Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Selecting the Best Model for Complex Interplay Between HIV and Nutrition

Vivek Sreejithkumar
Florida Atlantic University
Kia Ghods
Florida Atlantic University
Tharusha Bandara
University of Florida
Maia Martcheva
University of Florida
Necibe Tuncer (ntuncer@fau.edu)
Florida Atlantic University

Article

Keywords:

Posted Date: March 7th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2591977/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Selecting the Best Model for Complex Interplay Between HIV and Nutrition

Vivek Sreejithkumar[1], Kia Ghods[1], Tharusha Bandara[2], Maia Martcheva[2], and Necibe Tuncer[1,*]

[1]Florida Atlantic University, Department of Mathematical Sciences, Boca Raton, FL, 33431
[2]University of Florida, Department of Mathematics, Gainesville, FL, 32611
[*]corresponding author: ntuncer@fau.edu

ABSTRACT

HIV is a serious infectious disease. The role of nutrition in HIV progression has long been recognized. However, no mathematical models exist to describe the interplay of HIV with the nutritional components. In this paper, we present a family of models of HIV and the two protein components albumin and globulin. We use data from SIV-infected monkeys to perform model selection on the family of models. We discover that the simplest model accurately and uniquely describes the data. The selection of the simplest model leads to the observation that albumin and globulin do not impact the infection rate of target cells by the virus and the clearance of the infected target cells by the immune system. Furthermore, we obtain that the recruitment of target cells and immune cells do not depend on the globulin. Mathematical analysis of the selected model reveals that the model has an infection-free equilibrium and a unique infected equilibrium when the reproduction number is above one. The infection-free equilibrium is locally stable when the reproduction number is below one, and unstable when the reproduction number is greater than one. The infection equilibrium is locally stable whenever it exists. To determine the parameters of the best fitted model we perform structural and practical identifiability analysis. The structural identifiability analysis reveals that the model is identifiable when the immune cell infection rate is fixed at a value obtained from the literature. Practical identifiability reveals that only seven of the sixteen parameters are practically identifiable with the given data. Practical identifiability of parameters performed with synthetic data sampled a lot more frequently reveals that only two parameters are practically unidentifiable. We conclude that experiments that will improve the quality of the data can help improve the parameter estimates and lead to better understanding of the interplay of HIV and nutrition.

Introduction

Human immunodeficiency virus (HIV) is a serious public health concern in the USA and globally. Long term HIV infection progresses to the acquired immunodeficiency syndrome (AIDS). For individuals with HIV, maintaining proper nutrition is crucial in maintaining overall health and controlling the severity of the symptoms that are caused by HIV and AIDS[1]. A consistent healthy eating style sustains the immune system and helps support the strength of the immune responses when individuals are infected with HIV. Despite the fact that HIV and nutrition are strongly interlinked, no mathematical models seem to exist that model the feedbacks between HIV and nutrition. HIV primarily affects the immune system, and nutrition is critical for the immune system, which uses various nutritional elements to build immune system cells[2]. Thus, HIV infection and the medications taken to control it, increase nutritional need in HIV-infected individuals. However, not all HIV-infected individuals have proper access or interest in maintaining adequate eating habits, which can lead to malnutrition[3]. In contrast to other infections in which carbohydrates are first broken down to meet the increased needs of the immune system, proteins are first broken down in HIV[3], resulting in HIV-specific wasting and weight loss. Malnutrition, in turn, lowers immunity and increases viral load[3], as well as increases susceptibility to opportunistic infections. In this study, we develop within-host models of HIV to understand the relationship between nutrition and the immune response to HIV. Previous research suggests that improvements in nutrition acted as a form of early intervention for HIV patients; however, these studies lack mathematical models that incorporate this relation[4,5].

