U. Shemesh, I. HarTuv, Applying Laplace Adomian decomposition method (LADM) for solving a model of COVID-19, *Comput. Methods Biomech. Biomed. Eng.*, **24** (2021), 1618–1628. https://doi.org/10.1080/10255842.2021.1904399

4. A. Mateja, E. Chu, T. A. Murray, C. T. Bramante, C. Moser, N. Givens, et al., The choice of viral load end point in early phase trials of COVID-19 treatments aiming to reduce 28-day hospitalization and/or death, *J. Infect. Dis.*, **232** (2025), 60–68. https://doi.org/10.1093/infdis/jiaf282
5. D. M. Weinreich, S. Sivapalasingam, T. Norton, S. Ali, H. Gao, R. Bhore, et al., REGN-COV2, a neutralizing antibody cocktail, in outpatients with COVID-19, *N. Engl. J. Med.*, **384** (2021), 238–251. https://doi.org/10.1056/nejmoa2035002
6. A. S. Perelson, R. Ke, Mechanistic modeling of SARS-CoV-2 and other infectious diseases and the effects of therapeutics, *Clin. Pharmacol. Ther.*, **109** (2021), 829–840. https://doi.org/10.1002/cpt.2160
7. K. S. Kim, K. Ejima, S. Iwanami, Y. Fujita, H. Ohashi, Y. Koizumi, et al., A quantitative model used to compare within-host SARS-CoV-2, MERS-CoV, and SARS-CoV dynamics provides insights into the pathogenesis and treatment of SARS-CoV-2, *PLoS Biol.*, **19** (2021), e3001128. https://doi.org/10.1371/journal.pbio.3001128
8. R. Desikan, P. Padmanabhan, A. M. Kierzek, P. H. van der Graaf, Mechanistic models of COVID-19: Insights into disease progression, vaccines, and therapeutics, *Int. J. Antimicrob. Agents*, **60** (2022), 106606. https://doi.org/10.1016/j.ijantimicag.2022.106606
9. T. Phan, C. Zitzmann, K. W. Chew, D. M. Smith, E. S. Daar, D. A. Wohl, et al., Modeling the emergence of viral resistance for SARS-CoV-2 during treatment with an anti-spike monoclonal antibody, *PLoS Pathog.*, **20** (2024), e1011680. https://doi.org/10.1371/journal.ppat.1011680
10. E. Chigutsa, E. Jordie, M. Riggs, A. Nirula, A. Elmokadem, T. Knab, et al., A quantitative modeling and simulation framework to support candidate and dose selection of anti-SARS-CoV-2 monoclonal antibodies to advance bamlanivimab into a first-in-human clinical trial, *Clin. Pharmacol. Ther.*, **111** (2022), 595–604. https://doi.org/10.1002/cpt.2459
11. A. Goyal, E. F. Cardozo-Ojeda, J. T. Schiffer, Potency and timing of antiviral therapy as determinants of duration of SARS-CoV-2 shedding and intensity of inflammatory response, *Sci. Adv.*, **6** (2020), eabc7112. https://doi.org/10.1126/sciadv.abc7112
12. H. M. Dobrovolny, Quantifying the effect of remdesivir in rhesus macaques infected with SARS-CoV-2, *Virology*, **550** (2020), 61–69. https://doi.org/10.1016/j.virol.2020.07.015
13. K. Best, J. Guedj, V. Madelain, X. de Lamballerie, S. Y. Lim, C. E. Osuna, et al., Zika plasma viral dynamics in nonhuman primates provides insights into early infection and antiviral strategies, *Proc. Natl. Acad. Sci. U. S. A.*, **114** (2017), 8847–8852. https://doi.org/10.1073/pnas.1704011114
14. K. S. Kim, S. Iwanami, T. Oda, Y. Fujita, K. Kuba, T. Miyazaki, et al., Incomplete antiviral treatment may induce longer durations of viral shedding during SARS-CoV-2 infection, *Life Sci. Alliance*, **4** (2021), e202101049. https://doi.org/10.26508/lsa.202101049
15. R. Marlin, D. Desjardins, V. Contreras, G. Lingas, C. Solas, P. Roques, et al., Antiviral efficacy of favipiravir against zika and sars-cov-2 viruses in non-human primates, *Nat. Commun.*, **13** (2022), 5108. https://doi.org/10.1038/s41467-022-32565-w
16. J. Guedj, G. Piorkowski, F. Jacquot, V. Madelain, T. H. T. Nguyen, A. Rodallec, et al., Antiviral efficacy of favipiravir against ebola virus: A translational study in cynomolgus macaques, *PLoS Med.*, **15** (2018), e1002535. https://doi.org/10.1371/journal.pmed.1002535

17. *U.S. Food and Drug Administration*, Coronavirus (COVID-19) update: FDA authorizes monoclonal antibody treatment for COVID-19, 2020. Available from: https://www.fda.gov/drugs/drug-safety-and-availability/fda-authorizes-bamlanivimab-and-etesevimab-monoclonal-antibody-therapy-post-exposure-prophylaxis.

