Mathematical Biosciences and Engineering

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

MBE, 21(4): 5577–5603.
DOI: 10.3934/mbe.2024246
Received: 18 January 2024
Revised: 17 March 2024
Accepted: 29 March 2024
Published: 09 April 2024

Research article

A novel within-host model of HIV and nutrition

Archana N. Timsina[1], Yuganthi R. Liyanage[2], Maia Martcheva[3] and Necibe Tuncer[2,*]

[1] Department of Population Health and Pathobiology, North Carolina State University, Raleigh 27607, USA

[2] Department of Mathematical Sciences, Florida Atlantic University, Boca Raton 33431, USA

[3] Department of Mathematics, University of Florida, Gainesville 32611, USA

* **Correspondence:** Email: ntuncer@fau.edu.

Abstract: In this paper we develop a four compartment within-host model of nutrition and HIV. We show that the model has two equilibria: an infection-free equilibrium and infection equilibrium. The infection free equilibrium is locally asymptotically stable when the basic reproduction number $\mathcal{R}_0 < 1$, and unstable when $\mathcal{R}_0 > 1$. The infection equilibrium is locally asymptotically stable if $\mathcal{R}_0 > 1$ and an additional condition holds. We show that the within-host model of HIV and nutrition is structured to reveal its parameters from the observations of viral load, CD4 cell count and total protein data. We then estimate the model parameters for these 3 data sets. We have also studied the practical identifiability of the model parameters by performing Monte Carlo simulations, and found that the rate of clearance of the virus by immunoglobulins is practically unidentifiable, and that the rest of the model parameters are only weakly identifiable given the experimental data. Furthermore, we have studied how the data frequency impacts the practical identifiability of model parameters.

Keywords: HIV; within-host model; nutrition; structural identifiability; practical identifiability

1. Introduction

Human immunodeficiency virus (HIV) targets CD4 T cells, which are a type of white blood cell that plays a key role in the immune system. Once HIV enters the body, it binds to the surface of CD4 T cells, releases RNA into the host cell, and then replicates. The newly produced viral particles infect other CD4 cells and HIV spreads throughout the body [1]. HIV takes over the CD4 cells to reproduce itself, while the infected CD4 cells activate the immune system to fight off the virus [2, 3]. Mathematical modeling has been a useful tool to facilitate understanding of the HIV and CD4 cell dynamics. It provides a framework for modeling the virus interaction with the immune system, the reproduction of the virus and the decline in CD4 cells. So far, various deterministic mathematical models have been

developed, and the majority of these studies are focused on the interaction among HIV, the healthy cells and the infected CD4 T cells. More elaborate studies introduced the immune response. Among those, Nelson and Perelson in [4] have presented a model to show how some viruses may fail to produce an effective immune response. Further, Perelson et al, in [5], developed a dynamical model to study the depletion of CD4 cells by the HIV infection without considering the immune response. Again, Perelson and Nelson in [1] have used a modeling and parameter estimation technique to uncover the features of HIV pathogenesis and the impact of antiretroviral drug treatment. Moreover, Rong et al, in [6], presented a model to study the existence of latent reservoirs that prevent virus eradication. Over the past few decades several mathematical models have been developed to understand HIV within the human body, its pathogenesis with CD4 cells, and the nature of the virus, they have also been applied to investigate the use of medications for possible controls [1–3, 5–9]. The interplay between HIV and nutrition has not been investigated from the perspective of differential equation models, but some statistical studies, such as [10–14], indicate that the mean value of serum protein levels of HIV-positive individuals is higher than that of healthy individuals. Based on these observations, we have modeled the dynamics of the virus, the CD4 cells, and the protein in the body.

The novelty of our model is that it focuses precisely on this process and no mathematical model has been developed so far to study the interaction among CD4 cells (T), infected target cells (T_i), viral load (V), and protein (P). We have developed and analyzed a deterministic four compartment nutrition and HIV within-host model to understand the HIV and protein dynamics. One important observation that has helped in the development of the model presented here is that the protein level of the HIV infected individuals was frequently found to be significantly higher than that of normal healthy individuals [10, 12–14]. While the specific reasons for that are not known, we surmise that the need for immunoglobulins (antibodies) is higher which directs some of the ingested protein into that venue.

Our main goal with this study was to determine reliable parameter values that can be used in future studies. We fit the model to data from a multi-site study in Africa and Southeast Asia that was designed to track the HIV status to characterize the progression of acute HIV-1 infection [15, 16]. While fitting differential equation models to data, it is important to study models' identifiability. Identifiability analysis is a process that investigates whether the value of an unknown model parameter can be uniquely determined from a given set of observations [17–19]. There are two types of identifiability: *structural identifiability* and *practical identifiability*. Structural identifiability determines whether the parameters of the within-host model can be uniquely identified from the noise-free observations without the actual data [20].

Structural identifiability is becoming increasingly important in model formulation and is studied more frequently [17, 21–23]. It determines whether the model is structured such that its parameters can be determined from the given data. After structural identifiability, practical identifiability of the estimated parameters needs to be established. The practical identifiability determines whether the estimated parameters are identifiable when the data are noisy [17–19]. Lack of practical identifiability may still subject our predictions to significant errors [24]. Practical identifiability is typically determined by performing extensive simulations [18, 25, 26] using synthetic data with various levels of noise. Synthetic data are usually sampled with the frequency of the real data but assessing the identifiability with various frequencies tells us whether the identifiability can be improved with more experiments.

This paper is focused on a nutrition and HIV within-host model, its equilibrium points and their stability, identifiability, and estimation of the values of the parameters. Section 2 describes details about the derivation of a four compartment nutrition and HIV within-host model. Section 3 studies the

equilibria, and presents the derivation of the reproduction number, as well as the stability of the equilibrium points. Section 4 discusses the identifiability of the parameters, and shows the estimation method for the parameters employing real-life data. Finally, we discuss the results of this paper in Section 5.

2. Within-host model of HIV and nutrition

We model the interplay between HIV and nutrition by including the total protein to the well-studied target cell limited model [6, 8]. The target cell limited model is given by the following equations:

$$\begin{aligned} \frac{dT}{dt} &= r - \beta VT - dT, \\ \frac{dT_i}{dt} &= \beta VT - \delta T_i, \\ \frac{dV}{dt} &= \pi T_i - cV. \end{aligned} \tag{2.1}$$

In this model, $T(t)$, $T_i(t)$, and $V(t)$ represent the number of target cells, namely, CD4 cells, infected target cells, and the viral load, respectively. CD4 cells are produced at a constant rate r and cleared at a rate d. Target cells become infected upon contact with the HIV viral particles at a rate β. The mass-action term βVT states that the average number of new infected cells is proportional to the product of target cells and the viral particles. The infected cells are cleared at a rate δ. The elimination rate for viral particles is represented by c; hence, cV represents the number of viral particles eliminated per unit of time. In the model, replication of V is described by πT_i, where π is the number of new viral particles produced by an infected CD4 cell.

Total protein levels in HIV-infected people are higher [10–14]. To model the interaction between HIV and nutrition, we extend the within-host model, Eq (2.1), by adding a fourth compartment, $P(t)$, which is the total amount of protein at time t. The novel within-host model of HIV and nutrition can be described by the following equations:

$$\begin{aligned} \frac{dT}{dt} &= r - \frac{\beta}{1 + c_1 P} VT - dT, \\ \frac{dT_i}{dt} &= \frac{\beta}{1 + c_1 P} VT - \delta T_i, \\ \frac{dV}{dt} &= \pi T_i - cV - c_2 PV, \\ \frac{dP}{dt} &= \lambda + \gamma PV - \mu P. \end{aligned} \tag{2.2}$$

Here, the first equation of the system given by Eq (2.2) describes the dynamics of the uninfected CD4 cells population, where r is the recruitment rate of CD4 cells and d is their natural death rate. We chose to incorporate the impact of nutrition on the susceptibility of target cells and model the infection of target cells by HIV with the term $\frac{\beta}{1+c_1 P} VT$. An infected cell bursts virus particles at a rate π and dies at a rate δ. The term $c_2 PV$ denotes the viral clearance by the immunoglobulins. The last equation models the dynamics of the protein, where λ is the daily protein intake, μ is the total protein use and γ is the virus-driven enhancement of the total protein. The well established HIV within-host model (2.1), explores the dynamics of target cells, infected cells and the HIV. We have expanded this model to include the protein dynamics, introducing a novel model (2.2) that integrates HIV and nutrition. However, it is worth

mentioning that the within-host model (2.1) is not a sub-model of (2.2) unless $\lambda = 0$. Specifically, in the within-host model (2.2), the protein level never reaches to zero; therefore, $P = 0$ is not a solution for the model (2.2). The definitions of the state variables and parameters are summarized in Tables 1 and 2 respectively, and the model is presented schematically in Figure 1.

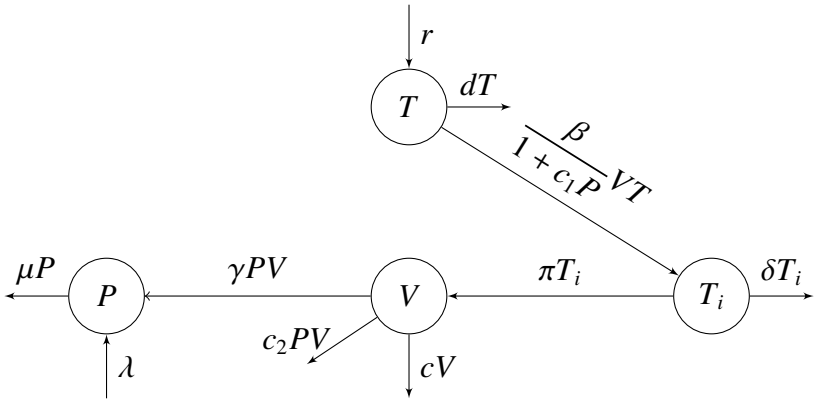

Figure 1. Flow chart of the within-host model Eq (2.2).

Table 1. State variables of the HIV within-host model Eq (2.2), and their definitions.

Variables	Definition
$T(t)$	Number of target cells at time t
$T_i(t)$	Number of infected cells at time t
$V(t)$	Number of viral particles at time t
$P(t)$	Protein at time t

Table 2. Definitions of the parameters of the HIV within-host model Eq (2.2).