Nutrition has been modeled before in the absence of HIV quite a bit both through mathematical models and through statistical methods[6–9]. Mathematical models in nutrition has been the topic of a conference whose selected papers are published in[10]. A within-host model of nutrition and challenges with its calibration to data has been discussed by Juillet *et al.*[11]. A model based on chemical kinetics for the rate of utilization and storage of carbohydrates, fats and proteins is introduced and discussed in[12]. A model of various determinants of feeding which would facilitate the simulation of food intake, thus providing information for existing models of nutrient absorption and metabolism, has been developed in[13]. Mathematical models of nutrition on population level have been investigated. Shah *et al.*[14] apply optimal control theory to a model of poor nutrition on

population level and compute the reproduction number. Baleanu *et al.*[15] investigate a fractional order differential equation model of poor nutrition again on population level. Nutrition in ruminant has been studied[16] in two books which focus on nutritional system of ruminants through modeling. Within-host modeling of nutrients have been considered in an edited book by Coburn and Townsend[17].

Our models here are built on the classical HIV models that incorporate target cells, infected target cells and virus[18,19]. We further incorporate the immune response in the form of CD8 cells. The immune response has also previously been considered in HIV models[18,20]. The novelty of our models is in the incorporation of nutritional elements, and more specifically protein which is composed of albumin and globulins[4,21]. We incorporate albumin and globulin separately as albumin and globulins have different response to HIV infection. In particular, albumin decreases in HIV-infected individuals while the globulins increase[5]. This basic interdependence between HIV and the albumins and globulins is known, however, it is less known how specifically albumins and globulins interact with the immune system. Thus, our main question in this article is "What kind of model will capture best the data of target cells, virus, albumin and globulin?" Because we model these interactions for the first time we would like to develop sound, well validated model that represents the available data[22]. Relatively little is known how proteins impact the infection of new target cells, how they impact the clearance of the immune response of the infected target cells or how the globulins affect the recruitment of CD4 cells and CD8 cells. To find that dependence, we hypothesize a general form of these terms. We obtain a general model that has 17 specific submodels. We perform model selection on these 17 models and use AIC[23] to select the model that best describes the data. We analyze the best model. We further fit the best model to the data and consider the well-posedness of the parameter estimation problem. In other words, we study the structural identifiability of the best fitted model. To evaluate its fit to the data and the quality of the parameter estimation from the real data, we perform Monte Carlo Simulations (MCS) and compute the Average Relative Estimation Eroors (AREs) in the estimated parameters.

Methods

Mathematical Models of HIV and Nutrition

Nutrition, in general, is made up of macronutrients and micronutrients, both of which are required for proper functioning of the immune system. The macronutrients are important, but in HIV the proteins are the key. The total protein in the body, often measured in grams per deciliter (g/dL)[24] consists of albumin and globulin. In healthy individuals the ratio of albumin to globulin is around 1.7[21]. Total protein levels in HIV-infected people are higher, as are globulin levels, but albumin levels are lower[5,25]. Globulin is required for the production of elements of the immune system, most notably antibodies, whereas albumin has been found to be a good predictor of HIV progression, morbidity, and mortality[26,27]. We begin modeling the interplay between HIV and nutrition by incorporating protein to the well studied target-cell limited model[19]. While modeling nutrition coupled with a disease like HIV is new, our within-host nutrition-HIV model is based on an extensive literature of within HIV model[19] and within-host HIV models with immune response[20,28]. Our advancement here consists of adding albumin and globulin to the well studied HIV within-host model[19].

We used six differential equations to describe the temporal dynamics of six populations: target cells (T), virus-producing infected target cells (T_i), viral load (V), immune cells (Z), albumin (A) and globulin (G). Figure 1 presents a flow diagram of the model. The within-host model of HIV and nutrition is then built with the following system of ordinary differential equations:

$$\textbf{Within-Host Model of HIV and Nutrition:} \quad \begin{cases} \frac{dT}{dt} &= h_1(G) - \rho(A,G)TV - dT, \\ \frac{dT_i}{dt} &= \rho(A,G)TV - \delta T_i - \psi(A,G)T_iZ, \\ \frac{dV}{dt} &= \pi T_i - cV - \mu_v GV, \\ \frac{dZ}{dt} &= h_2(G) + bT_iZ - \mu_z Z, \\ \frac{dA}{dt} &= \lambda_A - \gamma_A AV - \mu_A A, \\ \frac{dG}{dt} &= \lambda_G + \gamma_G GV - \mu_G G, \end{cases} \tag{1}$$