18. *U.S. Food and Drug Administration*, FDA authorizes REGEN-COV monoclonal antibody therapy for post-exposure prophylaxis (prevention) for COVID-19, 2021. Available from: https://www.fda.gov/drugs/drug-safety-and-availability/fda-authorizes-regen-cov-monoclonal-antibody-therapy-post-exposure-prophylaxis-prevention-covid-19.

19. P. Baccam, C. Beauchemin, C. A. Macken, F. G. Hayden, A. S. Perelson, Kinetics of influenza A virus infection in humans, *J. Virol.*, **80** (2006), 7590–7599. https://doi.org/10.1128/jvi.01623-05

20. A. S. Perelson, R. M. Ribeiro, T. Phan, An explanation for SARS-CoV-2 rebound after Paxlovid treatment, *MedRxiv*, 2023. https://doi.org/10.1101/2023.05.30.23290747

21. A. Gonçalves, J. Bertrand, R. Ke, E. Comets, X. De Lamballerie, D. Malvy, et al., Timing of antiviral treatment initiation is critical to reduce SARS-CoV-2 viral load, *CPT: Pharmacomet. Syst. Pharmacol.*, **9** (2020), 509–514. https://doi.org/10.1002/psp4.12543

22. P. Abuin, A. Anderson, A. Ferramosca, E. A. Hernandez-Vargas, A. H. Gonzalez, Characterization of SARS-CoV-2 dynamics in the host, *Annu. Rev. Control*, **50** (2020), 457–468. https://doi.org/10.1016/j.arcontrol.2020.09.008

23. M. Pérez, P. Abuin, M. Actis, A. Ferramosca, E. A. Hernandez-Vargas, A. H. González, *Optimal Control Strategies to Tailor Antivirals for Acute Infectious Diseases in the Host: a Study Case of COVID-19*, Elsevier, (2022), 11–39. https://doi.org/10.1016/b978-0-32-390171-0.00011-1

24. M. Perez, M. Actis, I. Sanchez, E. A. Hernandez-Vargas, A. H. González, A theory for viral rebound after antiviral treatment: A study case for SARS-CoV-2, *Math. Biosci.*, **379** (2025), 109339. https://doi.org/10.1016/j.mbs.2024.109339

25. J. E. Mittler, B. Sulzer, A. U. Neumann, A. S. Perelson, Influence of delayed viral production on viral dynamics in HIV-1 infected patients, *Math. Biosci.*, **152** (1998), 143–163. https://doi.org/10.1016/S0025-5564(98)10027-5

26. Y. Kakizoe, S. Nakaoka, C. A. Beauchemin, S. Morita, H. Mori, T. Igarashi, et al., A method to determine the duration of the eclipse phase for in vitro infection with a highly pathogenic SHIV strain, *Sci. Rep.*, **5** (2015), 10371. https://doi.org/10.1038/srep10371

27. O. Diekmann, J. A. P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*, John Wiley & Sons, 2000.

28. W. O. Kermack, A. G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. R. Soc. Lond. A*, **115** (1927), 700–721. https://doi.org/10.1098/rspa.1927.0118

29. S. Wiggins, *Normally Hyperbolic Invariant Manifolds in Dynamical Systems*, Springer New York, 1994. https://doi.org/10.1007/978-1-4612-4312-0

30. R. A. Cangelosi, E. J. Schwartz, D. J. Wollkind, A quasi-steady-state approximation to the basic target-cell-limited viral dynamics model with a non-cytopathic effect, *Front. Microbiol.*, **9** (2018), 54. https://doi.org/10.3389/fmicb.2018.00054

31. L. Shen, S. Peterson, A. R. Sedaghat, M. A. McMahon, M. Callender, H. Zhang, et al., Dose-response curve slope sets class-specific limits on inhibitory potential of anti-HIV drugs, *Nat. Med.*, **14** (2008), 762–766. https://doi.org/10.1038/nm1777