Parameters	Definition
r	Recruitment rate of target cells
β	Target cells infection rate
c_1	Reduction of the target cells' susceptibility in the presence of protein
d	Death rate of uninfected target cells
δ	Death rate of infected cells
π	Production rate of viral particles owing to infected cells
c	Virus elimination rate
c_2	Rate of clearance of virus by the immunoglobulins
λ	Protein intake
μ	Clearance rate of protein
γ	Virus-driven protein enhancement rate

3. Equilibria and stability

We analyze the model Eq (2.2) by determining its equilibrium points. At the equilibria, the model Eq (2.2) satisfies the following set of equations:

$$\begin{aligned} 0 &= r - \frac{\beta}{1 + c_1 P^*} V^* T^* - dT^*, \\ 0 &= \frac{\beta}{1 + c_1 P^*} V^* T^* - \delta T_i^*, \\ 0 &= \pi T_i^* - cV^* - c_2 P^* V^*, \\ 0 &= \lambda + \gamma P^* V^* - \mu P^*, \end{aligned} \tag{3.1}$$

where T^*, V^*, T_i^* and P^* denote the equilibrium values of target cells, the virus, infected target cells, and protein, respectively.

3.1. Stability analysis of infection-free equilibrium (IFE)

The infection-free equilibrium of the system occurs when no virus is present. In the absence of the virus, $V^* = 0$ and $T_i^* = 0$. We obtain the infection free equilibrium by solving for T^* and P^* in Eq (3.1). The infection free equilibrium is given by $E_0 = \left(\frac{r}{d}, 0, 0, \frac{\lambda}{\mu}\right)$.

Theorem 3.1. *The infection free equilibrium E_0 of the system Eq* (2.2) *is locally asymptotically stable, if $\mathcal{R}_0 < 1$, and unstable if $\mathcal{R}_0 > 1$, where*

$$\mathcal{R}_0 = \frac{r\pi\beta\mu^2}{\delta d(\mu + c_1\lambda)(c\mu + \lambda c_2)}.$$

Proof. The Jacobian of the system (2.2) at E_0 is given by

$$J(E_0) = \begin{bmatrix} -d & 0 & \frac{-\beta r\mu}{d(\mu + c_1\lambda)} & 0 \\ 0 & -\delta & \frac{\beta r\mu}{d(\mu + c_1\lambda)} & 0 \\ 0 & \pi & -c - c_2\frac{\lambda}{\mu} & 0 \\ 0 & 0 & \frac{\gamma\lambda}{\mu} & -\mu \end{bmatrix}$$

Clearly, $-d$ and $-\mu$ are eigenvalues of the matrix J and the remaining eigenvalues are given by the eigenvalues of the following matrix J_1:

$$J_1 = \begin{bmatrix} -\delta & \frac{\beta r\mu}{d(\mu + c_1\lambda)} \\ \pi & -c - c_2\frac{\lambda}{\mu} \end{bmatrix}$$

The characteristic equation of J_1 is $p(\Lambda) = \Lambda^2 + a_1\Lambda + a_2 = 0$, where

$$a_1 = \delta + c + \frac{c_2\lambda}{\mu} \quad \text{and} \quad a_2 = \delta(c + \frac{c_2\lambda}{\mu}) - \frac{\beta r\mu\pi}{d(\mu + c_1\lambda)}$$

The eigenvalues of J_1 will be negative or have a negative real part if $TrJ_1 < 0$ and $DetJ_1 > 0$. Clearly, the trace of J_1 is negative, i.e., $TrJ_1 < 0$. Setting the determinant to be positive gives the reproduction number in the following form:

$$\mathcal{R}_0 = \frac{r\pi\beta\mu^2}{\delta d(\mu + c_1\lambda)(c\mu + \lambda c_2)}.$$

If $\mathcal{R}_0 < 1$ then $TrJ_1 < 0$ and $DetJ_1 > 0$. Therefore, the infection free equilibrium E_0 of the system given by Eq (2.2) is locally asymptotically stable. Now, if $\mathcal{R}_0 > 1$ then $a_2 < 0$ and $p(\Lambda)$ has a positive real root. Therefore, the infection-free equilibrium is unstable if $\mathcal{R}_0 > 1$. □

3.2. Existence and local stability of infection equilibrium (IE)

At the infection equilibrium, $E_1 = (T^*, T_i^*, V^*, P^*)$, each state variable has a positive value. To obtain the infection equilibrium E_1, we first solve for P^* by using the forth equation of Eq (3.1) and obtain,

$$P^* = \frac{\lambda}{\mu - \gamma V^*}. \tag{3.2}$$

We are interested in a positive infection equilibrium; thus, each state variable must have a positive value. Therefore, by setting

$$P^* = \frac{\lambda}{\mu - \gamma V^*} > 0 \implies 0 < V^* < \frac{\mu}{\gamma}.$$

From the third equation in Eq (3.1), and by substituting Eq (3.2) for P^*, we obtain,

$$T_i^* = \frac{c}{\pi}V^* + \frac{c_2\lambda}{\pi(\mu - \gamma V^*)}V^*. \tag{3.3}$$

Since $0 < V^* < \frac{\mu}{\gamma}$, we have that $T_i^* > 0$. Next, we add the first two equations in Eq (3.1), and then solve for T^*. Using Eq (3.3), we obtain

$$T^* = \frac{r}{d} - \frac{\delta c}{d\pi}V^* - \frac{\delta c_2\lambda}{d\pi(\mu - \gamma V^*)}V^*.$$

In order to get a positive solution for T^* at the infection equilibrium, V^* must satisfy the condition of the following inequality:

$$\frac{\lambda c_2}{\pi}V^* < (\mu - \gamma V^*)\left(\frac{r}{\delta} - \frac{c}{\pi}V^*\right). \tag{3.4}$$

The left-hand side of Eq (3.4) is a straight line, which goes through the origin at a slope of $\frac{\lambda c_2}{\pi}$. The right hand side of the inequality is a parabola of V^* of roots $V_1^* = \frac{\mu}{\gamma}$ and $V_2^* = \frac{r\pi}{\delta c}$. Clearly, the parabola opens up and then the line meets the parabola at two distinct positive points, say V_1, V_2. Let $V_1 < V_2$; then, $V_1 < V_1^*$ and $V_1 < V_2^*$ for any V_1^*, V_2^*. Thus,

$$V_1 < \frac{\mu}{\gamma}.$$

We further assume that $V^* < V_1$ for $T^* > 0$. This satisfies our requirement to have positive values at the infection equilibrium. Next, from the second equation of Eq (3.1), we have that $\frac{\beta}{1+c_1P^*}V^*T^* = \delta T_i^*$. Substituting P^*, T^*, and T_i^* into this equation yields,

$$\frac{\beta(\mu-\gamma V^*)}{\mu-\gamma V^*+c_1\lambda}\left(\frac{r}{d}-\frac{\delta}{d}\left(\frac{c}{\pi}V^*+\frac{c_2\lambda}{\pi(\mu-\gamma V^*)}V^*\right)\right)=\frac{\delta}{\pi}\left(c+\frac{c_2\lambda}{\mu-\gamma V^*}\right). \tag{3.5}$$

Let $y_1(V^*) = \frac{\beta(\mu-\gamma V^*)}{\mu-\gamma V^*+c_1\lambda}\left(\frac{r}{d}-\frac{\delta}{d}\left(\frac{c}{\pi}V^*+\frac{c_2\lambda}{\pi(\mu-\gamma V^*)}V^*\right)\right)$ and $y_2(V^*) = \frac{\delta}{\pi}\left(c+\frac{c_2\lambda}{\mu-\gamma V^*}\right)$. Here, y_1 is a decreasing function since $y_1' < 0$ for $V^* < V_1 < \frac{\mu}{\gamma}$ and y_2 is an increasing function since $y_2' > 0$ for $V^* < V_1 < \frac{\mu}{\gamma}$. Thus, if there exists a solution, then it is unique. We continue with proving the existence of the solution. Since $0 < V^* < V_1 < \frac{\mu}{\gamma}$, we set $V^* = 0$ in $y_1(V^*)$ and $y_2(V^*)$. Then,

$$y_1(0) = \frac{\beta\mu r}{d(\mu+c_1\lambda)} \quad \text{and} \quad y_2(0) = \frac{\delta}{\pi}\left(c+\frac{c_2\lambda}{\mu}\right).$$

Since $\mathcal{R}_0 > 1$,

$$\mathcal{R}_0 = \frac{r\pi\beta\mu^2}{\delta d(\mu+c_1\lambda)(c\mu+\lambda c_2)} > 1 \quad \Longrightarrow \quad \frac{r\beta\mu}{d(\mu+c_1\lambda)} > \frac{\delta(c\mu+\lambda c_2)}{\pi\mu}.$$

Thus, $y_1(0) > y_2(0)$. Similarly, set $V^* = V_1$ in $y_1(V^*)$ and $y_2(V^*)$; then, we have that $y_1(V_1) = 0$ and $y_2(V_1) > 0$. Thus, $y_1(V_1) < y_2(V_1)$. Since y_1 is decreasing and y_2 is increasing for $0 < V^* < V_1 < \frac{\mu}{\gamma}$ and $y_1(0) > y_2(0)$ when $V^* = 0$ and $y_1(V_1) < y_2(V_1)$ when $V^* = V_1$, then $y_1(V^*)$ meets $y_2(V^*)$ at one point $V^* \in (0, V_1)$. When $\mathcal{R}_0 > 1$, Eq (3.5) has a unique root V^* in the interval $(0, V_1)$. This implies that Eq (3.5) has a unique positive solution $V^* < V_1 < \frac{\gamma}{\mu}$. Correspondingly, model Eq (2.2) has a unique infection equilibrium $E_1 = (T^*, T_i^*, V^*, P^*)$. Because $V^* \in (0, V_1)$, all components of this equilibrium are positive. We have proved the following Theorem 3.2.