Within-Host HIV and Nutrition Model:

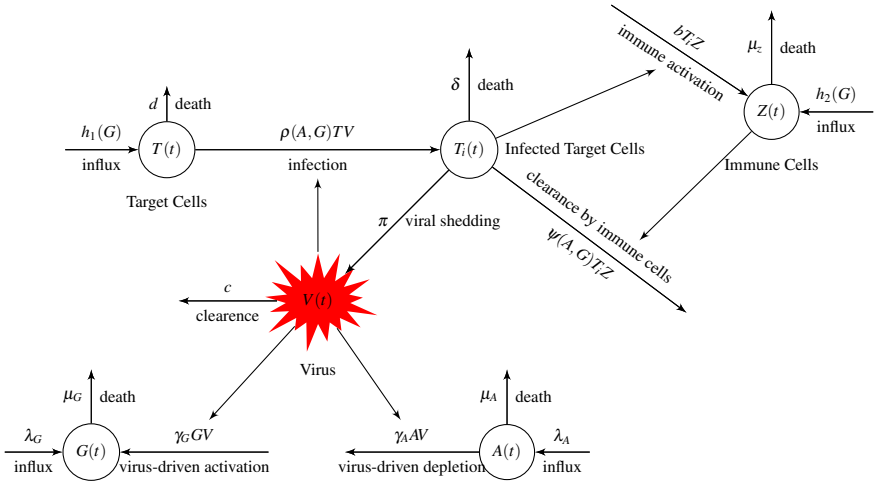


Figure 1. Flow diagram describing the interaction between target cells (T), infected target cells, (T_i), virus (V), immune cells (Z), albumin (A) and globulin (G).

HIV primarily targets CD4+ cells in the immune system. Since nutrition impacts the immune cell proliferation, we represent the production of target cells and immune cells as a function of nutrition. Therefore in the within-host model (1), we set the target and immune cell production depending on the individual's globulin level. Namely,

$$h_1(G) = r + r_0 G \quad \text{and} \quad h_2(G) = \lambda_z + \lambda_{z_0} G,$$

where, r_0 is the production of target cells per unit globulin, and r is the production rate of target cells. Similarly, λ_{z_0} is the recruitment rate of the immune cells per unit globulin, and λ_z is the recruitment rate of the immune cells. Virus particles, V, infect target cells, T, to produce infected cells, as expressed with the term $\rho(A,G)TV$. By setting the infection rate, $\rho(A,G)$, as a function of albumin and globulin, we account for the influence of nutrition on the target cell susceptibility[29]. Target cells die at rate d. Infected target die at rate δ and their clearance by the immune system is denoted by the term $\psi(A,G)T_iZ$. The ability of immune cells to attack infected target cells, $\psi(A,G)$, is considered to be a function of albumin and globulin[29]. Infected cells generate additional virus particles at rate π, while the virus particles are cleared from the system at rate c. Furthermore, immunoglobulins remove the viral particles at rate μ_v. Immune cells are activated at rate b in response to antigen and cleared at rate μ_z. Albumin and globulin intake are denoted by λ_A and λ_G respectively. The parameter μ_A represents the albumin consumption and clearance, whereas μ_G represents the globulin consumption and clearance. γ_A is the virus-driven depletion of albumin, while γ_G is the virus-driven increase in globulin. The functional forms of the infection rate of target cells $\rho(A,G)$ and the attack rate of immune cells $\psi(A,G)$ are defined as;

$$\rho(A,G) = \frac{\rho_0}{1 + A_1 A + A_2 G} \qquad \text{and} \qquad \psi(A,G) = \frac{\psi_{01} A + \psi_{02} G}{\Psi + A + G},$$

where ρ_0, A_i and ψ_{0i}, $i = 1,2$, are proportionality constants, and Ψ is a half-saturation constant. The constants $\Psi, A_1, A_2, \psi_{01}$ and ψ_{02} allow us to remove the dependence of the transmission rate ρ and killing rate ψ on the nutrition. We performed model selection on the global model (1) to determine more robust forms for the terms $h_1(G)$, $h_2(G)$, $\rho(A,G)$ and $\psi(A,G)$.