32. S. Esmaeili, K. Owens, J. Wagoner, S. J. Polyak, J. M. White, J. T. Schiffer, A unifying model to explain frequent SARS-CoV-2 rebound after nirmatrelvir treatment and limited prophylactic efficacy, *Nat. Commun.*, **15** (2024), 5478. https://doi.org/10.1038/s41467-024-49458-9

33. T. Aydillo, A. S. Gonzalez-Reiche, S. Aslam, A. van de Guchte, Z. Khan, A. Obla, et al., Shedding of viable SARS-CoV-2 after immunosuppressive therapy for cancer, *N. Engl. J. Med.*, **383** (2020), 2586–2588. https://doi.org/10.1056/NEJMc2031670

34. E. M. Scherer, A. Babiker, M. Engelbrecht-Zuebart, R. E. Geyer, H. Xie, S. J. Schrag, et al., SARS-CoV-2 evolution and immune escape in immunocompromised patients, *N. Engl. J. Med.*, **386** (2022), 2436–2438. https://doi.org/10.1056/NEJMc2202861

35. T. Phan, C. Zitzmann, K. W. Chew, D. M. Smith, E. S. Daar, D. A. Wohl, et al., Modeling suggests SARS-CoV-2 rebound after nirmatrelvir-ritonavir treatment is driven by target cell preservation coupled with incomplete viral clearance, *J. Virol.*, **99** (2025), e01623–24. https://doi.org/10.1128/jvi.01623-24

36. B. Killingley, A. J. Mann, M. Kalinova, A. Boyers, N. Goonawardane, J. Zhou, et al., Safety, tolerability and viral kinetics during SARS-CoV-2 human challenge in young adults, *Nat. Med.*, **28** (2022), 1031–1041. https://doi.org/10.1038/s41591-022-01780-9

37. S. A. Iyaniwura, A. Kunkel, P. Guo, J. B. Gutiérrez, S. Aris-Brosou, R. Ke, et al., The kinetics of SARS-CoV-2 infection based on a human challenge study, *Proc. Natl. Acad. Sci. U. S. A.*, **121** (2024), e2406303121. https://doi.org/10.1073/pnas.2406303121

38. R. Sachak-Patwa, E. I. Lafferty, C. J. Schmit, R. N. Thompson, H. M. Byrne, A target-cell limited model can reproduce influenza infection dynamics in hosts with differing immune responses, *J. Theor. Biol.*, **567** (2023), 111491. https://doi.org/10.1016/j.jtbi.2023.111491

39. N. M. Dixit, A. S. Perelson, Complex patterns of viral load decay under antiretroviral therapy: influence of pharmacokinetics and intracellular delay, *J. Theor. Biol.*, **226** (2004), 95–109. https://doi.org/10.1016/j.jtbi.2003.09.002

40. A. M. Smith, F. R. Adler, A. S. Perelson, An accurate two-phase approximate solution to an acute viral infection model, *J. Math. Biol.*, **60** (2009), 711–726. https://doi.org/10.1007/s00285-009-0281-8

41. S. Zhang, A. A. Agyeman, C. Hadjichrysanthou, J. F. Standing, SARS-CoV-2 viral dynamic modeling to inform model selection and timing and efficacy of antiviral therapy, *CPT: Pharmacomet. Syst. Pharmacol.*, **12** (2023), 1450–1460. https://doi.org/10.1002/psp4.13022

42. Y. Kuang, *Delay Differential Equations: with Applications in Population Dynamics*, Academic Press, 1993.

43. P. J. Hurtado, A. S. Kirosingh, Generalizations of the 'linear chain trick': incorporating more flexible dwell time distributions into mean field ODE models, *J. Math. Biol.*, **79** (2019), 1831–1883. https://doi.org/10.1007/s00285-019-01412-w

44. H. W. Hethcote, The mathematics of infectious diseases, *SIAM Rev.*, **42** (2000), 599–653. https://doi.org/10.1137/s0036144500371907

45. T. Phan, S. Brozak, B. Pell, A. Gitter, A. Xiao, K. D. Mena, et al., A simple SEIR-v model to estimate COVID-19 prevalence and predict SARS-CoV-2 transmission using wastewater-based surveillance data, *Sci. Total Environ.*, **857** (2023), 159326. https://doi.org/10.1016/j.scitotenv.2022.159326