Theorem 3.2. *There exists a unique infection equilibrium E_1 of the system Eq* (2.2) *when $\mathcal{R}_0 > 1$.*

Theorem 3.3. *Assume that $\mathcal{R}_0 > 1$ and*

$$\mu > \gamma V^*\left(1+\frac{c_1P^*}{1+c_1P^*}\right)$$

Then the infection equilibrium E_1 of the system given by Eq (2.2) *is locally asymptotically stable.*

Proof. The Jacobian of the system Eq (2.2) at E_1 is given by

$$J(E_1) = \begin{bmatrix} \frac{-\beta}{1+c_1P^*}V^*-d & 0 & \frac{-\beta}{1+c_1P^*}T^* & \frac{\beta c_1}{(1+c_1P^*)^2}V^*T^* \\ \frac{\beta}{1+c_1P^*}V^* & -\delta & \frac{\beta}{1+c_1P^*}T^* & \frac{-\beta c_1}{(1+c_1P^*)^2}V^*T^* \\ 0 & \pi & -c-c_2P^* & -c_2V^* \\ 0 & 0 & \gamma P^* & \gamma V^*-\mu \end{bmatrix}.$$

Then, the characteristic equation at E_1 is $det|J - kI| = 0$, where k is an eigenvalue. Thus, we have the following:

$$\begin{vmatrix} \frac{-\beta}{1+c_1P^*}V^* - d - k & 0 & \frac{-\beta}{1+c_1P^*}T^* & \frac{\beta c_1}{(1+c_1P^*)^2}V^*T^* \\ \frac{\beta}{1+c_1P^*}V^* & -\delta - k & \frac{\beta}{1+c_1P^*}T^* & \frac{-\beta c_1}{(1+c_1P^*)^2}V^*T^* \\ 0 & \pi & -c - c_2P^* - k & -c_2V^* \\ 0 & 0 & \gamma P^* & \gamma V^* - \mu - k \end{vmatrix} = 0. \tag{3.6}$$

Adding the first row to the second row of Eq (3.6), we have the following:

$$\begin{vmatrix} \frac{-\beta}{1+c_1P^*}V^* - d - k & 0 & \frac{-\beta}{1+c_1P^*}T^* & \frac{\beta c_1}{(1+c_1P^*)^2}V^*T^* \\ -d - k & -\delta - k & 0 & 0 \\ 0 & \pi & -c - c_2P^* - k & -c_2P^* \\ 0 & 0 & \gamma P^* & \gamma V^* - \mu - k \end{vmatrix} = 0. \tag{3.7}$$

Expanding the determinant by the last row, Eq (3.7) can be simplified to

$$\begin{aligned} 0 = {} & (\gamma V^* - \mu - k)(c + c_2P^* + k)(\delta + k)\left(\frac{-\beta}{1+c_1P^*}V^* - d - k\right) + \\ & (\gamma V^* - \mu - k)(d + k)\left(\frac{\beta\pi}{1+c_1P^*}T^*\right) + \\ & \gamma P^* c_2 V^* (\delta + k)\left(\frac{\beta V^*}{1+c_1P^*} + d + k\right) + \gamma P^*(d + k)\frac{\beta c_1 \pi V^* T^*}{(1+c_1P^*)^2}. \end{aligned}$$

This implies that

$$\frac{k + d + \frac{\beta}{1+c_1P^*}V^*}{k + d} = \frac{\left(k + \mu - \gamma V^*\left(1 + \frac{c_1P^*}{1+c_1P^*}\right)\right)\frac{\beta\pi}{1+c_1P^*}T^*}{(\delta + k)(k + \mu - \gamma V^*)\left(k + c + c_2P^* + \frac{\gamma c_2 P^* V^*}{k + \mu - \gamma V^*}\right)}. \tag{3.8}$$

Here, we assume that $\mu > \gamma V^*\left(1 + \frac{c_1P^*}{1+c_1P^*}\right)$.

Clearly, the left-hand side of Eq (3.8) is greater than one, since

$$\left|\frac{k + d + \frac{\beta}{1+c_1P^*}V^*}{k + d}\right| = \left|1 + \frac{\beta}{(1+c_1P^*)(k+d)}V^*\right| > 1.$$

To prove that the infection equilibrium is locally asymptotically stable, we need to show that all eigenvalues of the Jacobian have a negative real part. Using the notion of contradiction, we suppose that

k has a non-negative real part, i.e., $Re(k) \geq 0$. Then, from our assumption,

$$\left|k + \mu - \gamma V^* \left(1 + \frac{c_1 P^*}{1 + c_1 P^*}\right)\right| < |k + \mu - \gamma V^*|.$$

Thus,

$$\left|\frac{k + \mu - \gamma V^* \left(1 + \frac{c_1 P^*}{1 + c_1 P^*}\right)}{k + \mu - \gamma V^*}\right| < 1. \tag{3.9}$$

Let $k = x + iy$. Then,

$$\left|k + c + c_2 P^* + \frac{\gamma c_2 P^* V^*}{k + \mu - \gamma V^*}\right| = \left|x + iy + c + c_2 P^* + \frac{\gamma c_2 P^* V^*}{x + iy + \mu - \gamma V^*}\right|.$$

Multiplying by the complex conjugate, we obtain

$$\begin{aligned} &= \left|x + c + c_2 P^* + \frac{\gamma c_2 P^* V^* (x + \mu - \gamma V^*)}{(x + \mu - \gamma V^*)^2 + y^2} + iy\left(1 - \frac{\gamma c_2 P^* V^*}{(x + \mu - \gamma V^*)^2 + y^2}\right)\right| \\ &= \sqrt{\left(x + c + c_2 P^* + \frac{\gamma P^* V^* c_2 (x + \mu - \gamma V^*)}{(x + \mu - \gamma V^*)^2 + y^2}\right)^2 + y^2 \left(1 - \frac{\gamma c_2 P^* V^*}{(x + \mu - \gamma V^*)^2 + y^2}\right)^2} \\ &\geq c + c_2 P^*. \end{aligned} \tag{3.10}$$

Therefore, by inequality (3.10), it follows that

$$\left|\frac{\frac{\beta \pi T^*}{1 + c_1 P^*}}{(\delta + k)\left(k + c + c_2 P^* + \frac{\gamma c_2 P^* V^*}{k + \mu - \gamma V^*}\right)}\right| \leq \frac{\beta \pi T^*}{\delta (c + c_2 P^*)(1 + c_1 P^*)}.$$

From the second equation of Eq (3.1), we have that $\frac{\beta T^*}{1 + c_1 P^*} = \frac{\delta T_i}{V^*}$. Substituting this expression into the above inequality, and since $\frac{\pi T_i}{V^*(c + c_2 P^*)} = 1$ by the third equation of Eq (3.1), we obtain

$$\left|\frac{\frac{\beta \pi T^*}{1 + c_1 P^*}}{(\delta + k)\left(k + c + c_2 P^* + \frac{\gamma c_2 P^* V^*}{k + \mu - \gamma V^*}\right)}\right| \leq \frac{\pi T_i}{V^*(c + c_2 P^*)} \leq 1. \tag{3.11}$$

Thus, by Eqs (3.9) and (3.11), the right-hand side of the Eq (3.8) is less than 1. This contradicts that the left-hand side of the Eq (3.8) is greater than 1. Therefore, all of the eigenvalues have negative real parts, and, hence, the infection equilibrium is locally asymptotically stable. □

Remark 3.4. *If $\mu \leq \gamma V^* \left(1 + \frac{c_1 P^*}{1 + c_1 P^*}\right)$, then it is not hard to see that Eq* (3.8) *does not have real nonnegative solutions; however, it may have complex solutions with a positive real part. If this case occurs, then Hopf bifurcation occurs and equilibrium E_1 is unstable but there is an oscillatory solution that is stable and attracts the solutions.*

4. Structural and practical identifiability analysis

Identifiability analysis investigates whether a given set of observations can uniquely determine the model's parameters. There are two stages of identifiability analysis: *structural identifiability* and *practical identifiability*. Structural identifiability assesses whether the within-host model parameters can be uniquely determined from noise-free observations. This type of analysis relies on the relationship between the model and the observations and is performed prior to estimating model parameters based on any actual experimental data. Only if a model is structurally identifiable, then we can estimate the parameters of the model based on the actual data. Practical identifiability analysis investigates whether the within-host model parameters can be determined from experimental data that are contaminated with noise. In this section, we use the differential algebra method to examine structural identifiability and Monte Carlo simulations (MCSs) to assess practical identifiability [27–32].

4.1. Structural identifiability

We first analyze the structural identifiability of the within-host model. We rewrite the model Eq (2.2) in the following compact form:

$$x'(t) = f(x, p),$$

where t represents time, the vector $x = (T, T_i, V, P)$ denotes the state variables, and $p = (r, \beta, c_1, d, \delta, \pi, c, c_2, \lambda, \gamma, \mu)$ is the parameter vector. We denote the model output by $y(t)$, which represents the observations. The observations are the number of target cells $T(t)$, represented by $y_1(t)$, and the viral load $V(t)$, represented by $y_2(t)$. Then, the observations can be written as:

$$y_1(t) = T(t) \quad \text{and} \quad y_2(t) = V(t).$$

The goal of structural identifiability is to determine whether the model is structured to uniquely yield its parameters, denoted by p, based on observations $y_1(t)$ and $y_2(t)$. If so, we say that the within-host model is structurally identifiable. First, we give the definition of structural identifiability in terms of the observed variables $y_1(t)$ and $y_2(t)$ [17, 20].