Model Selection Process

We proposed a family of models that will aid in clarifying the complex relationship between HIV and protein dynamics. However, it is not clear which model would be the best for this purpose. It is standard to judge the suitability of a model by comparing its ability to match the experimental data, but this approach is known to favor model complexity. On the other hand, Akaike information criteria (AIC) balances model complexity and the goodness-of-fit. We use AIC to determine the robust

Model	A_1	A_2	ψ_{01}	ψ_{02}	λ_{z_0}	r_0	$\rho(A,G)$	$\psi(A,G)$	$h_1(G)$	$h_2(G)$
1	0	0	ψ_{01}	ψ_{01} and $\Psi=0$	0	0	ρ_0	ψ_{01}	r	λ_z
2	0	0	ψ_{01}	0	0	0	ρ_0	$\frac{\psi_{01}A}{\Psi+A+G}$	r	λ_z
3	0	0	0	ψ_{02}	0	0	ρ_0	$\frac{\psi_{02}G}{\Psi+A+G}$	r	λ_z
4	0	0	ψ_{01}	ψ_{02}	0	0	ρ_0	$\frac{\psi_{01}A+\psi_{02}G}{\Psi+A+G}$	r	λ_z
5	A_1	0	ψ_{01}	ψ_{01} and $\Psi=0$	0	0	$\frac{\rho_0}{1+A_1A}$	ψ_{01}	r	λ_z
6	A_1	0	ψ_{01}	0	0	0	$\frac{\rho_0}{1+A_1A}$	$\frac{\psi_{01}A}{\Psi+A+G}$	r	λ_z
7	A_1	0	0	ψ_{02}	0	0	$\frac{\rho_0}{1+A_1A}$	$\frac{\psi_{02}G}{\Psi+A+G}$	r	λ_z
8	A_1	0	ψ_{01}	ψ_{02}	0	0	$\frac{\rho_0}{1+A_1A}$	$\frac{\psi_{01}A+\psi_{02}G}{\Psi+A+G}$	r	λ_z
9	0	A_2	ψ_{01}	0	0	0	$\frac{\rho_0}{1+A_2G}$	ψ_{01}	r	λ_z
10	0	A_2	ψ_{01}	0	0	0	$\frac{\rho_0}{1+A_2G}$	$\frac{\psi_{01}A}{\Psi+A+G}$	r	λ_z
11	0	A_2	0	ψ_{02}	0	0	$\frac{\rho_0}{1+A_2G}$	$\frac{\psi_{02}G}{\Psi+A+G}$	r	λ_z
12	0	A_2	ψ_{01}	ψ_{02}	0	0	$\frac{\rho_0}{1+A_2G}$	$\frac{\psi_{01}A+\psi_{02}G}{\Psi+A+G}$	r	λ_z
13	A_1	A_2	ψ_{01}	ψ_{01}	0	0	$\frac{\rho_0}{1+A_1A+A_2G}$	ψ_{01}	r	λ_z
14	A_1	A_2	ψ_{01}	0	0	0	$\frac{\rho_0}{1+A_1A+A_2G}$	$\frac{\psi_{01}A}{\Psi+A+G}$	r	λ_z
15	A_1	A_2	0	ψ_{02}	0	0	$\frac{\rho_0}{1+A_1A+A_2G}$	$\frac{\psi_{02}G}{\Psi+A+G}$	r	λ_z
16	A_1	A_2	ψ_{01}	ψ_{02}	0	0	$\frac{\rho_0}{1+A_1A+A_2G}$	$\frac{\psi_{01}A+\psi_{02}G}{\Psi+A+G}$	r	λ_z
17	A_1	A_2	ψ_{01}	ψ_{02}	r_0	λ_{z_0}	$\frac{\rho_0}{1+A_1A+A_2G}$	$\frac{\psi_{01}A+\psi_{02}G}{\Psi+A+G}$	$r+r_0G$	$\lambda_z+\lambda_{z_0}G$

