mathematics

MDPI

Article

Identifiability and Parameter Estimation of Within-Host Model of HIV with Immune Response

Yuganthi R. Liyanage [1], Leila Mirsaleh Kohan [1], Maia Martcheva [2] and Necibe Tuncer [1,*]

[1] Department of Mathematical Sciences, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, USA; lmirsalehkoh2018@fau.edu (L.M.K.)
[2] Department of Mathematics, University of Florida, Gainesville, FL 32611, USA; maia@ufl.edu
* Correspondence: ntuncer@fau.edu

Abstract: This study examines the interactions between healthy target cells, infected target cells, virus particles, and immune cells within an HIV model. The model exhibits two equilibrium points: an infection-free equilibrium and an infection equilibrium. Stability analysis shows that the infection-free equilibrium is locally asymptotically stable when $\mathcal{R}_0 < 1$. Further, it is unstable when $\mathcal{R}_0 > 1$. The infection equilibrium is locally asymptotically stable when $\mathcal{R}_0 > 1$. The structural and practical identifiabilities of the within-host model for HIV infection dynamics were investigated using differential algebra techniques and Monte Carlo simulations. The HIV model was structurally identifiable by observing the total uninfected and infected target cells, immune cells, and viral load. Monte Carlo simulations assessed the practical identifiability of parameters. The production rate of target cells (λ), the death rate of healthy target cells (d), the death rate of infected target cells (δ), and the viral production rate by infected cells (π) were practically identifiable. The rate of infection of target cells by the virus (β), the death rate of infected cells by immune cells (Ψ), and antigen-driven proliferation rate of immune cells (b) were not practically identifiable. Practical identifiability was constrained by the noise and sparsity of the data. Analysis shows that increasing the frequency of data collection can significantly improve the identifiability of all parameters. This highlights the importance of optimal data sampling in HIV clinical studies, as it determines the best time points, frequency, and the number of sample points required to accurately capture the dynamics of the HIV infection within a host.

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

MSC: 92D30; 92-08; 92-10

Citation: Liyanage, Y.R.; Mirsaleh Kohan, L.; Martcheva, M.; Tuncer, N. Identifiability and Parameter Estimation of Within-Host Model of HIV with Immune Response. *Mathematics* **2024**, *12*, 2837. https://doi.org/10.3390/math12182837

Academic Editor: Junseok Kim

Received: 31 July 2024
Revised: 10 September 2024
Accepted: 11 September 2024
Published: 12 September 2024

1. Introduction

According to the Joint United Nations Programme on HIV/AIDS (UNAIDS), global HIV statistics indicate that approximately 39 million people were living with HIV worldwide in 2022 [1]. Among these individuals, approximately 37.5 million were adults, and 1.5 million were children under the age of 15. Furthermore, the data reveal that 53% of these individuals were women and girls. HIV still affects a significant amount of people across the world and is a major health problem worldwide.

Because of its significance, the Human Immunodeficiency Virus (HIV) is one of the most studied diseases. Extensive research has shown that HIV specifically targets CD4 cells, a crucial component of the immune system. Once the virus infiltrates the host cell, viral RNA is converted into DNA, which then integrates into the host cell's DNA. These healthy target cells subsequently become infected cells. Following this, the infected cells undergo cell death, releasing new virus particles, which proceed to infect other target cells. As a result of the viral infection process, the host's immune response is triggered. A major component of the immune response to HIV are CD8 cells, which directly target and attack

cells infected with the virus [2]. It is this understanding of the within-host processes that serves as a foundation of within-host mathematical models of HIV.

In the recent 30 years, numerous mathematical models of HIV have been developed to assess both the quantitative and qualitative aspects of HIV replication and treatment [3–6]. The model by Perelson et al. [4] is one of the most well-known mathematical models for HIV dynamics within a host. This model depicts the rate of change of uninfected cells, infected cells, and free virus particles by using a system of ordinary differential equations. Perelson et al.'s paper presents a mathematical modeling of the dynamics of HIV infection and the impact of antiretroviral therapy. This model demonstrates the rapid turnover of HIV in the body and the effectiveness of combination therapy in reducing viral loads. Nowak and May, in [7], created a mathematical model for HIV that investigates the interactions between healthy target cells, infected target cells, and free virus particles. However, because immune cells are not included in their model as a distinct variable, the immunological response is simplified. Conway and Ribeiro, in [8], examined developments in immune response modeling to HIV infection, with a specific emphasis on cell-mediated immunity. The application of viral dynamics models to understand the role of the latent reservoir and the multi-phasic decline of the viral load during antiretroviral therapy was examined by D'Orso and Forst in [9]. While these models have provided valuable insights, many of them either oversimplify the immune response or become highly complex with numerous variables and parameters. Additionally, some recent models incorporate the effects of antiretroviral drugs, which adds another layer of complexity. We develop the model introduced in this paper on the basis of many successful prior models, aiming to strike a compromise between it being as simple as possible and still capturing the function of the immune system.

The primary objective of this work is to identify parameter values of the model, thus providing reliable parameter estimates for future studies. Parameter estimates are obtained through fitting the model to data; however, this process is generally very ill-posed since multiple parameter combinations can provide the same output. Therefore, it is critical to examine the well-posedness of the fitting problem by studying the model's identifiability [10]. Structural identifiability and practical identifiability are the two categories of identifiability that examine whether a certain set of observations can be used to uniquely determine the values of the model's unknown parameters [11–14]. Structural identifiability analysis investigates whether the unknown parameters of a model can be recovered from perfect, noise-free, and unlimited data [15,16]. Structural identifiability evaluates if parameters can be derived from the available data. Once structural identifiability is confirmed, it is necessary to establish practical identifiability of the estimated parameters [11–14]. Practical identifiability analysis, on the other hand, ensures that the estimated parameters are robust and precise, given the often limited and noisy experimental data available [12,17–19]. While structural and practical identifiabilities are rarely studied when models are fitted to data, they are necessary to guarantee that the inverse problems being solved in recovering the parameters from the given data are well-posed and that the results are reliable.

This work explores a within-host model of HIV with an immune response, studying its equilibrium points and their stability, identifiability, and parameter estimation. We investigate equilibrium points, calculate the reproduction number, and assess the equilibrium points' stability in Section 2. Section 3 focuses on the identifiability of the parameters and presents an estimation method using the data obtained from HIV-infected individuals. Finally, Section 4 discusses this study's general conclusions and outcomes.

2. Within-Host Model of HIV

HIV in a within-host model commonly considers three variables: uninfected CD4 cells, T cells that are infected, and the presence of free virus particles [3,4,7]. The following model describes the dynamics of how target cells, infected cells, viruses, and immune cells interact.

$$\begin{aligned}
\frac{dT}{dt} &= \lambda - \beta VT - dT,\\
\frac{dT_i}{dt} &= \beta VT - \delta T_i - \Psi T_i Z,\\
\frac{dZ}{dt} &= \lambda_z + bT_i Z - d_z Z,\\
\frac{dV}{dt} &= \pi T_i - cV.
\end{aligned} \tag{1}$$

In this model, the variables T, T_i, Z, and V represent specific cell quantities: target cells (specifically referring to CD4 cells), infected target cells, immune cells, and the viral load, respectively. Target cells are generated at a rate λ and are removed from the system at a rate d. Similarly, immune cells are made at a rate of λ_z and eliminated at a rate d_z. Target cells get infected at rate β, and this infection process is represented mathematically by the term βVT, which states that the rate at which target cells become infected is directly proportional to the product of the number of target cells and the number of free virus particles in the system. Infected cells are eliminated at a rate δ. In the model, the term $\Psi T_i Z$ represents the clearance of infected cells by the immune system. It indicates that the rate at which infected cells are cleared is directly proportional to the product of the number of infected cells and the number of immune cells. Immune cells are activated in response to antigens at a rate denoted by b. Infected cells produce virus particles at a rate π, representing the number of new viral particles generated by each infected cell. Meanwhile, virus particles are removed or cleared from the system at a rate c. This balance between viral production and clearance contributes to the overall dynamics of the viral load in the model. Figure 1 presents a schematic representation of the state variables and parameters. Table 1 presents the state variables of the HIV within-host model along with their corresponding definitions and units. Table 2 outlines the definitions of the parameters and units used in the HIV within-host model.