Definition 1. Let p and $\hat{p}$ be the two distinct parameter vectors of the within-host model Eq (2.2). We say that the within-host model is structurally identifiable if and only if

$$y_1(t, p) = y_1(t, \hat{p}) \quad \text{and} \quad y_2(t, p) = y_2(t, \hat{p}) \implies p = \hat{p}.$$

Several methods have been proposed to analyze the structural identifiability of mathematical models [17, 20, 33]. Here, we use a technique called differential algebra method [20]. The method allows the elimination of the unobserved state variables and obtains the model as a function of the model parameters and the observed model state variables, which are called the input-output equations. The input-output equations are differential-algebraic polynomials of the outputs $y_1(t)$ and $y_2(t)$, and coefficients consist of model parameters. We can obtain the following input-output equations of the within-host model Eq (2.2) with the observations $y_1(t)$ and $y_2(t)$ using differential algebra for identifiability of systems (DAISY) software [34].

$$
\begin{aligned}
0 = {} & T''T^2V^5 - T'^2TV^5\frac{(\beta-\gamma)}{\beta} - T'^2TV^4\frac{(c_1\lambda+\mu)}{\beta} - T'V'T^2V^4 + T'T^2V^6\gamma - \\
& T'T^2V^5\frac{(\beta\mu-2d\gamma)}{\beta} - 2T'T^2V^4\frac{d(c_1\lambda+\mu)}{\beta} - T'TV^5\frac{r(-\beta+2\gamma)}{\beta} + 2T'TV^4\frac{r(c_1\lambda+\mu)}{\beta} - \\
& V'T^3V^4d + V'T^2V^4r + T^3V^6d\gamma - T^3V^5\frac{d(\beta\mu-d\gamma)}{\beta} - T^3V^4\frac{d^2(c_1\lambda+\mu)}{\beta} - T^2V^6\gamma r - \\
& T^2V^5\frac{r(-\beta\mu+2d\gamma)}{\beta} + 2T^2V^4dr(c_1\lambda+\mu)/\beta + TV^5\frac{\gamma r^2}{\beta} - TV^4\frac{r^2(c_1\lambda+\mu)}{\beta}.
\end{aligned} \tag{4.1}
$$

$$
\begin{aligned}
0 = {} & T'^2TV^3 + \frac{T'V''TV^3}{\pi} - T'V'TV^3\frac{(cc_1 - c_1\delta + c_2)}{c_1\pi} + 2T'T^2V^3d - T'TV^5\frac{c_2\gamma}{c_1\pi} - \\
& T'TV^4\frac{(-cc_1\delta - c_1c_2\lambda + c_2\delta - c_2\mu)}{c_1\pi} - 2T'TV^3r + V''T^2V^3\frac{d}{\pi} - V''TV^3\frac{r}{\pi} - V'T^2V^4\frac{\beta c_2}{c_1\pi} - \\
& V'T^2V^3\frac{d(-cc_1 - c_1\delta + c_2)}{c_1\pi} - V'TV^3\frac{r(cc_1 + c1\delta - c_2)}{c_1\pi} + T^3V^3d^2 - T^2V^6\frac{\beta c_2\gamma}{c_1\pi} - \\
& T^2V^5\frac{c_2(\beta\delta - \beta\mu + d\gamma)}{c_1\pi} - T^2V^4\frac{d(-cc_1\delta - c_1c_2\lambda + c_2\delta - c_2\mu)}{c_1\pi} - 2T^2V^3dr + TV^5\frac{c_2\gamma r}{c_1\pi} - \\
& TV^4\frac{r(cc_1\delta + c_1c_2\lambda - c_2\delta + c_2\mu}{c_1\pi} + TV^3r^2.
\end{aligned} \tag{4.2}
$$

Structural identifiability of a within-host model using the differential algebra approach is defined as follows [17, 18].

Definition 2. Let $c(p)$ denote the coefficients of the input-output equations, where p is the vector of model parameters. We say that the within-host model Eq (2.2) is structurally identifiable based on the observations $y_1(t)$ and $y_2(t)$ if and only if

$$c(p) = c(\hat{p}) \quad \Longrightarrow \quad p = \hat{p}.$$

According to Definition 2, we need to show that the map from the parameter space to the coefficients of the input-output equations is one to one. So, suppose that there is another parameter vector $\hat{p} = (\hat{r}, \hat{\beta}, \hat{c_1}, \hat{d}, \hat{\delta}, \hat{\pi}, \hat{c}, \hat{c_2}, \hat{\lambda}, \hat{\gamma}, \hat{\mu})$ which has produced the same target cell and viral load observations. Hence, setting $c(p) = c(\hat{p})$, we obtain the following system of equations:

$$
\begin{aligned}
& \gamma = \hat{\gamma}, \quad \pi = \hat{\pi}, \quad r = \hat{r}, \quad d = \hat{d}, \quad (\beta-\gamma)/\beta = (\hat{\beta}-\hat{\gamma})/\hat{\beta}, \quad (\beta\mu - 2d\gamma)/\beta = (\hat{\beta}\hat{\mu} - 2\hat{d}\hat{\gamma})/\hat{\beta}, \\
& (-cc_1\delta - c_1c_2\lambda + c_2\delta - c_2\mu)/c_1\pi = (-\hat{c}\hat{c_1}\hat{\delta} - \hat{c_1}\hat{c_2}\hat{\lambda} + \hat{c_2}\hat{\delta} - \hat{c_2}\hat{\mu})/\hat{c_1}\hat{\pi}, \quad c_2\gamma/c_1\pi = \hat{c_2}\hat{\gamma}/\hat{c_1}\hat{\pi}, \\
& 2r(c_1\lambda+\mu)/\beta = 2\hat{r}(\hat{c_1}\hat{\lambda}+\hat{\mu})/\hat{\beta}, \quad c_2\beta/c_1\pi = \hat{c_2}\hat{\beta}/\hat{c_1}\hat{\pi}, \\
& (-cc_1 - c_1\delta + c_2)/c_1\pi = (-\hat{c}\hat{c_1} - \hat{c_1}\hat{\delta} + \hat{c_2})/\hat{c_1}\hat{\pi}.
\end{aligned}
$$

Solving the above nonlinear system in `MATHEMATICA`, we obtain a set of positive solutions as follows:

$$\Big\{\gamma = \hat{\gamma}, \quad d = \hat{d}, \quad \beta = \hat{\beta}, \quad \delta = \hat{\delta}, \quad \pi = \hat{\pi}, \quad c = \hat{c}, \quad r = \hat{r}, \quad \mu = \hat{\mu}, \quad c_1\lambda = \hat{c}_1\hat{\lambda}, \quad \frac{c_2}{c_1} = \frac{\hat{c_2}}{\hat{c_1}}\Big\}.$$

This concludes that the within-host model Eq (2.2) is not structurally identifiable. The model only reveals the parameters $\gamma, d, \beta, \delta, \pi, c, r$, and μ from the CD4 cells counts and the viral load observations. The parameters c_1, c_2, and λ are correlated. Infinitely many values of the parameters, c_1, c_2, and λ will give the same observations as long as the product $c_1\lambda$ and the ratio $\frac{c_2}{c_1}$ remain constant. We summarize the structural identifiability result in the following proposition.

Proposition 1. The within-host model Eq (2.2) is not structurally identifiable given the CD4 cell count and viral load observations.

When a model is structurally unidentifiable, like the within-host model Eq (2.2), there are basically two approaches to obtain a structurally identifiable model. We show both of these approaches here. The differential algebra approach gives the correlations between the parameters λ, c_1, and c_2; hence, fixing one of these parameters, we can obtain a structurally identifiable model. For instance, if we fix $\lambda = \hat{\lambda}$, from the parameter correlations we obtain that $c_1 = \hat{c_1}$ and $c_2 = \hat{c_2}$. Similarly, fixing c_1 or c_2 will reveal the other parameters uniquely. Summarizing the above analysis, we have the following proposition.

Proposition 2. If either λ, c_1, or c_2 is fixed, then the within-host model Eq (2.2) uniquely reveals all of its parameters based on the CD4 cell count and viral load observations.

The second way to obtain a structurally identifiable model is to add another observation. For this study, we chose to add the total protein levels in infected individuals, $y_3(t) = P(t)$, as the third observation, as well as perform the structural identifiability analysis again. We can obtain the following input-output equations of Eq (2.2) with observations $y_1(t), y_2(t)$, and $y_3(t)$ by using DAISY.

$$
\begin{aligned}
0 &= T'Pc_1 + T + TV\beta + TPc_1d + Td - Pc_1r - r \\
0 &= P' - VP\gamma + P\mu - \lambda \\
0 &= V''Pc_1 + V'' + V'P^2c_1c_2 + V'P(cc_1 + c_1\delta + c_2) + V'(c + \delta) - \\
&\quad TV\beta\pi + V^2P^2c_1c_2\gamma + V^2Pc_2\gamma + VP^2c_1c_2(\delta - \mu) + \\
&\quad VP(cc_1\delta + c_1c_2\lambda + c_2\delta - c_2\mu) + V(c\delta + c_2\lambda)
\end{aligned}
$$

Based on the definition of structural identifiability, we can solve the nonlinear system $c(p) = c(\hat{p})$ for p and obtain the following,

$$
\{\gamma = \hat{\gamma}, \quad d = \hat{d}, \quad \beta = \hat{\beta}, \quad \delta = \hat{\delta}, \quad \pi = \hat{\pi}, \quad c = \hat{c}, \quad r = \hat{r}, \quad \mu = \hat{\mu}, \quad \lambda = \hat{\lambda}, \quad c_1 = \hat{c}_1, \quad c_2 = \hat{c_2}\},
$$

which means that all of the parameters can be uniquely determined. Hence, the within-host model Eq (2.2) is structurally identifiable given the observations of the CD4 cell counts, the viral load, and the protein level. We summarize the structural identifiability analysis in the following proposition.

Proposition 3. The within-host model Eq (2.2) is structurally identifiable given the observations of the CD4 cell counts, the viral load, and the protein level.

4.2. Parameter estimation

Data: In this study, we have used data from a multi-site study in Africa and Southeast Asia, designed to track the HIV status to characterize the progression of acute HIV-1 infection [15]. The HIV viral load was determined from 22 HIV-1 acutely infected people, and the CD4 cell count was obtained from 24 infected individuals. For our analysis, we used the average of all viral load measurements and all CD4 cell counts from these infected individuals. We used MATLAB to extract the data from Figure 1(A) in [15]. Additionally, we obtained the protein levels of HIV-infected individuals from another study [16]. In Table 3, we present the logarithmic values of the extracted data and the protein levels used in parameter estimation.

Table 3. Logarithmic values of average viral loads (RNA copies/ ml), average CD4 cell counts (cells/μl) and average value of total protein (g/L).