46. T. Phan, S. Brozak, B. Pell, J. Oghuan, A. Gitter, T. Hu, et al., Making waves: Integrating wastewater surveillance with dynamic modeling to track and predict viral outbreaks, *Water Res.*, **243** (2023), 120372. https://doi.org/10.1016/j.watres.2023.120372

47. T. Phan, S. Brozak, B. Pell, S. M. Ciupe, R. Ke, R. M. Ribeiro, et al., Post-recovery viral shedding shapes wastewater-based epidemiological inferences, *Commun. Med.*, **5** (2025), 193. https://doi.org/10.1038/s43856-025-00908-5

Appendix

A. Proof of $\frac{\partial r}{\partial k_i} > 0$

For ease of notation, we assume $n = 1$ without loss of generality. Let $b = \beta p T_\infty k$ and $m = \min\{\delta, c\}$. Based on (3.5), we consider the following equation:

$$(k + x)(\delta + x)(c + x) = bk,$$

where $b < \delta c$ and $k > m$. We will show that on $(-m, 0)$ there exists a unique root, denoted by x^*, and $\frac{dx^*}{dk} < 0$.

Define the following:

$$f(x, k) = (k + x)(\delta + x)(c + x) - bk, \qquad g(x) = (\delta + x)(c + x) = x^2 + (\delta + c)x + \delta c.$$

Evaluate f at the endpoints of $(-m, 0)$ as follows:

$$f(-m, k) = -bk < 0, \qquad f(0, k) = k(\delta c - b) > 0,$$

since $b < \delta c$. By continuity, there exists at least one root in

$$x^*(k) \in (-m,\, 0).$$

For $x \in (-m, 0)$, we have $\delta + x > 0$, $c + x > 0$, hence $g(x) > 0$. Additionally, $k + x > k - m > 0$. Differentiate

$$\partial_x f(x, k) = g(x) + (k + x)g'(x), \qquad g'(x) = 2x + (\delta + c).$$

Since $g'(x) > 0$ on this interval, it follows that $\partial_x f(x, k) > 0$, so $f(\cdot, k)$ is strictly increasing. Thus, there is exactly one root $x^*(k)$ in the interval.

Differentiating implicitly,

$$\frac{dx^*}{dk} = -\frac{\partial_k f}{\partial_x f} = -\frac{g(x^*) - b}{g(x^*) + (k + x^*)g'(x^*)}.$$

At the root, we have $(k + x^*)g(x^*) = bk$, which implies the following:

$$g(x^*) = \frac{bk}{k + x^*} > b.$$

Hence, the numerator is positive, the denominator is positive, and therefore

$$\frac{dx^*}{dk} < 0.$$

Thus, $x^*(k)$ is strictly decreasing in k. Moreover, by the Implicit Function Theorem, $x^*(k)$ is C^1. Since $r = -x^*$, we have $\frac{dr}{dk} > 0$.

B. Model matching

We demonstrate that the modified model with a sigmoid infection term can replicate the viral load dynamics of the standard TCL model through parameter re-estimation. Consider the following system with a general infection function $f(T)$:

$$\frac{dT}{dt} = -\beta V f(T), \qquad \frac{dI_1}{dt} = \beta V T - kI_1, \qquad \frac{dI_2}{dt} = kI_1 - \delta I_2, \qquad \frac{dV}{dt} = pI_2 - cV. \tag{A.1}$$

The standard TCL model corresponds to $f(T) = T$, while the modified model uses $f(T) = S(T;\, x_0, \kappa)$, where S is the sigmoid function defined in (3.22), with T replaced by T/T_0 and scaled by T_0, thus ensuring $S(0;\, x_0, \kappa) = 0$ and $S(T_0;\, x_0, \kappa) = T_0$, where T_0 is an initial target cell count.

The reference viral load curve $V_{\text{std}}(t)$ is obtained by solving the standard model ($f(T) = T$) with the baseline parameter values from Table 1. The viral production rate p is held fixed, while the remaining kinetic parameters $\theta = (\beta, k, \delta, c)$ are re-estimated for the modified model by minimizing the mean squared error in $\log_{10}$ VL:

$$\theta^* = \arg\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \left[\log_{10} V_{\text{std}}(t_i) - \log_{10} V_{\text{mod}}(t_i;\, \theta)\right]^2,$$

where $V_{\text{mod}}(t_i;\, \theta)$ denotes the viral load of the modified model evaluated at $N = 300$ uniformly spaced time points $t_i \in [0, 15]$ days. Optimization is performed in $\log_{10}$-parameter space using the Nelder–Mead simplex algorithm (`fminsearch` in MATLAB), which is initialized at the standard model parameter values. Three configurations of the sigmoid parameters are examined: $(x_0, \kappa) \in \{(0.5, 2),\ (0.7, 2),\ (0.7, 4)\}$, spanning from a mild, nearly linear deviation to a steep threshold-like response.