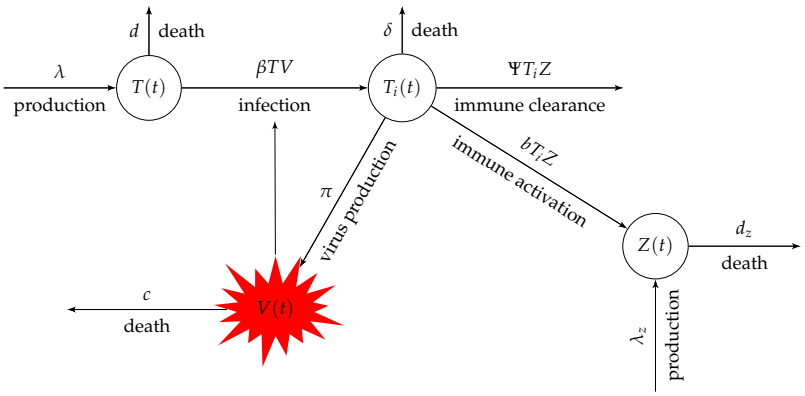

Figure 1. Flow chart of the HIV within-host model (1).

Table 1. State variables of the HIV within-host model (1) and their definitions and units. The units of state variables are determined by the data used in this study [10]. Target cells are counted in 1 µL of blood, while the viral RNA particles are measured in 1 mL of blood.

Variables	Definition	Units
$T(t)$	Number of target cells (CD4) at time t	$\frac{\text{cells}}{\mu\text{L}}$
$T_i(t)$	Number of infected cells (CD8) at time t	$\frac{\text{cells}}{\mu\text{L}}$
$Z(t)$	Number of immune cells at time t	$\frac{\text{cells}}{\mu\text{L}}$
$V(t)$	Number of viral particles at time t	$\frac{\text{RNA copies}}{\text{mL}}$

Table 2. Definitions and units of the HIV within-host model (1) parameters. Units were determined based on datasets, where target cells and infected target cells are measured in $\frac{\text{cells}}{\mu\text{L}}$ and viral load is measured in $\frac{\text{RNA copies}}{\text{mL}}$, from Figure 1(A) in [10].

Parameters	Definition	Unit
λ	Production rate of target cells	$\frac{\text{cells}}{\mu\text{L} \times \text{day}}$
β	Rate of infection of target cells by virus	$\frac{\text{mL}}{\text{RNA copies} \times \text{day}}$
d	Death rate of target cells	day^{-1}
δ	Death rate of infected cells	day^{-1}
Ψ	Death rate of infected cells by immune cells	$\frac{\mu\text{L}}{\text{cells} \times \text{day}}$
λ_z	Production rate of immune cells	$\frac{\text{cells}}{\mu\text{L} \times \text{day}}$
b	Antigen-driven proliferation rate of immune cell	$\frac{\mu\text{L}}{\text{cells} \times \text{day}}$
d_z	Death rate of immune cells	day^{-1}
π	Virus production rate by infected cells	$\frac{\text{RNA copies}}{\text{cells} \times \text{day}}$
c	Death rate of the virus	day^{-1}

2.1. Equilibria and Stability

We analyze the equilibrium points for Model (1). At equilibrium points, Model (1) satisfies the following set of equations.

$$\begin{aligned} 0 &= \lambda - \beta V^* T^* - dT^*, \\ 0 &= \beta V^* T^* - \delta T_i^* - \Psi T_i^* Z^*, \\ 0 &= \lambda_z + bT_i^* Z^* - d_z Z^*, \\ 0 &= \pi T_i^* - cV^*, \end{aligned} \tag{2}$$

where T^*, T_i^*, Z^*, and V^* represent the equilibrium values for target cells, infected target cells, immune cells, and viral load, respectively.

2.2. Stability Analysis of Infection-Free Equilibrium

The infection-free equilibrium, denoted as E_0, is a specific state where there is no infection. This occurs when both V and T_i are equal to zero. We determine the infection-free equilibrium by solving for T^* and Z^* in Equation (2). Mathematically, the infection-free equilibrium is defined as $E_0 = \left(\frac{\lambda}{d}, 0, \frac{\lambda_z}{d_z}, 0\right)$.

Theorem 1. *The infection-free equilibrium E_0 is locally asymptotically stable if $\mathcal{R}_0 < 1$, and it is unstable if $\mathcal{R}_0 > 1$. $\mathcal{R}_0$ is given by*

$$\mathcal{R}_0 = \frac{\beta \lambda \pi d_z}{dc(d_z \delta + \Psi \lambda_z)}.$$

Proof. The Jacobian matrix of the system (1) at the infection-free equilibrium E_0 is

$$J(E_0) = \begin{bmatrix} -d & 0 & 0 & \frac{-\beta\lambda}{d} \\ 0 & -\delta - \frac{\Psi\lambda_z}{d_z} & 0 & \frac{\beta\lambda}{d} \\ 0 & \frac{b\lambda_z}{d_z} & -d_z & 0 \\ 0 & \pi & 0 & -c \end{bmatrix}$$

Two of the four eigenvalues are $-d$ and $-d_z$, both of which are negative. The remaining two are the eigenvalues of the following matrix J.

$$J = \begin{bmatrix} -\delta - \frac{\Psi\lambda_z}{d_z} & \frac{\beta\lambda}{d} \\ \pi & -c \end{bmatrix}$$

The trace and determinant of the characteristic equation of matrix J are represented by the following equations:

$$Tr(J) = -\delta - \frac{\Psi\lambda_z}{d_z} - c$$

and

$$Det(J) = \delta c + \frac{c\Psi\lambda_z}{d_z} - \frac{\beta\lambda\pi}{d}$$

For an equilibrium point in a system to be locally asymptotically stable, the real parts of all the eigenvalues of the Jacobian matrix must be negative. This condition ensures that small perturbations around the equilibrium will decay over time. In two-dimensional systems, this stability criterion can be expressed more simply: the equilibrium is locally asymptotically stable if two conditions are met. First, the trace of the Jacobian matrix ($Tr(J)$) must be negative. Second, the determinant of the Jacobian matrix ($Det(J)$) must be positive. Together, these conditions imply that both eigenvalues have negative real parts, leading to the local stability of the equilibrium. Applying these conditions for stability leads to an expression for the reproduction number. It is clear that $Tr(J) < 0$. By setting $Det(J) > 0$, this condition yields a reproduction number in the form

$$\mathcal{R}_0 = \frac{\beta\lambda\pi d_z}{dc(d_z\delta + \Psi\lambda_z)}.$$

The basic reproduction number, $\mathcal{R}_0$, represents the average number of secondary infections generated by a single infected cell when the virus is introduced into a population of uninfected target cells. If $\mathcal{R}_0 < 1$, both $Tr(J) < 0$ and $Det(J) > 0$. This indicates that the real parts of all eigenvalues are negative and that, consequently, the infection-free equilibrium is locally asymptotically stable. However, if $\mathcal{R}_0 > 1$, $Det(J) < 0$, implying that there is an eigenvalues with a positive real part, which results in the infection-free equilibrium being unstable. □

The expression of the basic reproduction number has a clear interpretation as the number of secondary infections by a single infected cell during its life time when one viral particle is introduced to the population of infection-free target cells. The life span of an infected cell is $\frac{1}{\delta + \Psi Z_0}$, where Z_0 is the immune response in the absence of infection, that is $Z_0 = \frac{\lambda_z}{d_z}$. One infected cell will produce $\frac{\pi d_z}{d_z\delta + \Psi\lambda_z}$ viral particles during its lifetime. Similarly, one viral particle introduced to the system will infect $\frac{\beta T_0}{c}$ cells during its lifetime, where $T_0 = \frac{\lambda}{d}$ is the population of infection-free target cells. Therefore, one infected cell will produce $\frac{\pi d_z}{d_z\delta + \Psi\lambda_z}\frac{\beta\lambda}{cd}$ new infected cells during its time in an infection-free population.