	Day 1.9	Day 5.8	Day 9.7	Day 13.7	Day 17.5	Day 20.8	Day 24.7	Day 27.7
Viral load (RNA copies/ml)	3.5273	5.0143	5.9952	6.2903	5.9324	5.4763	5.0858	4.7602
	Day 31.7	Day 4.7	Day 48.8	Day 63.3	Day 94.1	Day 174.55	Day 257.39	
Viral load (RNA copies/ml)	4.5493	4.4706	4.3752	4.4809	4.4172	4.1996	3.8672	
	Day 1.9	Day 17.8	Day 32.1	Day 49.0	Day 94.0	Day 259.28		
CD4 count (cells/μl)	2.9532	2.7728	2.8006	2.8616	2.8105	2.7416		
	Day 2.2	Day 6.3	Day 10.2	Day 14.0	Day 18.2	Day 21.2	Day 25.0	Day 28.3
Total Protein (g/L)	2.1339	1.9446	2.0056	2.0270	2.0065	1.8830	1.9389	1.9592
	Day 32.2	Day 41.2	Day 49.2	Day 68.0	Day 93.5	Day 178.5	Day 254.1	
Total Protein (g/L)	1.9803	1.9625	1.9420	1.9398	1.9299	1.9762	2.0659	

Parameter estimation: For the parameter estimation, we fit the model Eq (2.2) to the CD4 cell count, viral load, and total protein data. We chose to use the least squares (LS) method to perform the parameter estimation problem. The LS principle is based on minimizing the objective function given by Eq (4.4). Let $\{t_1, t_2, ..., t_n\}$ represent the discrete times when the viral load is observed, let $\{t_1, t_2, ..., t_m\}$ represent the discrete times when the CD4 cell counts are observed, and let $\{t_1, t_2, ..., t_l\}$ represent the discrete times when the total protein level is observed. Then, the measurements can be written as,

$$Y_1^i = T(t_i) + \epsilon_i \quad Y_2^j = V(t_j) + \epsilon_j \quad \text{and} \quad Y_3^k = P(t_k) + \epsilon_k. \tag{4.3}$$

where, $i = 1, 2, ...n$, $j = 1, 2, .., m$, and $k = 1, 2, .., l$, and ϵ_i,ϵ_j, and ϵ_k represent the observation or measurement errors. We estimate the parameter $\hat{p}$ of the model Eq (2.2) by solving the following optimization problem,

$$\hat{p} = \min_{p} \left(\sum_{i=1}^{n} (\log_{10}(T(t_i)) - \log_{10}(Y_1^i))^2 + \sum_{j=1}^{m} (\log_{10}(V(t_j)) - \log_{10}(Y_2^j))^2 + \sum_{k=1}^{l} (\log_{10}(P(t_k)) - \log_{10}(Y_3^k))^2 \right), \tag{4.4}$$

where $T(t_i)$, $V(t_j)$, and $P(t_k)$ are the solutions to the system Eq (2.2) at times t_i, t_j, and t_k, respectively. To minimize and evaluate the objective function Eq (4.4), we use the MATLAB function fminsearchbnd. We chose parameter values based on previous studies, and the initial values of state variables were chosen as $(T(0), T_i(0), V(0), P(0)) = (2613, 0, 1048, 69)$. Then, we perform iterations until the error did not change and the optimization tolerance was reached. The estimated parameter values are given in Table 4.

Table 4. Estimated parameter values of the within-host models Eqs (2.2) and (4.7), obtained by solving the LS problem Eq (4.4).

Parameter	Value Eq (2.2)	Value Eq (4.7)	Unit
r	82.6352	80.3611	$\frac{\text{cells}}{\mu\text{l} \times \text{day}}$
d	0.0986	0.0931	day^{-1}
β	1.9109×10^{-5}	2.0681×10^{-5}	$\frac{\text{ml}}{vRNA \times \text{day}}$
δ	0.9071	0.9311	day^{-1}
π	1.0975×10^{4}	1.1726×10^{4}	$\frac{10^{-3} \times vRNA}{\text{cells} \times \text{day}}$
c	1.1813	1.2123	day^{-1}
c_1	1.6883	1.9417	$\frac{\text{L}}{\text{g}}$
c_2	7.7271×10^{-11}	–	$\frac{\text{L}}{\text{g} \times \text{day}}$
γ	1.1391×10^{-8}	8.933×10^{-9}	$\frac{\text{ml}}{vRNA \times \text{day}}$
μ	0.0129	0.0131	day^{-1}
λ	1.2381	1.2410	$\frac{\text{g}}{\text{l} \times \text{day}}$

Remark 4.1. *Using the estimated values, we compute the basic reproduction number as*

$$\mathcal{R}_0 = \frac{r\pi\beta\mu^2}{\delta d(\mu + c_1\lambda)(c\mu + \lambda c_2)} = 1.0061.$$

There exist three solutions of Eq.(3.5) for V^*:

$$V_1^* = 2933.78, \quad V_2^* = 1132470, \quad \textit{and} \quad V_3^* = 1970740.$$

To have positive values for state variables, we require that $V^* < \frac{\mu}{\gamma} = 1132470$. Thus, $V^* = 2933.78$. Then, $P^* = \dfrac{\lambda}{\mu - \gamma V^*} = 96.226$, and

$$\gamma V^* \left(1 + \frac{c_1 P^*}{1 + c_1 P^*}\right) = 0.0000666329 < \mu = 0.0129.$$

4.3. Practical identifiability

In Section 4.1, we demonstrated that the within-host model of HIV and nutrition, i.e., Eq (2.2), is not structurally identifiable based on the observations of CD4 cells and viral load only. To obtain a structurally identifiable model, we incorporated an additional observation, namely, the protein levels. With the inclusion of protein levels, the within-host model (2.2) became structurally identifiable. Structural identifiability analysis provides a theoretical base for practical identifiability analysis, which accounts for the masurement errors in actual observations. Although a structurally identifiable model can be determined from exact observations without measurement errors, practical identifiability is also crucial in real-world scenarios, where actual observations are subject to measurement errors. Hence, a structurally identifiable model may not necessarily be practically identifiable [17, 20, 35]. In this study, we evaluate the practical identifiability of the HIV within-host model (2.2) to assess the accuracy of the estimated parameters. Various methods have been employed in previous studies to determine the practical identifiability of ODE model parameters, such as those described in [17, 18, 20, 35–39]. In our study, we conducted MCSs to assess the practical identifiability of the estimated parameter vector $\hat{p}$. Specifically, we performed simulations based on a mathematical algorithm that generates $M = 1000$ sets of synthetic data by using the estimated parameters as the true parameter set $\hat{p}$, with noise added at increasing levels of 1, 5, 10, and 20%. The following steps describe our MCS process:

- Solve the HIV within-host model (2.2) numerically with the true parameter values denoted by $\hat{p}$ that have been obtained by fitting the model to the experimental data from the previous section. Then get the observations (output) at the experimental time points.
- Generate $M = 1000$ data sets for each of the experimental time points with 4 different measurement error levels, $\sigma = 1, 5, 10, 20\%$. Each measurement error is assumed to be distributed with a normal distribution with a zero mean and standard deviation σ.

$$Y_1^i = T(t_i, \hat{p}) + \epsilon_i \quad Y_2^j = V(t_j, \hat{p}) + \epsilon_j \quad \text{and} \quad Y_3^k = P(t_k, \hat{p}) + \epsilon_k, \tag{4.5}$$

  where measurement errors $\epsilon_i, \epsilon_j, \epsilon_k \sim N(0, \sigma)$.

- Fit the model to each of the 1000 data sets to estimate the parameter set p_m.

$$p_s = \min_{p} \left(\sum_{i=1}^{n} (\log_{10}(T(t_i)) - \log_{10}(Y_1^i))^2 + \sum_{j=1}^{m} (\log_{10}(V(t_j)) - \log_{10}(Y_2^j))^2 + \sum_{k=1}^{l} (\log_{10}(P(t_k)) - \log_{10}(Y_3^k))^2 \right) \quad s = 1, 2, \ldots, 1000. \tag{4.6}$$

- Calculate the average relative estimation error (ARE) for each parameter in the within-host model by as follows:

$$ARE(p^k) = 100\% \times \frac{1}{M} \sum_{s=1}^{M} \frac{|\hat{p}^k - p_s^k|}{|\hat{p}^k|},$$

  where $\hat{p}^k$ is the k^{th} element of the true parameter set $\hat{p}$ and p_s^k is the k^{th} element of p_s.
- Use $ARE s$ to determine whether each of the model's parameters is practically identifiable.

As the level of measurement error in the data increases, so will the ARE values for the estimated parameters. However, some parameter's ARE values may see a rapid increase, whereas others may see a more gradual increase. If the ARE value for a given parameter estimate is a constant multiple of the noise level, the parameter is considered to be practically identifiable; otherwise, it is considered unidentifiable. More information on the MCS algorithm can be found in articles such as [17, 35]. We define practical identifiability in this study as follows:

Definition 3. Let the ARE(p) be the average relative estimation error of the parameter p. The practical identifiability of parameter p is determined by comparing the ARE with the measurement error σ. More precisely, The parameter p is (strongly) practically identifiable if

$$0 \leq ARE(p) \leq \sigma.$$

The parameter p is weakly practically identifiable if

$$\sigma < ARE(p) \leq 10\sigma.$$

The parameter p is practically unidentifiable if

$$ARE(p) > 10\sigma.$$

Here, we would like to mention that, in our previous studies, we defined the parameter p to be practically identifiable if its ARE value falls between zero and σ, and practically unidentifiable if it lies outside of this range. Recently, we refined this definition [27, 28] to include the weakly practically identifiable case. This revised definition of practical identifiability does not conflict with the earlier definitions, rather, it includes an additional condition for weak identifiability.

We first performed the MCS for the actual data points presented in Table 3, and we present the results in Table 5. As the noise in the data increased, the ARE of each parameter increases gradually, except for the parameter c_2. Notably, the ARE of parameter c_2 experienced a rapid increase to 17049, even for a measurement error of 1%. Additionally, the ARE of c_2 exhibited an irregular trend, increasing

to 26955 for 5% noise in data and then decreasing to 9606.2 for 20% noise. This indicates that the accuracy of the c_2 parameter is not solely impacted by the noise in the data, as there may be some other hidden factor influencing these high AREs. Overall, the parameter c_2 is practically unidentifiable. Based on Definition 3, the rest of the parameters were found to be only weakly practically identifiable. Further, fixing c_2 and running the MCS again for the actual data did not improve the identifiability of the remaining parameters (see Table 6). All parameters of the within-host model (2.2) remained weakly practically identifiable, even when c_2 was fixed. We summarize the practical identifiability of the within-host model parameters in the following proposition.

Table 5. MCS results: *ARE* of each parameter of the within-host model Eq (2.2) based on the actual data frequency.