Table 1. Formulation of various model forms

Time of Measurement	Viral Load vRNA copies per mL log scale	CD4 Cells CD4 cells per mL log scale	Albumin mg/dL	Globulin mg/dL
Week 0	-	6.1153	3.4430	2.5718
Week 1	6.9299	5.7920	2.9426	-
Week 2	7.0531	5.5696	3.1272	-
Week 3	5.9624	5.1809	2.9204	-
Week 4	5.5821	5.3726	-	-
Week 5	-	-	3.8768	-
Week 6	5.4362	5.3823	3.2235	-
Week 8	5.1959	5.4314	3.5133	-
Week 10	5.1903	5.5788	-	-
Week 12	4.9782	5.3912	3.3405	2.5920
Week 16	-	-	3.0636	-
Week 24	-	-	-	2.7064

Table 2. Average Data Measurements - Data for the Average Viral Load (viral RNA copies per mL) and CD4+ concentration (CD4 cells per mL) came from a study on simian immunodeficiency virus (SIV) in 6 rhesus monkeys over 12 weeks[30]. Albumin given in mg/dL obtained from an SIV study on 8 juvenile pig-tailed macaques over 16 weeks[31]. Globulin, given in mg/dL, was measured through peripheral blood examination on 4 rhesus monkeys with SIV[32].

forms of the terms depending on albumin and globulin; $\lambda(G)$, $\lambda_z(G)$, $\rho(A,G)$ and $\psi(A,G)$. The set of candidate models is obtained by setting $r_0 = 0$ and/or $\lambda_{z0} = 0$ and/or $A_1 = 0$ and/or $A_2 = 0$ and/or $\psi_{01} = 0$ and/or $\psi_{02} = 0$. We also explored the scenario when $\psi_{01} = \psi_{02} = \psi_0$ and $\Psi = 0$. In total, we compared 17 models (see Table 1).

To validate the models, we used data from a rhesus monkey study of simian immunodeficiency virus (SIV), in which CD4+ T-cell and plasma viral load were measured for 12 weeks following SIV infection[30]. We took albumin levels from another SIV experiment in which albumin concentrations were measured for 16 weeks after SIV infection in 8 pig-tailed macaques[31]. For globulin levels, we obtained data from the published study[32] in which globulin levels were assessed in the peripheral blood of rhesus monkeys infected with SIV (see Table 2).

The observations are the state variables of the model, namely CD4+ cells is the state variable $T(t,p)$, virus is $V(t,p)$, albumin is $A(t,p)$, and globulin is $G(t,p)$ in (1), with p denoting the parameters of the model. Observations vary as the model parameters p vary. Clearly, data collected in the experimental setting do not fall on the smooth path of the observations $T(t,p)$, $V(t,p)$, $A(t,p)$ and $G(t,p)$. We represent the relationship between the experimental data and model observations with the following statistical model,

$$\begin{aligned} Y_1^i &= T(t_i,p) + E_i \quad i = 1,\cdots,n_1 \qquad Y_2^j = V(t_j,p) + E_j \quad j = 1,\cdots,n_2 \\ Y_3^k &= A(t_k,p) + E_k \quad k = 1\cdots,n_3 \qquad Y_4^l = G(t_l,p) + E_l \quad l = 1,\cdots,n_4 \end{aligned} \tag{2}$$

where E_i, E_j, E_k, E_l are the measurement errors with Gaussian structure. The measurement errors are independent and normally distributed with mean zero, and standard deviation σ. Parameter estimation based on a maximum likelihood principle and least square principle are equivalent when the measurement errors are independent and normally distributed. The least square estimate can be obtained by minimizing the following objective function, which will be refered as sum of squared errors (SSE),

$$\hat{p} = \min_p \omega_1 \sum_{i=1}^{n_1} (Y_1^i - \log_{10} T(t_i,p))^2 + \omega_2 \sum_{i=1}^{n_2} (Y_2^i - \log_{10} V(t_i,p))^2 + \omega_3 \sum_{i=1}^{n_3} (Y_3^i - A(t_i,p))^2 + \omega_4 \sum_{i=1}^{n_4} (Y_4^i - G(t_i,p))^2 \tag{3}$$