2.3. Existence and Local Stability of Infection Equilibrium

Theorem 2. *A unique infection equilibrium E_1 of the system (1) exists when $\mathcal{R}_0 > 1$ and $V^* < \frac{\pi d_z}{bc}$.*

Proof. At the infection equilibrium $E_1 = (T^*, T_i^*, Z^*, V^*)$, all state variables have positive values. To obtain this infection equilibrium E_1, our initial step involves solving for T_i using the fourth equation in (2), allowing us to obtain

$$T_i^* = \frac{cV^*}{\pi}. \tag{3}$$

Then, we solve for Z^* using the third equation of the system (2) and substitute T_i^* from Equation (3). We obtain

$$Z^* = \frac{\pi\lambda_z}{\pi d_z - bcV^*}. \tag{4}$$

We focus on the positive state variable $Z^* > 0$. Therefore,

$$0 < V^* < \frac{\pi d_z}{bc}. \tag{5}$$

To find T^* in terms of V^*, we solve for T^* using the first equation of (2) and obtain

$$T^* = \frac{\lambda}{\beta V^* + d}. \tag{6}$$

Substituting Equations (4)–(6) into the second equation of the system (2) and simplifying and rearranging terms, we obtain the following equation:

$$\frac{\beta\lambda}{\beta V^* + d} = \frac{\delta c}{\pi} + \frac{\Psi c\lambda_z}{\pi d_z - bcV^*} \tag{7}$$

Let $y_1(V^*) = \frac{\beta\lambda}{\beta V^* + d}$ and $y_2(V^*) = \frac{\delta c}{\pi} + \frac{\Psi c\lambda_z}{\pi d_z - bcV^*}$. Since $y_1' = \frac{-\beta^2\lambda}{(\beta V^* + d)^2} < 0$ and $y_2' = \frac{\Psi c^2\lambda_z b}{(\pi d_z - bcV^*)^2} > 0$, this implies that the function $y_1(V^*)$ is decreasing and $y_2(V^*)$ is increasing for all $V^* > 0$. This implies that if a solution exists, it is necessarily unique. Next, we prove that the unique solution V^* is in the interval $(0, \frac{\pi d_z}{bc})$. In order to do that, we must firstly assign the value of $V^* = 0$ into $y_1(V^*)$ and $y_2(V^*)$. Hence,

$$y_1(0) = \frac{\beta\lambda}{d} \quad \text{and} \quad y_2(0) = \frac{\Psi c\lambda_z}{\pi d_z} + \frac{\delta c}{\pi} = \frac{c(\Psi\lambda_z + \delta d_z)}{\pi d_z}.$$

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

$$\mathcal{R}_0 = \frac{\beta\lambda\pi d_z}{dc(\Psi\lambda_z + \delta d_z)} > 1 \quad \Longrightarrow \quad \frac{\beta\lambda}{d} > \frac{c(\Psi\lambda_z + \delta d_z)}{\pi d_z}.$$

Therefore, $y_1(0) > y_2(0)$. Similarly, by substituting $V^* = \frac{\pi d_z}{bc}$ into $y_1(V^*)$ and $y_2(V^*)$, we have $y_1\left(\frac{\pi d_z}{bc}\right) = \frac{\beta\lambda bc}{\beta\pi d + dbc} > 0$—a positive value and $y_2(V^*) \to \infty$ as V^* approaches $\frac{\pi d_z}{bc}$.

Hence, $y_1\left(\frac{\pi d_z}{bc}\right) < y_2\left(\frac{\pi d_z}{bc}\right)$. Thus, this implies that $y_1(V^*)$ intersects $y_2(V^*)$ at a point in $\left(0, \frac{\pi d_z}{bc}\right)$. This shows that the model has a positive, unique infection equilibrium $E_1 = (T^*, T_i^*, Z^*, V^*)$ for $V^* < \frac{\pi d_z}{bc}$. □

Theorem 3. *Assume $\mathcal{R}_0 > 1$. Then, the infection equilibrium $E_1 = (T^*, T_i^*, Z^*, V^*)$, where $V^* < \frac{\pi d_z}{bc}$, in (1) is locally asymptotically stable.*

Proof. The Jacobian matrix of the system (1) at E_1 is

$$J(E_1) = \begin{bmatrix} -\beta V^* - d & 0 & 0 & -\beta T^* \\ \beta V^* & -\delta - \Psi Z^* & -\Psi T_i^* & \beta T^* \\ 0 & bZ^* & bT_i^* - d_z & 0 \\ 0 & \pi & 0 & -c \end{bmatrix}.$$

At point E_1, the characteristic equation is derived by setting $det|J - kI| = 0$, where k represents an eigenvalue.

$$\begin{vmatrix} -\beta V^* - d - k & 0 & 0 & -\beta T^* \\ \beta V^* & -\delta - \Psi Z^* - k & -\Psi T_i^* & \beta T^* \\ 0 & bZ^* & bT_i^* - d_z - k & 0 \\ 0 & \pi & 0 & -c - k \end{vmatrix} = 0, \tag{8}$$

In order to simplify the expression, we perform the operation of adding the first row to the second row of Equation (8); thus, we have

$$\begin{vmatrix} -\beta V^* - d - k & 0 & 0 & -\beta T^* \\ -d - k & -\delta - \Psi Z^* - k & -\Psi T_i^* & 0 \\ 0 & bZ^* & bT_i^* - d_z - k & 0 \\ 0 & \pi & 0 & -c - k \end{vmatrix} = 0. \tag{9}$$

By expanding the determinant Equation (9), we obtain the following equation:

$$\begin{aligned}(-\beta V^* - d - k)((-\delta - \Psi Z^* - k)(bT_i^* - d_z - k)(-c - k) + b\Psi T_i^* Z^*(-c - k)) + \\ \pi\beta T^*((d + k)(bT_i^* - d_z - k)) = 0,\end{aligned}$$

We rearrange the terms in the above equation and solve for $\frac{\beta V^* + d + k}{d + k}$; thus, we have

$$\frac{\beta V^* + d + k}{d + k} = \frac{\pi\beta T^*(bT_i^* - d_z - k)}{(c + k)((\delta + \Psi Z^* + k)(bT_i^* - d_z - k) - \Psi T_i^* bZ^*)},$$

By dividing both the numerator and denominator on the right-hand side of the equation by $bT_i^* - d_z - k$, we obtain

$$\frac{\beta V^* + d + k}{d + k} = \frac{\pi\beta T^*}{(c + k)\left((\delta + \Psi Z^* + k) - \frac{\Psi T_i^* bZ^*}{(bT_i^* - d_z - k)}\right)}. \tag{10}$$

To prove the stability of the infection equilibrium, we aim to show that all eigenvalues of Equation (10) have negative real parts. We employ a method of contradiction by assuming the existence of an eigenvalue k with a non-negative real part, that is, $Re(k) \geq 0$. Since $Re(k) \geq 0$, then $|\beta V^* + d + k| > |d + k|$. This implies that

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

This means that for k with $Re(k) \geq 0$, the left-hand side of Equation (10) is always grater than 1.