Parameter	r	d	β	δ	π	c	c_1	c_2	γ	μ	λ
ARE $\sigma = 1\%$	1	1	6.1	4.4	2.9	5.9	6.6	17049	4.3	4.9	5.1
ARE $\sigma = 5\%$	5.2	5.2	23.1	22.7	13.9	26.4	26.1	26955	19.4	18.5	19.4
ARE $\sigma = 10\%$	10.2	10.1	60.6	40.8	31.1	49.7	63.5	23721	35.3	29.4	31.2
ARE $\sigma = 20\%$	19.1	18.2	114.3	70.2	91.1	108.2	125.5	9606.2	62.9	44.9	49.1

Table 6. MCS results: *ARE* of each parameter of the within-host model Eq (2.2), as obtained by using the actual data frequency with c_2 fixed at the fitted value 7.73×10^{-11}.

Parameter	r	d	β	δ	π	c	c_1	γ	μ	λ
ARE $\sigma = 1\%$	1	0.9	5.6	4.4	2.8	5.8	6.1	3.6	3.2	3.5
ARE $\sigma = 5\%$	5.2	4.9	18.8	22.6	13.4	26.4	20.8	15.0	9.8	11.1
ARE $\sigma = 10\%$	10.3	10.0	49.1	39.3	31.0	49.4	53.5	31.4	22.4	25
ARE $\sigma = 20\%$	19.2	17.9	92.2	71.9	87.6	107.2	98.1	58.5	38	41.7

Proposition 4. The clearance rate of immunoglobulins, c_2, is not practically identifiable based on the data given in Table 3. The remaining parameters of the within-host model (2.2) are weakly practically identifiable based on the data. When c_2 is fixed, the parameters of the within-host model (2.2) remain weakly practically identifiable.

In structural identifiability analysis, we make two assumptions. The first assumption is that the model is correct and the experimental data can be explained by the model. The second assumption is that there are an infinite number of noise-free data points. However, the experimental data are typically collected at a low frequency, resulting in a limited number of available data points. To put this in perspective, for the experimental data we used 15 measurements for the viral load and the total protein level, as well as 6 measurements for the CD4 cell counts (see Table 3). A total of 36 data points were available for parameter estimation. The within-host model (2.2) is structurally identifiable given the observations of viral load, CD4 count, and protein level. However, we found that the parameter c_2 is practically unidentifiable and the rest of the model parameters are only weakly practically identifiable based on the experimental data (see Table 5).

Given the scarcity of the experimental data, it was deemed necessary to investigate how the frequency of data collection affects the accuracy of estimates. To investigate the impact of data frequency on the practical identifiability of the model parameters, we repeated the MCS by gradually increasing the total number of data points. We first doubled the actual data frequency, generated a total of 72 data points and computed the AREs for each parameter; the results are shown in Table 13. We found that the parameter c_2 is practically unidentifiable, parameters r and d are practically identifiable, and the rest of the parameters remain weakly identifiable. We repeated the MCS, this time by generating one data point per day for 300 days. This yielded 300 data points for each observation, namely, the viral load, CD4 count, and total protein levels, resulting in a total of 900 data points. So, we increased the data frequency from 72 to 900 and found that c_2 is still practically unidentifiable, parameters δ and c remain weakly identifiable, and the rest of the parameters are practically identifiable (see Table 7). We continued increasing the data frequency, this time repeating the MCS by generating 10 data points per day, which resulted in a total of 9000 data points. With this high data frequency all of the parameters became practically identifiable, except for the parameter c_2 (see Table 8). With the exception of the parameter c_2, we found that the ARE of each of the parameters decreases as the number of data points increases.

Table 7. MCS results: *ARE* of each parameter of the within-host model Eq (2.2), as obtained by generating one data point per day for a period of 300 days.

Parameter	r	d	β	δ	π	c	c_1	c_2	γ	μ	λ
ARE $\sigma = 1\%$	0.2	0.2	0.9	2.0	0.8	2.6	1.2	11052	1.0	0.9	0.8
ARE $\sigma = 5\%$	1.1	1.1	2.8	7	2.7	9.0	4.0	8647.1	4.0	3.3	3.3
ARE $\sigma = 10\%$	2.3	2.2	5.5	11.4	4.9	14.7	7.4	8671.5	8.5	7.5	7.6
ARE $\sigma = 20\%$	4.5	4.2	9.5	17.1	9.3	23.6	12.8	30584	16.6	14.5	14.6

Table 8. MCS results: *ARE* of each parameter of the within-host model (2.2) by generating 10 data points per day for 300 days, resulting in total of 9000 data points.

Parameter	r	d	β	δ	π	c	c_1	c_2	γ	μ	λ
ARE $\sigma = 1\%$	0.1	0.1	0.3	0.6	0.2	0.8	0.4	5742.5	0.3	0.3	0.3
ARE $\sigma = 5\%$	0.4	0.3	1.3	3.1	1.2	4.0	1.7	10776	1.5	1.2	1.2
ARE $\sigma = 10\%$	0.7	0.7	2.4	5.4	2	7.0	3.2	8336.6	2.8	2.3	2.3
ARE $\sigma = 20\%$	1.5	1.4	3.6	8.5	3.4	10.8	5	17316	5.4	4.6	4.6

The clearance rate of immunoglobulins denoted by c_2, is the only parameter that remained practically unidentifiable even for 9000 data points. Based on the MCS results and the validation of the model with data, it appears that the rate of clearance of the virus by immunoglobulins is negligible and practically unidentifiable. As such, we proceeded with adjusting the within-host HIV and nutrition model Eq (2.2) by setting $c_2 = 0$. The modified model becomes

$$\begin{aligned}
\frac{dT}{dt} &= r - \frac{\beta}{1 + c_1 P} VT - dT, \\
\frac{dT_i}{dt} &= \frac{\beta}{1 + c_1 P} VT - \delta T_i, \\
\frac{dV}{dt} &= \pi T_i - cV, \\
\frac{dP}{dt} &= \lambda + \gamma PV - \mu P.
\end{aligned} \tag{4.7}$$

We chose to first study the structural idenfitiability analysis of the within-host model (4.7) based on the observations of the viral load, CD4 cells, and protein levels. We first derive the input-output equations:

$$\begin{aligned}
0 &= T' P c_1 + T' + TV\beta + TPc_1 d + Td - Pc_1 r - r, \\
0 &= P' - VP\gamma + P\mu - \lambda, \\
0 &= V'' P c_1 + V'' + V' P c_1 (c + \delta) + V'(c + \delta) - TV\beta\pi + VPcc_1\delta + Vc\delta.
\end{aligned}$$

Based on the definition of structural identifiability, we solve the nonlinear system $c(p) = c(\hat{p})$ for p and obtain the following two sets of solutions,

$$S_1 : \{\gamma = \hat{\gamma}, \quad d = \hat{d}, \quad \beta = \hat{\beta}, \quad \delta = \hat{\delta}, \quad \pi = \hat{\pi}, \quad c = \hat{c}, \quad r = \hat{r}, \quad \mu = \hat{\mu}, \quad \lambda = \hat{\lambda}, \quad c_1 = \hat{c}_1\},$$

$$S_2 : \{\gamma = \hat{\gamma}, \quad d = \hat{d}, \quad \beta = \hat{\beta}, \quad \delta = \hat{c}, \quad \pi = \hat{\pi}, \quad c = \hat{\delta}, \quad r = \hat{r}, \quad \mu = \hat{\mu}, \quad \lambda = \hat{\lambda}, \quad c_1 = \hat{c}_1\}.$$

These results imply that the parameters c and δ are locally identifiable, whereas all other parameters are globally identifiable.

Proposition 5. The within-host model Eq (4.7) is structurally locally identifiable based on the observations of the CD4 cell counts, the viral load and the protein level. In particular, the parameters c and δ are locally identifiable, whereas all other parameters are globally identifiable.

We fit the new modified within-host model of HIV and nutrition Eq (4.7), to the data given in Table 3 and estimated its parameters. The estimated values are presented in Table 4. Next, we performed the MCS with the actual data frequency; the results are presented in Table 9. As can be seen in Table 9, the parameters r and d are strongly practically identifiable, whereas all of the other parameters are weakly practically identifiable.

We performed the same analysis as we did for the within-host model (2.2) and repeated the MCS by gradually increasing the data points. We observed that practical identifiablity conclusion does not change when we double the data frequency, meaning that the parameters r and d are practically identifiable given 72 data points, while the rest of the parameters remain weakly practically identifiable (see Table 10). When the data frequency increased to a total of 900 data points all of the parameters became practically identifiable, except for δ and c (see Table 11). Finally, with a total of 9000 data points, all of the parameters were found to be become practically identifiable (see Table 12). The simulation results presented above demonstrate how the number of data points impacts the identifiability of the model parameters. These findings highlight the importance of carefully considering the frequency of data collection in experimental design to ensure that the resulting data is informative and suitable for accurate model parameter estimation.

Table 9. MCS results: *ARE* of each parameter of the within-host model (4.7) with the actual data frequency.

Parameter	r	d	β	δ	π	c	c_1	γ	μ	λ
ARE $\sigma = 1\%$	1.0	0.9	5.0	4.2	2.8	5.5	5.3	4.0	2.4	2.7
ARE $\sigma = 5\%$	5.0	4.9	17.8	21.6	13.8	25.5	19.8	18.9	9.3	10.8
ARE $\sigma = 10\%$	9.9	9.7	43.7	38.1	28.9	45.8	47.2	38.4	20.0	22.4
ARE $\sigma = 20\%$	18.1	17.4	135.0	65.5	73.8	93.6	136.0	67.2	39.5	44.9

Table 10. MCS results: *ARE* of each parameter of the within-host model (4.7), as obtained by generating 72 data points.

Parameter	r	d	β	δ	π	c	c_1	γ	μ	λ
ARE $\sigma = 1\%$	0.7	0.6	2	3.4	2.2	4.5	2.4	2.5	1.2	1.4
ARE $\sigma = 5\%$	3.6	3.1	10.9	15.5	9.8	19.8	11.9	12.8	6.1	7
ARE $\sigma = 10\%$	7.2	6.6	24.6	29	19.1	36.6	26.3	25	10.8	12.4
ARE $\sigma = 20\%$	14	12.1	52.3	46.2	54.4	78.1	60.6	51.6	25.6	28.5

Table 11. MCS results: *ARE* of each parameter of the within-host model (4.7), as obtained by generating 1 data point per day, resulting in a total of 900 data points.