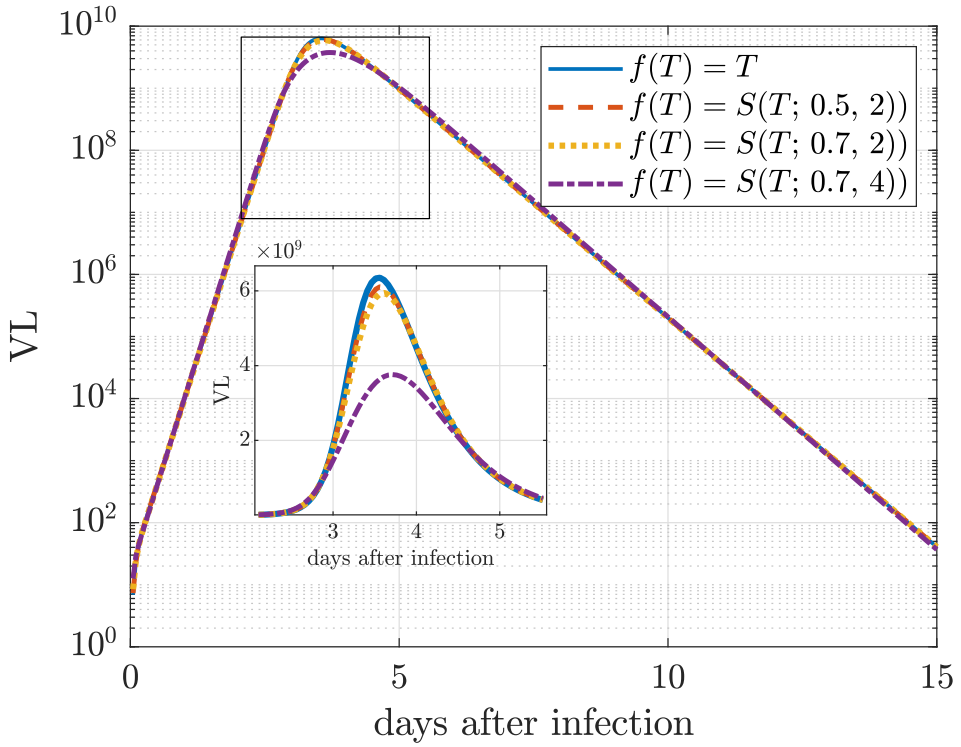


Figure B.1. Model matching results. The standard TCL model VL trajectory (solid) is compared with the re-estimated modified model trajectories for three choices of sigmoid parameters (x_0, κ). The modified model closely replicates the standard model's VL dynamics in all cases. The inset shows the same curves on a linear scale near the VL peak.

Although a visible mismatch near the VL peak appears on the linear scale (inset of Figure B.1), the fitting is performed in $\log_{10}$ VL, which is standard practice since VL data span many orders of magnitude and are routinely analyzed on a logarithmic scale [19, 20]. Additionally, we note that the sigmoid infection term is proposed here as a simple, proof-of-concept modification to illustrate that the non-monotonic behavior can be mitigated by a suitable choice of $f(T)$. It is not intended as a definitive model and should not be overstretched beyond this purpose.

C. Target cell dynamics

Figure C.1 shows the target cell dynamics that correspond to the VL simulations in Figure 1. The final target cell count T_∞ monotonically increases with the drug effect, which is consistent with $\frac{\partial T_\infty}{\partial \beta} < 0$ and $\frac{\partial T_\infty}{\partial p} < 0$ shown in (3.4). This is a final size result: stronger drug effects reduce the infection and preserve more target cells. However, target cell preservation alone does not explain the non-monotonic VL reduction rate. As shown in Section 3, the asymptotic reduction rate r depends on the product $\beta p T_\infty$ through the characteristic equation (3.5), and the sign of $\frac{\partial r}{\partial \beta}$ (or $\frac{\partial r}{\partial p}$) reflects a trade-off between decreasing β (or p) and increasing T_∞. Very high drug effects ($u \to 1$) also preserve the target cells, yet the VL rapidly declines because $\beta p T_\infty \to 0$. Thus, the non-monotonicity arises from the interplay of these competing factors, not from target cell preservation per se.