Let $k = x + iy$ and consider the modulus of the denominator of the right-hand side of Equation (10).

$$|c + k|\left|\delta + \Psi Z^* + k - \frac{\Psi T_i^* bZ^*}{bT_i^* - d_z - k}\right| = |c + k|\left|\delta + \Psi Z^* + x + iy - \frac{\Psi T_i^* bZ^*}{bT_i^* - d_z - x - iy}\right|,$$

To simplify the complex fraction, we multiply by the complex conjugate; thus we obtain

$$= |c+k|\left|\delta + \Psi Z^* + x + iy + \frac{\Psi T_i^* bZ^*(d_z - bT_i^* + x - iy)}{(bT_i^* - d_z - x)^2 + y^2}\right|,$$

By separating the real and imaginary parts of the equation, we obtain

$$= |c+k|\left|\delta + \Psi Z^* + x + \frac{\Psi T_i^* bZ^*(d_z - bT_i^* + x)}{(bT_i^* - d_z - x)^2 + y^2} + i\left(y - \frac{\Psi T_i^* bZ^* y}{(bT_i^* - d_z - x)^2 + y^2}\right)\right|,$$

Since $V^* < \frac{\pi d_z}{bc} \implies \frac{cV^*}{\pi} < \frac{d_z}{b} \implies T_i^* < \frac{d_z}{b} \implies d_z - bT_i^* > 0,$

$$\geq |c+k|\left|\delta + \Psi Z^* + x + \frac{\Psi T_i^* bZ^*(d_z - bT_i^* + x)}{(bT_i^* - d_z - x)^2 + y^2}\right| \geq |c+k||\delta + \Psi Z^*| \geq c(\delta + \Psi Z^*),$$

Thus, the modulus of the right-hand side of Equation (10) is

$$\left|\frac{\pi\beta T^*}{(c+k)\left((\delta + \Psi Z^* + k) - \frac{\Psi T_i^* bZ^*}{(bT_i^* - d_z - k)}\right)}\right| \leq \frac{\pi\beta T^*}{c(\delta + \Psi Z^*)},$$

From the second and fourth equations of model (2), $\frac{\pi\beta T^*}{c(\delta + \Psi Z^*)} = 1$, which implies that the modulus of the right-hand side of Equation (10) is less than 1, that is,

$$\left|\frac{\pi\beta T^*}{(c+k)\left((\delta + \Psi Z^* + k) - \frac{\Psi T_i^* bZ^*}{(bT_i^* - d_z - k)}\right)}\right| \leq \frac{\pi\beta T^*}{c(\delta + \Psi Z^*)} = 1.$$

This implies that for k with $Re(k) \geq 0$, the left-hand side of Equation (10) is always greater than the right-hand side. This leads to a contradiction. Therefore, the eigenvalues have negative real parts. Thus, we have proven the theorem. □

3. Structural and Practical Identifiability Analysis and Parameter Estimation

Structural and practical identifiability analyses are two crucial aspects of mathematical modeling. Structural identifiability analysis investigates whether a model can uniquely reveal its unknown parameters given perfect, noise-free, and unlimited data. This theoretical analysis, performed prior to collecting any experimental data, involves methods such as the differential algebra approach, generating series approach, Taylor series approach, and a method based on the Implicit Function Theorem [15,20–23].

Practical identifiability analysis further ensures that the estimated parameters are robust and precise given the available experimental data, which is often limited and noisy. Similar to structural identifiability, it is important to see if the model's structure causes practical identifiability problems. Techniques such as Monte Carlo simulations, sensitivity analysis, Bayesian methods, correlation matrix, and profile likelihoods method are employed to assess the robustness and precision of parameter estimates [12,15,24–26]. We employ the differential algebra approach for structural identifiability analysis and Monte Carlo simulations to evaluate practical identifiability [11,13,25].

3.1. Structural Identifiability

Analyzing the structural identifiability of the within-host model, we rewrite the model (1) in the following compact form.

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

where t represents time, the vector $x = (T, T_i, Z, V)$ represents the dependent variables, $p = (\lambda, \beta, d, \delta, \Psi, \lambda_z, b, d_z, \pi, c)$ denotes the model's parameters, and $y(t)$ is the model output, corresponding to the observations. The observations of this study include the count of the uninfected and infected target cells denoted as $y_1(t)$, the immune cells as $y_2(t)$, and the viral load as $y_3(t)$. Therefore, the expressions of the observations are as follows.

$$y_1(t) = T(t) + T_i(t), \quad y_2(t) = Z(t), \quad \text{and} \quad y_3(t) = V(t).$$

The within-host model (1) is considered identifiable if the parameter vector p can be uniquely determined from the system output $y(t)$. Structural identifiability is defined as follows.

Definition 1. *Let p and $\hat{p}$ be distinct model parameter vectors. The within-host model is structurally identifiable if*

$$g(x, p) = g(x, \hat{p}) \quad \textit{implies} \quad p = \hat{p}.$$

The differential algebra approach requires the removal of state variables for which data are not given, resulting in the model being expressed as a function of both the parameters of the model and the state variables for which data are given, referred to as the input–output equations. These input–output equations are expressed as algebraic polynomials involving the outputs $y_1(t)$, $y_2(t)$, and $y_3(t)$, their derivatives, and their coefficients composed of parameters of the model. We derive the input–output equations for the within-host model (1) based on the observations $y_1(t)$, $y_2(t)$, and $y_3(t)$ using DAISY (Differential Algebra for Identifiability of SYstem, https://daisy.org/ Mountain View, CA, USA) software [20].

$$0 = y_1' y_2 + y_2' y_2 \frac{\Psi}{b} + y_2' \frac{(\delta - d)}{b} + y_1 y_2 d + y_2^2 \frac{d_z \Psi}{b} + y_2 \left(-\lambda - \frac{d d_z}{b} + \frac{\delta d z}{b} - \lambda_z \frac{\Psi}{b} \right) + \lambda_z \frac{(d - \delta)}{b}. \tag{11}$$

$$0 = y_1' + y_3' y_2 \frac{\Psi}{\pi} + y_3' \frac{(\delta - d)}{\pi} + y_1 d + y_2 y_3 \frac{c\Psi}{\pi} + y_3 \frac{c(\delta - d)}{\pi} - \lambda. \tag{12}$$

$$0 = y_1' y_2 + y_1 y_3 \frac{\beta}{\Psi} + y_1' \frac{(c + \delta)}{\Psi} - y_3'' y_2 \frac{1}{\pi} + y_3'' \frac{(d - \delta)}{\pi \Psi} + y_1 y_2 y_3 \beta + y_1 y_2 d + y_1 y_3 \frac{\beta \delta}{\Psi} + y_1 \frac{d(c + \delta)}{\Psi} + y_2 y_3 \frac{c^2}{\pi} - y_2 \lambda - y_3 \left(\frac{\beta \lambda}{\Psi} + \frac{c^2 d}{\pi \Psi} - \frac{c^2 \delta}{\pi \Psi} \right) - \lambda \frac{(c + \delta)}{\pi \Psi}. \tag{13}$$

The definition of identifiability for a model based on the differential algebra approach is provided below.

Definition 2. *Let $c(p)$ represent the coefficients of the input–output Equations (11)–(13). The within-host model is structurally identifiable from observations $y_1(t)$, $y_2(t)$, and $y_3(t)$ if and only if*

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

To illustrate, assume the existence of another parameter vector, denoted as $\hat{p} = (\hat{\lambda}, \hat{\beta}, \hat{d}, \hat{\delta}, \hat{\Psi}, \hat{\lambda_z}, \hat{b}, \hat{d_z}, \hat{\pi}, \hat{c})$. Establishing a one-to-one mapping between the parameter space and the coefficients of the input–output equations is essential. Therefore, we derive the following set of equations:

$$d = \hat{d}, \quad \lambda = \hat{\lambda}, \quad \pi = \hat{\pi}, \quad \beta = \hat{\beta}, \quad \frac{\beta}{\Psi} = \frac{\hat{\beta}}{\hat{\Psi}}, \quad \frac{\Psi}{b} = \frac{\hat{\Psi}}{\hat{b}}, \quad \frac{\delta - d}{b} = \frac{\hat{\delta} - \hat{d}}{\hat{b}}$$
$$\frac{d_z \Psi}{b} = \frac{\hat{d}_z \hat{\Psi}}{\hat{b}}, \quad \frac{d(c+\delta)}{\Psi} = \frac{\hat{d}(\hat{c}+\hat{\delta})}{\hat{\Psi}}, \quad \frac{dd_z}{b} + \frac{\delta d_z}{b} - \frac{\lambda d_z \Psi}{b} = \frac{\hat{d}\hat{d}_z}{\hat{b}} + \frac{\hat{\delta}\hat{d}_z}{\hat{b}} - \frac{\hat{\lambda}_z \hat{\Psi}}{\hat{b}}.$$

After solving the nonlinear equation with Definition 2 described above, we find a collection of positive solutions, as outlined below.

$$\{\lambda = \hat{\lambda}, \quad \beta = \hat{\beta}, \quad d = \hat{d}, \quad \delta = \hat{\delta}, \quad \Psi = \hat{\Psi}, \quad \lambda_z = \hat{\lambda}_z, \quad b = \hat{b}, \quad d_z = \hat{d}_z, \quad \pi = \hat{\pi}, \quad c = \hat{c}\}$$

As a result, the within-host model (1) is structurally identifiable. The model reveals its parameters $\lambda, \beta, d, \delta, \Psi, \lambda_z, b, d_z, \pi$, and c from the observations of the CD4/CD8 cell counts and viral load. Structural identifiability analysis is summarized in the following proposition.