Parameter	r	d	β	δ	π	c	c_1	γ	μ	λ
ARE $\sigma = 1\%$	0.2	0.2	0.6	1.4	0.7	1.9	0.9	1.1	0.8	0.8
ARE $\sigma = 5\%$	1.1	1	2.5	6.2	3.1	8.2	3.4	4.5	2.9	3
ARE $\sigma = 10\%$	2.1	2	4.5	10.6	5.6	14.6	6	8.4	5	5.1
ARE $\sigma = 20\%$	4.1	3.9	8.5	16.5	9.6	23.1	11.7	16.1	8.5	8.7

Table 12. MCS results: *ARE* of each parameter of the within-host model (4.7), as obtained by generating 10 data points per day, resulting in a total of 9000 data points.

Parameter	r	d	β	δ	π	c	c_1	γ	μ	λ
ARE $\sigma = 1\%$	0.1	0.1	0.2	0.5	0.2	0.6	0.3	0.4	0.3	0.3
ARE $\sigma = 5\%$	0.3	0.3	0.9	2.1	1.1	2.7	1.3	1.7	1.1	1.1
ARE $\sigma = 10\%$	0.6	0.6	1.8	4.5	2.2	6	2.5	3.2	2.1	2.1
ARE $\sigma = 20\%$	1.4	1.3	3.3	7.7	3.7	9.8	4.6	6	3.8	3.9

5. Discussion

In this paper, the dynamics of CD4 cells (T), infected target cells (T_i), virus particles (V), and protein (P) were studied. Taking into consideration the significant elevation of the protein level in an HIV infected individuals, we extended the well-studied target cell limited model in the form of Eq (2.1) to include protein and introduce the novel HIV within-host model Eq (2.2). We then studied the equilibria, stability, and structural and practical identifiability of the model parameters.

Following the model formation, we analyzed the equilibria and their stability for the model Eq (2.2). We determined that this model has two equilibrium points: an infection-free equilibrium point (E_0), where there is no presence of infected cells and viruses, and infection equilibrium point (E_1), where each of the variables has a positive value. Working with the E_0 equilibrium point, to study its stability, we linearized Eq (2.2) and found that the equilibrium is locally asymptotically stable if the basic reproduction number $\mathcal{R}_0 < 1$, and unstable if $\mathcal{R}_0 > 1$. Similarly, for the infection equilibrium E_1, we proved that the equilibrium point is unique and locally asymptotically stable when $\mathcal{R}_0 > 1$ an additional condition holds.

If a model is structurally unidentifiable based on the observations, then multiple parameter combinations can give the same observations. Therefore, it is not possible to estimate a unique set of parameters by fitting a structurally undentifiable model. We studied the structural identifiability of the model Eq (2.2). The structural identifiability can be studied independently of the data by using DAISY. With the help of DAISY, and with the CD4 cell count (T) and viral load (V) as the two observation functions, we found that the parameter that reduces the target cells' susceptibility in the presence of protein, the rate of clearance of virus by immunoglobulins, and the daily protein intake, namely, the parameters c_1, c_2, and λ, are unidentifiable. To remedy this situation we added a third data set, namely the protein level, and considered the identifiability of the model with three observation functions: CD4 cell (T), viral load (V), and protein level (P). In this case, we observed that the model is structurally identifiable, that is, all parameters are structurally identifiable. We then estimated all parameters of Eq (2.2) by employing the logarithmic values of the average viral load (RNA copies/ml), average CD4 cell count (cells/μ l), and the average value of total protein (g/L) data from a multi-site study in Africa and Southeast Asia that was designed to track the HIV status to characterize the progression of acute HIV-1 infection [15, 16]. These data are presented in Table 3. By fitting the model Eq (2.2) to the viral load, CD4 cell count, and total protein data by using the LS approach, we obtained the parameter values presented in Table 4. The graphs of the fitted curves are presented in Figure 2.

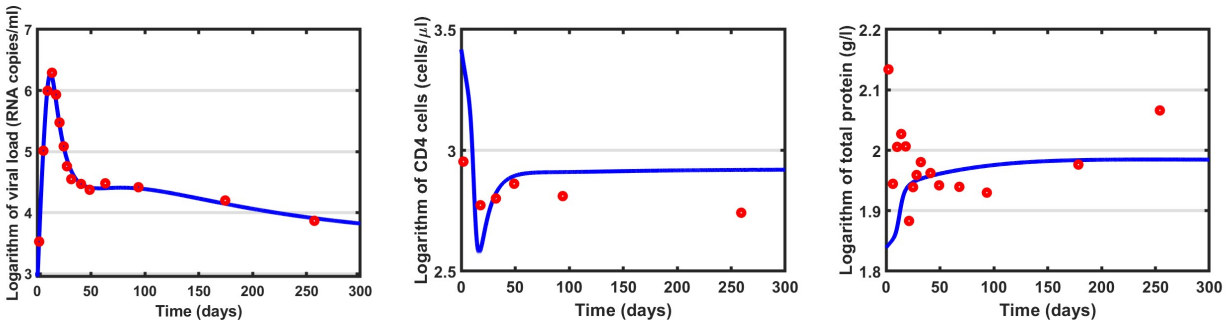

Figure 2. Logarithmic values of viral load, CD4 cell count, and total protein data (red dots) plotted along the model prediction (blue curve) according to the estimated values of the parameters presented in Table 4.

Even though the estimated parameters were obtained based on data and we used a structurally identifiable model, we found that the effect of the noise in real data on the parameters may lead to practical unidentifiability. Thus, we performed practical identifiability analysis. For the practical identifiability analysis, we used the MCS algorithm and generated 1000 sets of data with [1%, 5%, 10%, 20%] synthetic noise for the CD4 cells, viral load, and protein level. We then fitted the 1000 data sets for each noise level, recovered 1000 sets of parameters for each noise level, and calculated the *ARE* for each parameter.

Data frequency plays an important role in the practical identifiability of each parameter. We applied MCS for four different data frequencies: data collected ten times a day (total of 9000 data points), data collected daily (total of 900 data points), double the frequency of actual data (total of 72 data points) and more sparsely collected data, relative to actual data (total of 36 data points). First, we conducted as MCS with the actual data frequency, which resulted in c_2 being unidentifiable and all other parameters being only weakly identifiable with *ARE*s above the noise level (see Table 5). To improve the results, we doubled the actual data frequency and reran the MCS, and the results are presented in Table 13. The results still indicated that c_2 was unidentifiable, while r and d became practically identifiable; also, all other parameters were only weakly identifiable, though with decreased *ARE* values. Next, we increased the total number of data points to 900 and 9000 and performed MCSs (see Tables 7 and 8). The results indicated that the parameter c_2 is still not practically identifiable, but every other parameter is practically identifiable. As a result, we fixed c_2 to the fitted value and performed the MCS for both high-frequency and the actual data frequency cases; we found that all other parameters are practically identifiable with high-frequency data and all parameters are weakly identifiable under the actual data frequency. The results are presented in Tables 6 and 14, respectively.

To further investigate the parameters of the model Eq (2.2), we recast the model into another form, i.e., Eq (4.7) by setting $c_2 = 0$, since the value of c_2 was originally estimated to be very small. We showed that the new within-host model of HIV and protein given by Eq(4.7) is structurally locally identifiable for the observations of CD4 cell count, viral load and the protein level. Then, we fitted the model (4.7) and estimated its parameters based on the experimental data. We conducted MCSs by using the experimental data frequency and by doubling the frequency of the actual data. The results are presented in Tables 9 and 10, respectively. We observed that the identifiability conclusion of the parameters did nit change from 36 to 72 data points. In particular, parameters r and d were found to be practically identifiable, while the remaining parameters were weakly practically identifiable. To enhance these results, we performed MCSs with 900 and 9000 data points (see Tables 11 and 12). When the data frequency increased to a total of 900 data points, all parameters became practically identifiable, except for δ and c. Finally, with a total of 9000 data points all of the parameters became practically identifiable. Taking all results of the MCSs of this work into account, we determined that, if the frequency of data collection is high, then more parameters of the model will be more practically identifiable.

In conclusion, we studied the dynamics, equilibria, and stability of CD4 cells, infected target cells, virus load, and protein level by developing a nutrition and HIV within-host model. Using this model, we not only estimated the parameters by using real-life data we also studied both the structural and practical identifiability of the parameters. Overall, the novel nutrition and HIV within-host model has two equilibrium points, its parameters are structurally identifiable with three observations, and practical identifiability of the estimated parameters depends upon the frequency of the measured data. In summary, this model can be a useful tool to understand the HIV infection and protein's function within the human body.

Table 13. MCS results: *ARE* of each parameter of the within-host model Eq (2.2), obtained by generating 72 data points (i.e., doubled the actual data frequency).

Parameter	r	d	β	δ	π	c	c_1	c_2	γ	μ	λ
ARE $\sigma = 1\%$	0.8	0.7	2.8	3.6	2.2	4.9	3.2	14568	2.9	3.4	3.6
ARE $\sigma = 5\%$	3.7	3.4	14.6	16	10.3	21.3	16.9	10483	14.2	14.3	14.8
ARE $\sigma = 10\%$	7.5	6.9	33.2	28.9	20.9	39.5	37.1	10534	26.7	22.3	23.1
ARE $\sigma = 20\%$	14.3	12.4	80	45.3	56.8	84.9	90.4	13005	46	32.1	35.3

Table 14. MCS results: *ARE* of each parameter in the within-host model Eq (2.2), as obtained by generating one data point per day for a period of 300 days, and with c_2 fixed at a value of 7.73×10^{-11}.

Parameter	r	d	β	δ	π	c	c_1	γ	μ	λ
ARE $\sigma = 1\%$	0.2	0.2	0.6	1.9	0.8	2.6	0.7	0.9	0.8	0.8
ARE $\sigma = 5\%$	1.1	1.0	2.4	6.9	3.4	9.1	3.1	3.5	3.0	3.1
ARE $\sigma = 10\%$	2.2	2.0	4.2	11.0	5.5	14.4	5.3	6.0	4.5	4.7
ARE $\sigma = 20\%$	4.3	4.1	8.8	16.5	11.3	25.1	11.4	12.7	8.7	9.0

Use of AI tools declaration

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

Conflict of interest

Maia Martcheva and Necibe Tuncer are guest editors for Mathematical Biosciences and Engineering and were not involved in the editorial review or the decision to publish this article. All authors declare that thereare no competing interests.