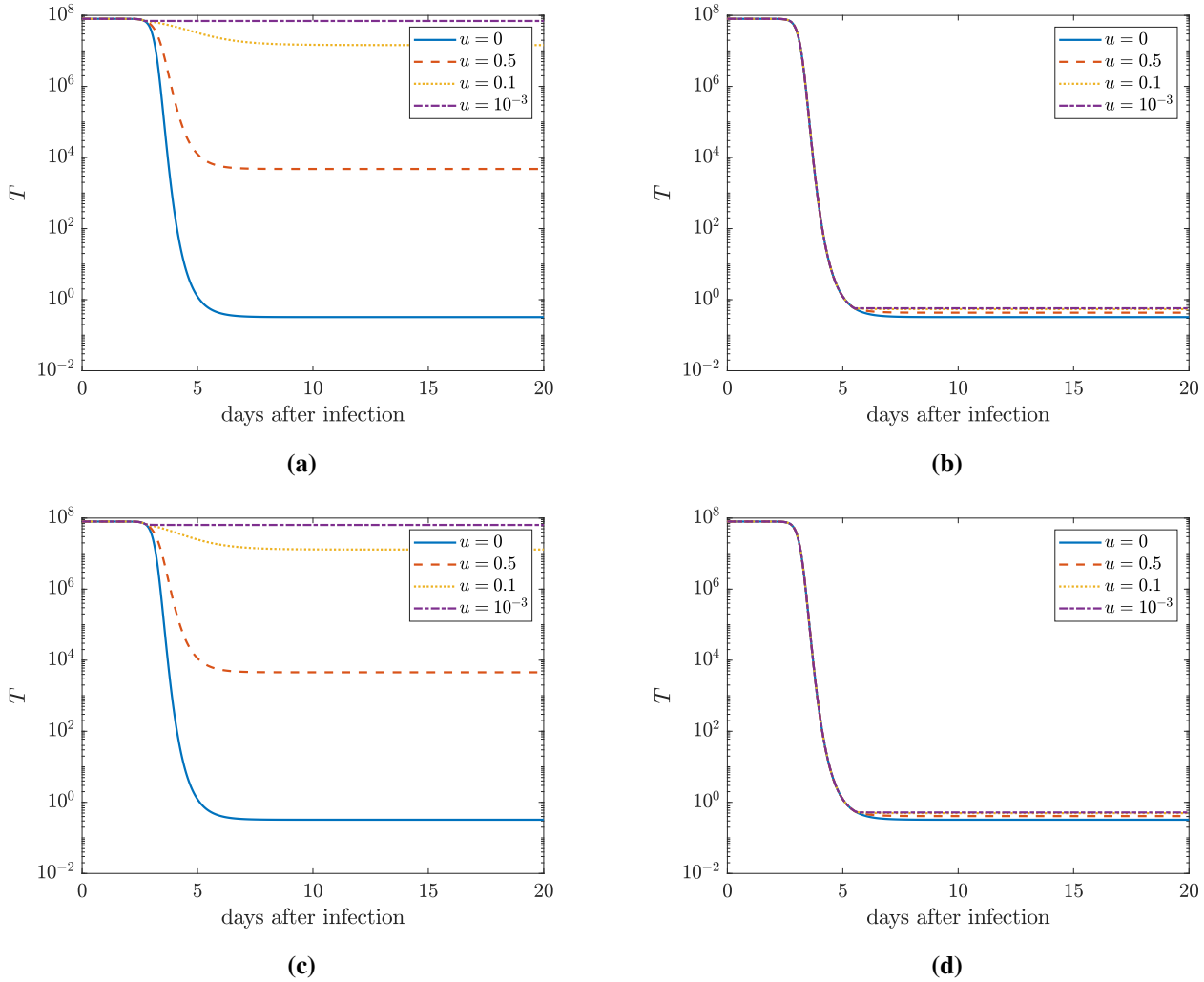


Figure C.1. Target cell dynamics corresponding to the VL simulations in Figure 1. (a) Drug effects on β started at 2.7 days. (b) Drug effects on β started at 5.5 days. (c) Drug effects on p started at 2.7 days. (d) Drug effects on p started at 5.5 days. The final target cell count T_∞ monotonically increases with drug effect strength.