Proof. The within-host model (1) is structurally identifiable based on the given observations of the CD4/CD8 cell counts and viral load. □

Parameter Estimation

In this study, we gathered data from Figure 1(A) in [10]. Data points were extracted from Figure 1(A) in [10] using the Grabit tool in MATLAB. Table 3 presents the logarithmic values of the obtained data. We fit the model (1) to CD4/CD8 cell counts and viral load data. We apply the least squares (LS) principle to estimate the parameters. This principle involves minimizing the objective function defined by Equation (15). Each observable value, denoted as $y_1(t_i)$, $y_2(t_j)$, and $y_3(t_k)$, is obtained at discrete time points t_i, t_j, and t_k. The measurements include inherent measurement noise, represented by ϵ_i, ϵ_j, and ϵ_k, respectively. The relationship between observations and measurements is defined as follows:

$$\begin{aligned} Y_1^i &= T(t_i) + T_i(t_i) + (T(t_i) + T_i(t_i))\epsilon_i, \quad i = 1, 2, \ldots, n \\ Y_2^j &= Z(t_j) + Z(t_j)\epsilon_j, \quad j = 1, 2, \ldots, m \\ Y_3^k &= V(t_k) + V(t_k)\epsilon_k. \quad k = 1, 2, \ldots, l \end{aligned} \tag{14}$$

The measurement errors are normally distributed with zero mean and variances, where

$$\epsilon_i \sim \mathcal{N}(0, \sigma_i^2), \quad \epsilon_j \sim \mathcal{N}(0, \sigma_j^2), \quad \epsilon_k \sim \mathcal{N}(0, \sigma_k^2)$$

In our statistical model, the expected values and variances for Y_1^i, Y_2^j, and Y_3^k are, respectively, given by [27]

$$\begin{aligned} \mathbb{E}(Y_1^i) &= T(t_i) + T_i(t_i) && \text{and} \quad \mathrm{Var}(Y_1^i) = (T(t_i) + T_i(t_i))^2 \sigma_i^2 \\ \mathbb{E}(Y_2^j) &= Z(t_j) && \text{and} \quad \mathrm{Var}(Y_2^j) = Z^2(t_j)\sigma_j^2 \\ \mathbb{E}(Y_3^k) &= V(t_k) && \text{and} \quad \mathrm{Var}(Y_3^k) = V^2(t_k)\sigma_k^2 \end{aligned}$$

To estimate the $\hat{p}$ parameters, we minimize the following objective function [27].

$$\begin{aligned} \hat{p} = \min_p \Bigg(& \frac{1}{n} \sqrt{\sum_{i=1}^{n} \left(\frac{\log_{10}(T(t_i) + T_i(t_i)) - \log_{10} Y_1^i}{\log_{10}(T(t_i) + T_i(t_i))} \right)^2} + \frac{1}{m} \sqrt{\sum_{j=1}^{m} \left(\frac{\log_{10} Z(t_j) - \log_{10} Y_2^j}{\log_{10} Z(t_j)} \right)^2} \\ & + \frac{1}{l} \sqrt{\sum_{k=1}^{l} \left(\frac{\log_{10} V(t_k) - \log_{10} Y_3^k}{\log_{10} V(t_k)} \right)^2} \Bigg), \quad \delta - d \geq 0 \end{aligned} \tag{15}$$

where n, m, and l represent the number of data points for CD4/CD8 cells and viral load, respectively. $T(t_i)$, $Z(t_j)$, and $V(t_k)$ are the solutions to the system (1). In our model, we introduce a constraint $\delta - d \geq 0$ to balance between the death rate of uninfected target cells d and the death rate of infected cells δ. This constraint ensures that the natural decay rate of uninfected cells does not exceed that of infected cells. Additionally, we establish lower and upper limits for each parameter in our model, which are listed in Table 4. We estimate the parameters and initial conditions using the least squares fitting (fmincon) in MATLAB to the data. The initial conditions for the dependent variables were fixed as $(T(0), T_i(0), Z(0), V(0)) = (918, 0, 583, 1003)$. The fitting iteration continues until the error stabilizes, reaching the optimization tolerance. The best parameter fits are given in Table 4. Figure 2 shows the graphs of the fitted curves for viral load, target cells, and immune cells.

Table 3. Logarithms of the viral loads (RNA copies/mL), CD4 cell counts (cells/μL), and CD8 cell counts (cells/μL)) are presented.

	Day 2	Day 6	Day 10	Day 14	Day 18	Day 21	Day 25	Day 28
Viral load (RNA copies/mL)	3.5521	5.0611	5.9734	6.3262	6.0449	5.5590	5.2032	4.8850
	Day 32	Day 41	Day 49	Day 67	Day 96	Day 179	Day 259	
Viral load (RNA copies/mL)	4.6970	4.6190	4.5042	4.4079	4.3850	4.1678	3.8765	
	Day 2	Day 18	Day 32	Day 49	Day 90	Day 257		
CD4 count (cells/μL)	2.957	2.7758	2.7944	2.8526	2.8401	2.7805		
	Day 2	Day 18	Day 32	Day 49	Day 90	Day 255		
CD8 count (cells/μL)	2.7507	3.0818	3.0541	3.0367	2.9906	2.9689		

Table 4. Estimated parameter values of the model (1) derived by solving the least squares problem given in (15).

Parameters	**Parameter Space**	**Value**
λ	$(0, 100)$	96.7
β	$(0, 1)$	2.3×10^{-7}
d	$(0.01, 1)$	0.13
δ	$(0.1, 1)$	0.31
Ψ	$(0, 10)$	0.002
λ_z	$(0, 100)$	4.23
b	$(0, 1)$	0.006
d_z	$(0.01, 1)$	0.027
π	$(0, 5 \times 10^5)$	12,725.36
c	$(0, 100)$	0.68

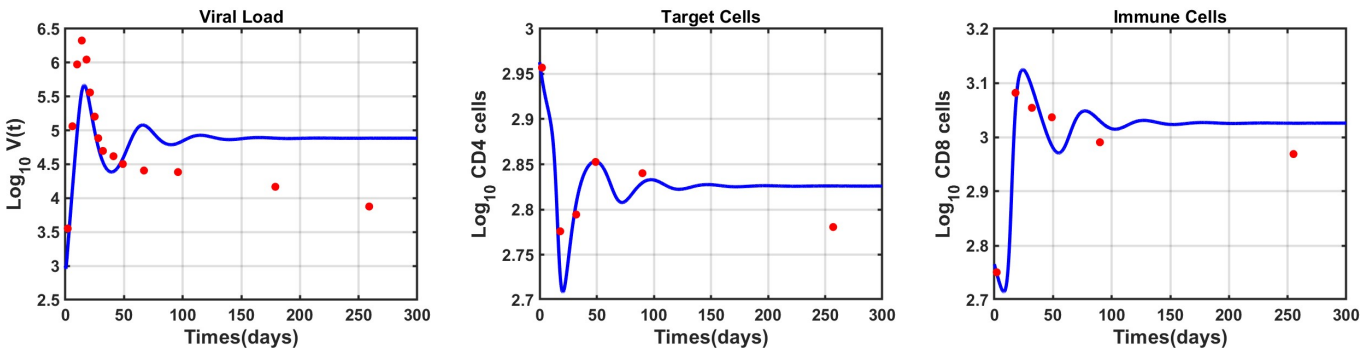

Figure 2. Logarithms of viral load, CD4 cells, and immune cells (red dots) plotted along the solutions of the model (blue curves) with the estimated parameter values in Table 4. The initial values are $(T(0), T_i(0), Z(0), V(0)) = (918, 0, 583, 1003)$.