References

1. A. S. Perelson, P. W. Nelson, Mathematical analysis of HIV-1 dynamics in vivo, *SIAM Rev.*, **41** (1999), 3–44. https://doi.org/10.1137/S0036144598335107
2. M. Nowal, R. M. May, Virus dynamics: mathematical principles of immunology and virology, Oxford University Press, Oxford, 2000.
3. A. S. Perelson, R. M. Ribeiro, Modeling the within-host dynamics of HIV infection, *BMC Biol.*, **11** (2013), 1–10. https://doi.org/10.1186/1741-7007-11-96
4. G. W. Nelson, A. S. Perelson, A mechanism of immune escape by slow-replicating HIV strains, *J. Acq. Imm. Def.*, **5** (1992), 82–93.
5. A. S. Perelson, D. E. Kirschner, R. D. Boer, Dynamics of HIV infection of CD4+ T cells, *Math. Biosci.*, **114** (1993), 81–125. https://doi.org/10.1016/0025-5564(93)90043-A
6. L. Rong, A. S. Perelson, Modeling HIV persistence, the latent reservoir, and viral blips, *J. Theor. Biol.*, **260** (2009), 308–331. https://doi.org/10.1016/j.jtbi.2009.06.011
7. N. K. Vaidya, R. M. Ribeiro, A. S. Perelson, A. Kumar, Modeling the effects of morphine on simian immunodeficiency virus dynamics, *PLoS Comput. Biol.*, **12** (2016), e1005127. https://doi.org/10.1371/journal.pcbi.1005127
8. M. A. Stafford, L. Corey, Y. Cao, E. S. Daar, D. D. Ho, A. S. Perelson, Modeling plasma virus concentration during primary HIV infection, *J. Theor. Biol.*, **203** (2000), 285–301. https://doi.org/10.1006/jtbi.2000.1076
9. A. S. Perelson, P. W. Nelson, Modeling viral infections, in *Proceedings of Symposia in Applied Mathematics*, **59** (2002), 139–172.
10. R. Patil, U. Raghuwanshi, Serum protein, albumin, globulin levels, and A/G ratio in HIV positive patients, *Biomed. Pharmacol. J.*, **2** (2009), 321–325.
11. V. T. Sowmyanarayanan, S. Jun, A. Cowan, R. L. Bailey, The nutritional status of HIV-Infected US adults, *Curr. Dev. Nutr.*, **1** (2017), e001636. https://doi.org/10.3945/cdn.117.001636
12. R. K. Chandra, Nutrition and immunity: I. Basic considerations. II. Practical applications, *ASDC J. Dent. Child.*, **54** (1987), 193–197.
13. W. R. Beisel, Nutrition in pediatric HIV infection: setting the research agenda, *J. Nutr.*, **126** (1996), 2611–2615.
14. O. O. Oguntibeju, W. M. Van den Heever, F. E. Van Schalkwyk, The interrelationship between nutrition and the immune system in HIV infection: a review, *Pak. J. Biol. Sci.*, **10** (2007), 4327–4338. https://doi.org/10.3923/pjbs.2007.4327.4338
15. M. A. Eller, N. Goonetilleke, B. Tassaneetrithep, L. A. Eller, M. C. Costanzo, S. Johnson, et al., Expansion of inefficient HIV-specific CD8 T cells during acute infection, *J. Virol.*, **90** (2016), 4005–4016. https://doi.org/10.1128/jvi.02785-15
16. T. Were, J. O. Jesca, E. Munde, C. Ouma, T. M. Titus, F. Ongecha-Owuor, et al., Clinical chemistry profiles in injection heroin users from Coastal Region, Kenya, *BMC Clin. Pathol.*, **14** (2014), 1–9. https://doi.org/10.1186/1472-6890-14-32

17. N. Tuncer, T. T. Le, Structural and practical identifiability analysis of outbreak models, *Math. Biosci.*, **299** (2018), 1–18. https://doi.org/10.1016/j.mbs.2018.02.004

18. N. Tuncer, M. Marctheva, B. LaBarre, S. Payoute, Structural and practical identifiability analysis of Zika epidemiological models, *Bull. Math. Biol.*, **80** (2018), 2209–2241. https://doi.org/10.1007/s11538-018-0453-z

19. N. Tuncer, A. Timsina, M. Nuno, G. Chowell, M. Martcheva, Parameter identifiability and optimal control of an SARS-CoV-2 model early in the pandemic, *J. Biol. Dyn.*, **16** (2022), 412–438. https://doi.org/10.1080/17513758.2022.2078899

20. H. Miao, X. Xia, A. S. Perelson, H. Wu, On the identifiability of nonlinear ODE models and applications in viral dynamics, *SIAM Rev.*, **53** (2011), 3–39. https://doi.org/10.1137/090757009

21. E. A. Dankwa, A. F. Brouwer, C. A. Donnelly, Structural identifiability of compartmental models for infectious disease transmission is influenced by data type, *Epidemics*, **41** (2022), 100643. https://doi.org/10.1016/j.epidem.2022.100643

22. G. Massonis, J. R. Banga, A. F. Villaverde, Structural identifiability and observability of compartmental models of the COVID-19 pandemic, *Annu. Rev. Control*, **51** (2021), 441–459. https://doi.org/10.1016/j.arcontrol.2020.12.001

23. M. Renardy, D. Kirschne, M. Eisenberg, Structural identifiability analysis of age-structured PDE epidemic models, *J. Math. Biol.*, **84** (2022). https://doi.org/10.1007/s00285-021-01711-1

24. L. Gallo, M. Frasca, V. Latora, G. Russo, Lack of practical identifiability may hamper reliable predictions in COVID-19 epidemic models, *Sci. Adv.*, **8** (2022), eabg5234. https://doi.org/10.1126/sciadv.abg5234

25. K. Roosa, G. Chowell, Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models, *Theor. Biol. Med. Model.*, **16** (2019). https://doi.org/10.1186/s12976-018-0097-6

26. C. Tönsing, J. Timmer, C. Kreutz, Profile likelihood-based analyses of infectious disease models, *Stat. Methods Med. Res.*, **27** (2018), 1979–1998. https://doi.org/10.1177/0962280217746444

27. N. Heitzman-Breen, Y. R. Liyanage, N. Duggal, N. Tuncer, S. M. Ciupe, The effect of model structure and data availability on Usutu virus dynamics at three biological scales, *Roy. Soc. Open Sci.*, **11** (2024), 231146. https://doi.org/10.1098/rsos.231146

28. V. Sreejithkumar, K. Ghods, T. Bandara, M. Martcheva, N. Tuncer, Modeling the interplay between albumin-globulin metabolism and HIV infection, *Math. Biosci. Eng.*, **20** (2023), 19527–19552. https://doi.org/10.1098/rsos.231146

29. N. Tuncer, M. Martcheva, Determining reliable parameter estimates for within-host and within-vector models of Zika virus, *J. Biol. Dyn.*, **15** (2021), 430–454. https://doi.org/10.1080/17513758.2021.1970261

30. N. Tuncer, M. Martcheva, B. LaBarre, S. Payoute, Structural and practical identifiability analysis of Zika epidemiological models, *Bull. Math. Biol.*, **80** (2018), 2209–2241. https://doi.org/10.1007/s11538-018-0453-z

31. N. Tuncer, H. Gulbudak, V. L. Cannataro, M. Martcheva, Structural and practical identifiability issues of immuno-epidemiological vector–host models with application to rift valley fever, *Bull. Math. Biol.*, **78** (2016), 1796–1827. https://doi.org/10.1007/s11538-016-0200-2

32. S. M. Ciupe, N. Tuncer, Identifiability of parameters in mathematical models of SARS-CoV-2 infections in humans, *Sci. Rep.*, **12** (2022), 14637. https://doi.org/10.1038/s41598-022-18683-x

33. M. C. Eisenberg, S. L. Robertson, J. H. Tien, Identifiability and estimation of multiple transmission pathways in Cholera and waterborne disease, *J. Theor. Biol.*, **324** (2013), 84–102. https://doi.org/10.1016/j.jtbi.2012.12.021

34. G. Bellu, M. P. Saccomani, S. Audoly, L. D'Angiò, DAISY: A new software tool to test global identifiability of biological and physiological systems, *Comput. Meth. Prog. Bio.*, **88** (2007), 52–61. https://doi.org/10.1016/j.cmpb.2007.07.002

35. H. Miao, C. Dykes, L. M. Demeter, J. Cavenaugh, S. Y. Park, A. S. Perelson, et al., Modeling and estimation of kinetic parameters and replicative fitness of HIV-1 from flow-cytometry-based growth competition experiment, *Bull. Math. Biol.*, **70** (2008), 1749–1771. https://doi.org/10.1007/s11538-008-9323-4

36. H. Wu, H. Zhu, H. Miao, A. S. Perelson, Parameter identifiability and estimation of HIV/AIDS dynamic models, *B. Math. Biol.*, **70** (2008), 785–799. https://doi.org/10.1007/s11538-007-9279-9

37. F. G. Wieland, A. L. Hauber, M. Rosenblatt, C. Tonsing, J. Timmer, On structural and practical identifiability, *Curr. Opin Syst. Biol.*, **25** (2021), 60–69. https://doi.org/10.1016/j.coisb.2021.03.005

38. A. Pironet, P. D. Docherty, P. C. Dauby, J. G. Chase, T. Desaive, Practical identifiability analysis of a minimal cardiovascular system model, *Comput. Meth. Prog. Bio.*, **171** (2019), 53–65. https://doi.org/10.1016/j.cmpb.2017.01.005

39. A. Raue, J. Karlsson, M. P. Saccomani, M. Jirstrand, J. Timmer, Comparison of approaches for parameter identifiability analysis of biological systems, *Bioinformatics*, **30** (2014), 1440–1448. https://doi.org/10.1093/bioinformatics/btu006

AIMS Press