3.2. Practical Identifiability

We have shown that the within-host model of HIV (1) is structurally identifiable if observations of the uninfected and infected target cell counts, immune cells, and viral load are given. Structural identifiability refers to a characteristic of the model's structure concerning the specific output it produces, assuming the model does not have errors, and the output is not affected by noise. Practical parameter identification is not solely determined by the model's structure but also by factors such as the quantity and quality of the data, as well as the specific numerical optimization algorithm used for parameter estimation (15). If the model lacks structural identifiability, it is not practically identifiable either. However, even if a model is structurally identifiable, it might still be practically non-identifiable [28]. Here, we explore the practical identifiability of the HIV within-host model (1) to evaluate the reliability of estimated parameters. Multiple techniques, including Monte Carlo simulations (MCS), are employed to evaluate the practical identifiability of parameters in an ODE model [12,15,29,30]. We conduct Monte Carlo simulations by generating the 1000 synthetic datasets with the estimated parameters viewed as "true" parameter set $\hat{p}$ and including noise at progressively growing levels 1%, 5%, 10%, and 20%. We conduct Monte Carlo simulations according to the following steps.

1. We numerically solve the model using the true parameters $\hat{p}$ and collect the output vector $g(x, \hat{p})$ at the specific discrete experimental time points.
2. We generate 1000 datasets from the statistical model (14) with a given measurement error. These datasets are generated from a normal distribution with the mean corresponding to the output vector obtained in step 1, represented as $E(Y_i) = g(x(t_i), \hat{p})$. The standard deviation is calculated as a σ_0% of the mean using the formula $Var(Y_i) = g(x(t_i), \hat{p})^2\sigma^2$. Figure 3 presents a series of 12 graphs illustrating the impact of varying measurement errors on the distribution of the generated datasets. Each graph corresponds the viral load, CD4 cell count, and CD8 cell count to a different value of σ.
3. We approximate the parameter set p_q by fitting the within-host model (1) to all simulated datasets.

$$p_q = \min_p \left(\frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(\frac{\log_{10}(T(t_i)+T_i(t_i)) - \log_{10} Y_1^i}{\log_{10}(T(t_i)+T_i(t_i))}\right)^2} + \frac{1}{m}\sqrt{\sum_{j=1}^{m}\left(\frac{\log_{10} Z(t_j) - \log_{10} Y_2^j}{\log_{10} Z(t_j)}\right)^2} + \frac{1}{l}\sqrt{\sum_{k=1}^{l}\left(\frac{\log_{10} V(t_k) - \log_{10} Y_3^k}{\log_{10} V(t_k)}\right)^2}\right), \quad q = 1, ..., 1000. \tag{16}$$

4. Next, we compute the average relative estimation error (ARE) for each parameter in the model via

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

where $p^{(k)}$ represents the kth element of p, $\hat{p}^{(k)}$ denotes the kth parameter in the true parameter set $\hat{p}$, and $\hat{p}_q^{(k)}$ indicates the kth parameter in the set p_q.
5. We repeat steps 1 to 5, gradually increasing the level of noise by considering σ values of 1%, 5%, 10%, and 20%.

The average relative estimation errors $(AREs)$ obtained are valuable for understanding the identifiability of the parameters of the within-host model. When the model is structurally identifiable, meaning its parameters can be accurately determined from the data, we expect AREs to approach zero when the measurement error σ is zero. In noise-free data, $AREs$ ideally should be minimal or close to zero. However, if a parameter is not practically identifiable, its ARE will be significantly elevated even with reasonable measurement error levels. Some parameters may be highly sensitive to data noise, resulting in disproportionately high $AREs$. In such cases, increasing measurement errors will cause

the *AREs* to rise significantly, indicating that the parameter is practically unidentifiable. Practical identifiability is defined in the following way [31].

Definition 3. *The practical identifiability of a parameter p is determined by comparing its average relative estimation error ARE with the measurement error σ. If*

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

then the parameter p is considered (strongly) practically identifiable. If

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

then p is weakly practically identifiable. However, if $ARE(p)$ *exceeds* 10σ*, then p is practically unidentifiable.*

We apply the Monte Carlo simulation algorithm in MATLAB and calculate the *AREs* of all parameters of the model (1). Table 5 presents the results of Monte Carlo simulations (MCSs) conducted with actual data points presented in Table 3. According to the *ARE* values presented in Table 5, it is observed that parameters λ, d, δ, and π are practically identifiable. On the other hand, parameters λ_z, d_z, and c exhibit weak identifiability. Parameters β, Ψ, and b are not identifiable.

Table 5 that shows three parameters (β, Ψ, and b) are practically non-identifiable. With measurement error levels of 1% and 5%, all parameters show average relative estimation errors that meet the practical identifiability criteria. This suggests that the simulation effectively estimates the parameters even with minimal noises, ensuring practical identifiability. However, when the noise level increases to 10% and 20%, three parameters (β, Ψ, and b) become highly sensitive to noise, and their ARE values increase rapidly. Additionally, experimental data are frequently obtained at a low frequency, which leads to few data points. This low frequency of data collection further complicates the practical identifiability of parameters. To understand how data collection frequency impacts the practical identifiability of a model's parameters, we conducted Monte Carlo simulations and increased the total number of data points. While high-frequency data collection may not reflect real-world experimental conditions, evaluating the accuracy of our estimates is essential. Initially, high-frequency data were simulated at specific time intervals over a period ranging from $t = 1$ to $t = 300$ days, resulting in 3000 data points. These data points were derived through curve fitting, providing estimated true parameter values. Subsequently, we determined the ARE values for all 10 parameters across four levels of measurement errors, as outlined in Table 6. The Monte Carlo simulation using high-frequency data provided valuable insights into the identifiability of parameters. At a measurement error level of $\sigma = 1\%$, all parameters exhibited average relative estimation errors *AREs* well within the bounds of practical identifiability. This indicates that even with a low level of noise, the simulation accurately estimates the parameters, making them practically identifiable. The results remain consistent when moving to a slightly higher noise level of $\sigma = 5\%$. Despite the increased noise, all parameters maintained *ARE* values that fell within the acceptable range for identifiability. At a measurement error level of $\sigma = 10\%$, all parameters were identifiable. At a measurement error level of $\sigma = 20\%$, parameters such as λ, β, d, δ, Ψ, λ_z, b, d_z, π, and c showed practical identifiability. Parameters become identifiable at a high frequency, which suggests that the identifiability of parameters may vary depending on the frequency of the data.

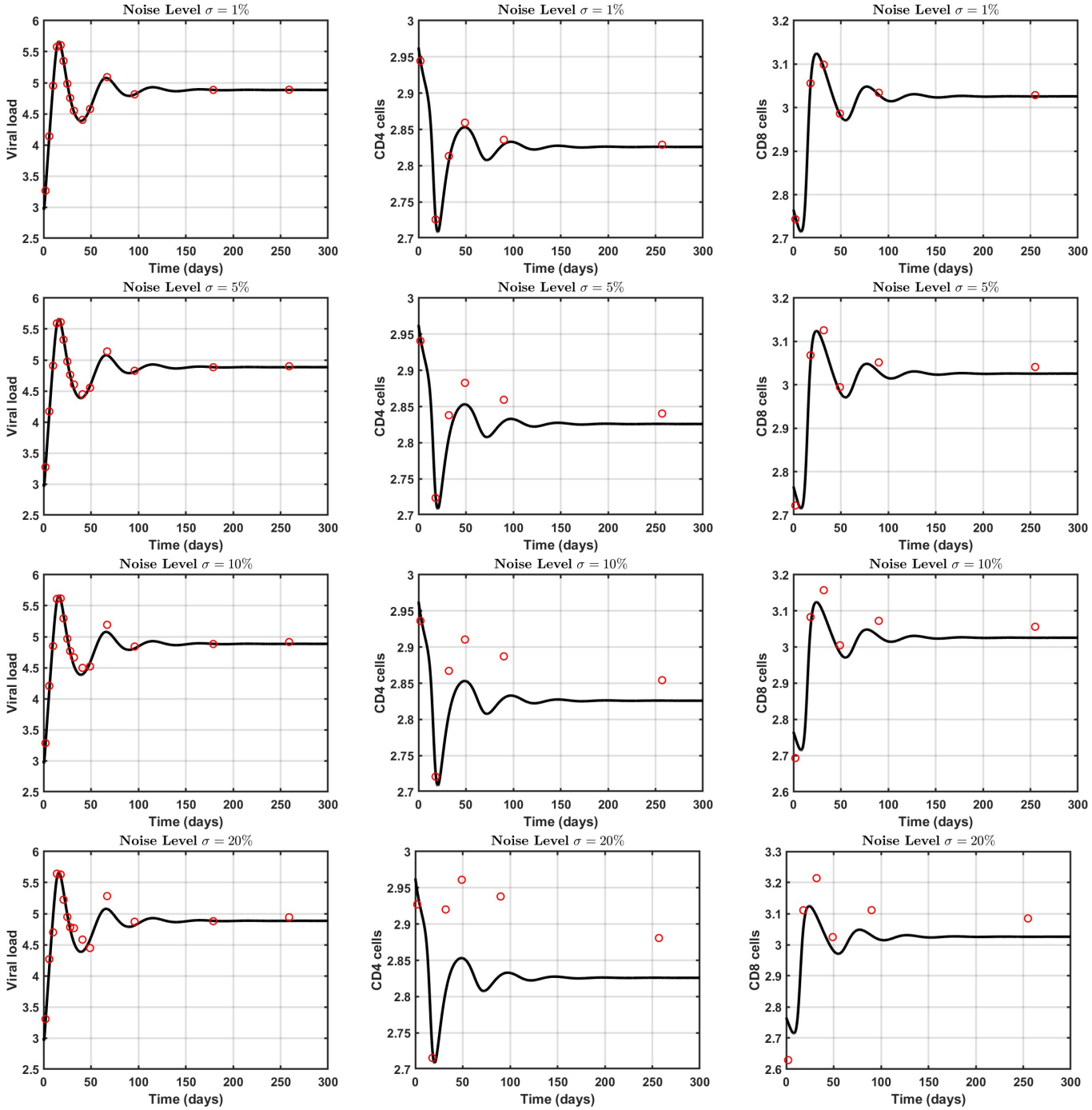

Figure 3. Comparison of within-host model predictions (black curve) and data observations (red dots) for viral load, CD4 cell count, and CD8 cell count at different noise levels (σ = 1%, 5%, 10%, and 20%) at a logarithmic scale.

Table 5. Monte Carlo simulation *MCS* results: Absolute relative error *ARE* for each parameter within the within-host model (1) computed based on the actual data frequency.

Parameter	λ	β	d	δ	Ψ	λ_z	b	d_z	π	c
ARE $\sigma = 1\%$	2×10^{-6}	0.07	0.01	0.004	0.02	4×10^{-5}	0.02	0.003	1×10^{-10}	0.01
ARE $\sigma = 5\%$	4×10^{-5}	0.7	0.3	0.07	0.3	6×10^{-4}	0.28	0.095	2×10^{-9}	0.2
ARE $\sigma = 10\%$	0.009	27	1.8	1.3	310.7	2	27.7	12	8×10^{-5}	1.5
ARE $\sigma = 20\%$	0.54	1618	10.9	11.7	1957	34.8	537.7	199	0.002	54

Table 6. The Monte Carlo simulation uses high-frequency data, which means that data are collected ten times per day. In total, 3000 data points gathered over 300 days are utilized for the Monte Carlo simulation.

Parameter	λ	β	d	δ	Ψ	λ_z	b	d_z	π	c
ARE $\sigma = 1\%$	2×10^{-8}	0.007	1×10^{-4}	4×10^{-5}	2×10^{-4}	4×10^{-7}	5×10^{-5}	8×10^{-6}	2×10^{-12}	1×10^{-4}
ARE $\sigma = 5\%$	2×10^{-7}	0.04	0.001	4×10^{-4}	0.002	3×10^{-6}	0.001	8×10^{-5}	2×10^{-11}	0.002
ARE $\sigma = 10\%$	7×10^{-7}	0.098	0.005	0.001	0.01	1×10^{-5}	0.007	3×10^{-4}	8×10^{-11}	0.006
ARE $\sigma = 20\%$	1.5×10^{-5}	0.28	0.09	0.02	0.16	1×10^{-4}	0.07	0.004	9×10^{-10}	0.1

4. Discussion

In this study, we looked into how target cells T, infected target cells T_i, virus particles V, and immune cells Z interact. We examined the equilibrium points and their stability for our model (1). We found that our model only has an infection-free equilibrium if $\mathcal{R}_0 < 1$, and the model has two equilibrium points when $\mathcal{R}_0 > 1$: the infection-free equilibrium E_0, where there are no infected cells or virus, and an infection equilibrium E_1, where all variables have positive values. Focusing on the equilibrium point E_0, we analyzed its stability for our model (1). The analysis revealed that E_0 is locally asymptotically stable when the basic reproduction number $\mathcal{R}_0 < 1$ and becomes unstable when $\mathcal{R}_0 > 1$. For the infection equilibrium E_1, we were able to show that it is unique. Further, the infection equilibrium is locally asymptotically stable whenever it exists.

We also investigated the structural and practical identifiabilities of the within-host model for HIV dynamics using the differential algebra method for structural identifiability and Monte Carlo simulations to evaluate practical identifiability. The analysis revealed several key findings regarding parameter identifiability and the implications for understanding HIV infection dynamics. In this study, we employed the differential algebra method to analyze the structural identifiability of the within-host HIV model. The results suggested that the model is structurally identifiable, meaning the parameters can be uniquely determined from the observations of the sum of uninfected and infected target cells, the immune cells, and the viral load. We proceeded to estimate all parameters of (1) by using the least squares approach applied to the logarithmic values of the average viral load RNA (copies/mL), average CD4 cell counts (cells/μL), and average CD8 cell counts (cells/μL). Employing the least squares principle, we addressed the parameter estimation problem by minimizing the objective function defined by Equation (15), resulting in the parameter values presented in Table 4. Practical identifiability, on the other hand, deals with the reality of limited and noisy data. Even if a model is structurally identifiable, practical identifiability must be assessed to ensure that the parameters can still be estimated accurately under realistic conditions. Practical identifiability analysis involved generating synthetic datasets with varying levels of measurement noise and assessing the robustness and precision of the parameter estimates. For practical identifiability analysis, the MCS algorithm was employed to generate 1000 sets

of data with varying levels of measurement noise for CD4 cell counts, CD8 cell counts, and viral load. Each dataset at each noise level was fitted to recover 1000 sets of parameters. The average relative error (ARE) for each parameter was then calculated. Monte Carlo simulations were conducted to evaluate the practical identifiability of the model parameters under different data collection frequencies and noise levels. Initial simulations with actual data points revealed that parameters λ, d, δ, and π are identifiable, while parameters λ_z, d_z, and c are weakly identifiable, and parameters β, Ψ, and b are not identifiable. To investigate further, we simulated high-frequency data collection over 300 days, resulting in 3000 data points. At a low measurement error level $\sigma = 1\%$, all parameters exhibited average relative estimation errors $AREs$ within the bounds of practical identifiability, indicating practical identifiability. This trend continued with slightly higher noise levels $\sigma = 5\%$, where all parameters maintained acceptable ARE values. Even at a higher noise level $\sigma = 10\%$, all parameters were still practically identifiable. At the highest noise level tested $\sigma = 20\%$, all parameters remained practically identifiable, although with varying degrees of precision. This high-frequency data simulation aimed to provide a more precise estimation of the parameters and to highlight the importance of data collection frequency. The findings in this study emphasize the importance of optimal data sampling in HIV clinical studies. By analyzing different sampling frequencies and their impact, we provide insights into how sampling strategies can enhance the applicability of mathematical models in infectious disease modeling. Furthermore, the estimated parameters can be used as a baseline for HIV treatment models, which then can be used to evaluate and compare different treatment strategies. By simulating how different therapies impact viral dynamics, the HIV treatment model can be used in optimizing treatment plans and assessing potential outcomes, such as viral suppression.

Overall, our findings highlight the relationship between model structure, data quality, and parameter identifiability in modeling HIV dynamics. By clarifying the identifiability of model parameters, our study offers insights into model predictions and the limitations of parameter estimation methods in HIV research.

Author Contributions: Conceptualization, M.M. and N.T.; Methodology, M.M. and N.T.; Software, Y.R.L., L.M.K. and N.T.; Validation, N.T.; Formal analysis, Y.R.L., L.M.K., M.M. and N.T.; Writing—original draft, L.M.K. and N.T.; Writing—review & editing, Y.R.L. and M.M.; Supervision, N.T.; Funding acquisition, M.M. and N.T. All authors have read and agreed to the published version of the manuscript.

Funding: N.T. and Y.R.L acknowledge partial support from National Institute of Health grant, NIH NIGMS 1R01GM152743-01.

Data Availability Statement: The data for the research is available at https://github.com/LeilaKohan/HIV-Within-Host-model.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Joint United Nations Programme on HIV/AIDS (UNAIDS). Global HIV & AIDS Statistics—Fact Sheet. Available online: https://www.unaids.org/en/resources/fact-sheet (accessed on 20 August 2024).
2. Nowak, M.A.; Bangham, C.R. Population Dynamics of Immune Responses to Persistent Viruses. *Science* **1996**, *272*, 74–79. [CrossRef] [PubMed]
3. Perelson, A.S.; Kirschner, D.E.; De Boer, R. Dynamics of HIV Infection of CD4+ T Cells. *Math. Biosci.* **1993**, *114*, 81–125. [CrossRef] [PubMed]
4. Perelson, A.S.; Neumann, A.U.; Markowitz, M.; Leonard, J.M.; Ho, D.D. HIV-1 Dynamics In Vivo: Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time. *Science* **1996**, *271*, 1582–1586. [CrossRef] [PubMed]
5. Perelson, A.S.; Nelson, P.W. Mathematical Analysis of HIV-1 Dynamics In Vivo. *SIAM Rev.* **1999**, *41*, 3–44. [CrossRef]
6. Rong, L.; Gilchrist, M.A.; Feng, Z.; Perelson, A.S. Modeling Within-Host HIV-1 Dynamics and the Evolution of Drug Resistance: Trade-Offs Between Viral Enzyme Function and Drug Susceptibility. *J. Theor. Biol.* **2007**, *247*, 804–818. [CrossRef]
7. Nowak, M.; May, R.M. *Virus Dynamics: Mathematical Principles of Immunology and Virology*; Oxford University Press: Oxford, UK, 2000.
8. Conway, J.M.; Ribeiro, R.M. Modeling the Immune Response to HIV Infection. *Curr. Opin. Syst. Biol.* **2018**, *12*, 61–69. [CrossRef]

9. D'Orso, I.; Forst, C.V. Mathematical Models of HIV-1 Dynamics, Transcription, and Latency. *Viruses* **2023**, *15*, 2119. [CrossRef]
10. Eller, M.A.; Goonetilleke, N.; Tassaneetrithep, B.; Eller, L.A.; Costanzo, M.C.; Johnson, S.; Betts, M.R.; Krebs, S.J.; Slike, B.M.; Nitayaphan, S.; et al. Expansion of Inefficient HIV-Specific CD8 T Cells During Acute Infection. *J. Virol.* **2016**, *90*, 4005–4016. [CrossRef] [PubMed]
11. Tuncer, N.; Gulbudak, H.; Cannataro, V.L.; Martcheva, M. Structural and Practical Identifiability Issues of Immuno-Epidemiological Vector-Host Models with Application to Rift Valley Fever. *Bull. Math. Biol.* **2016**, *78*, 1796–1827. [CrossRef]
12. Tuncer, N.; Le, T.T. Structural and Practical Identifiability Analysis of Outbreak Models. *Math. Biosci.* **2018**, *299*, 1–18. [CrossRef]
13. Tuncer, N.; Martcheva, M.; LaBarre, B.; Payoute, S. Structural and Practical Identifiability Analysis of Zika Epidemiological Models. *Bull. Math. Biol.* **2018**, *80*, 2209–2241. [CrossRef] [PubMed]
14. Chowell, G.; Dahal, S.; Liyanage, Y.R.; Tariq, A.; Tuncer, N. Structural Identifiability Analysis of Epidemic Models Based on Differential Equations: A Tutorial-Based Primer. *J. Math. Biol.* **2023**, *87*, 79. [CrossRef] [PubMed]
15. Miao, H.; Xia, X.; Perelson, A.S.; Wu, H. ON Identifiability of Nonlinear ODE Models and Applications in Viral Dynamics. *SIAM Rev. Soc. Ind. Appl. Math.* **2011**, *53*, 3–39. [CrossRef] [PubMed]
16. Dankwa, E.A.; Brouwer, A.F.; Donnelly, C.A. Structural identifiability of compartmental models for infectious disease transmission is influenced by data type. *Epidemics* **2022**, *41*, 100643. [CrossRef]
17. Renardy, M.; Kirschner, D.; Eisenberg, M. Structural Identifiability Analysis of Age-Structured PDE Epidemic Models. *J. Math. Biol.* **2022**, *84*, 9. [CrossRef]
18. Massonis, G.; Banga, J.R.; Villaverde, A.F. Structural Identifiability and Observability of Compartmental Models of the COVID-19 Pandemic. *Annu. Rev. Control* **2021**, *51*, 441–459. [CrossRef]
19. Gallo, L.; Frasca, M.; Latora, V.; Russo, G. Lack of practical identifiability may hamper reliable predictions in COVID-19 epidemic models. *Sci. Adv.* **2022**, *8*, eabg5234. [CrossRef]
20. Bellu, G.; Saccomani, M.P.; Audoly, S.; D'Angiò, L. DAISY: A New Software Tool to Test Global Identifiability of Biological and Physiological Systems. *Comput. Methods Programs Biomed.* **2007**, *88*, 52–61. [CrossRef]
21. Eisenberg, M.C.; Robertson, S.L.; Tien, J.H. Identifiability and estimation of multiple transmission pathways in Cholera and waterborne disease. *J. Theor. Biol.* **2013**, *324*, 84–102. [CrossRef]
22. Ljung, L.; Glad, T. On Global Identifiability for Arbitrary Model Parametrizations. *Automatica* **1994**, *30*, 265–276. [CrossRef]
23. Pohjanpalo, H. System Identifiability Based on the Power Series Expansion of the Solution. *Math. Biosci.* **1978**, *41*, 21–33. [CrossRef]
24. Walter, E.; Pronzato, L.; Norton, J. *Identification of Parametric Models from Experimental Data*; Springer: Berlin/Heidelberg, Germany, 1997; Volume 1.
25. Ciupe, S.M.; Tuncer, N. Identifiability of Parameters in Mathematical Models of SARS-CoV-2 Infections in Humans. *Sci. Rep.* **2022**, *12*, 14637. [CrossRef] [PubMed]
26. Raue, A.; Kreutz, C.; Maiwald, T.; Bachmann, J.; Schilling, M.; Klingmüller, U.; Timmer, J. Structural and Practical Identifiability Analysis of Partially Observed Dynamical Models by Exploiting the Profile Likelihood. *Bioinformatics* **2009**, *25*, 1923–1929. [CrossRef] [PubMed]
27. Banks, H.; Hu, S.; Thompson, W. *Modeling and Inverse Problems in the Presence of Uncertainty*; CRC Press: Boca Raton, FL, USA, 2014.
28. Liyanage, Y.R.; Heitzman-Breen, N.; Tuncer, N.; Ciupe, S.M. Identifiability Investigation of Within-Host Models of Acute Virus Infection. *bioRxiv* **2024**. [CrossRef]
29. Wu, H.; Zhu, H.; Miao, H.; Perelson, A.S. Parameter Identifiability and Estimation of HIV/AIDS Dynamic Models. *Bull. Math. Biol.* **2008**, *70*, 785–799. [CrossRef]
30. Raue, A.; Karlsson, J.; Saccomani, M.P.; Jirstrand, M.; Timmer, J. Comparison of Approaches for Parameter Identifiability Analysis of Biological Systems. *Bioinformatics* **2014**, *30*, 1440–1448. [CrossRef]
31. Timsina, A.N.; Liyanage, Y.R.; Martcheva, M.; Tuncer, N. A Novel Within-Host Model of HIV and Nutrition. *Math. Biosci. Eng.* **2024**, *21*, 5577–5603. [CrossRef]