SSE values alone may indicate the better model when models have the same number of parameters. Since all 17 models considered have different set of parameters, in this study we employed the AIC to assess which model explains the data best. The corrected AIC for the finite sample size is given by[23],

$$\mathrm{AIC}_c = n \ln\left(\frac{\mathrm{SSE}}{n}\right) + 2K + \frac{2K(K+1)}{n-K-1} \tag{4}$$

Model	SSE	Number of Parameters	AIC_c Score	Δ	Relative likelihood
1	3.32733	17	41.6114	-	0.996357534
2	3.32703	18	59.6532	18.0418	0.000120417
3	3.32728	18	59.6554	18.044	0.000120285
4	3.32741	19	82.2121	40.6007	1.52085×10^{-9}
5	3.35822	18	59.9238	18.3124	0.000105179
6	3.34985	19	82.4069	40.7955	1.3797×10^{-9}
7	3.33751	19	82.3	40.7955	1.45545×10^{-9}
8	3.32727	20	111.211	69.5996	7.67457×10^{-16}
9	2.64807	18	53.0339	11.4225	0.003296483
10	2.72119	19	76.3795	34.7681	2.80943×10^{-8}
11	2.67312	19	75.8625	34.2511	3.63817×10^{-8}
12	2.24495	20	99.8002	58.1888	2.3061×10^{-13}
13	2.69078	19	76.0535	34.4421	3.3068×10^{-8}
14	2.57026	20	103.725	62.1136	3.24055×10^{-14}
15	2.69456	20	105.094	63.4826	1.63435×10^{-14}
16	2.69917	21	143.81	102.1986	6.40134×10^{-23}
17	3.23414	23	284.388	242.7766	1.9061×10^{-53}

Table 3. Model selection from the collection of models.

where n is the number of data points and K is the number of parameters fit by the model plus one[23]. AIC balances the model complexity and the SSE by penalizing the models with higher parameters. Due to this penalty, a more complex model needs to improve its fit to compensate for its higher number of parameters. For instance model 12 has a higher AIC_c despite having the minimum SSE value (see Table 3). Since the AIC score of each model varies greatly; it is standard to rescale them with the minimum AIC score by calculating the difference between the AIC_c score and $AIC_{c_{min}}$. The rescaled AIC_c scores of each model, $\Delta = AIC_c - AIC_{c_{min}}$, are reported in Table 3. The rescaled AIC values are a direct way to compare the models, only the models with $\Delta < 2$ have strong support[23]. Based on our analysis, model 1 is the only model that has strong support for the experimental data. The last column of Table 3 represents the relative likelihood of the model j, given by

$$w_j = \frac{e^{\Delta j/2}}{\sum_{i=1}^{17} e^{\Delta i/2}}. \tag{5}$$

The models that best describe the data are those whose relative likelihood sum up to $\sum w_j > 0.95$. The model 1 exclusively explains the data since its relative probability is higher than the threshold of 0.95.

Analysis of the Best Selected Model

The best model takes the form:

$$\textbf{Selected Model via AIC Criteria:} \quad \begin{cases} \frac{dT}{dt} &= r - \rho_0 TV - dT, \\ \frac{dT_i}{dt} &= \rho_0 TV - \delta T_i - \psi_{01} T_i Z, \\ \frac{dV}{dt} &= \pi T_i - cV - \mu_v GV, \\ \frac{dZ}{dt} &= \lambda_z + bT_i Z - \mu_z Z, \\ \frac{dA}{dt} &= \lambda_A - \gamma_A AV - \mu_A A, \\ \frac{dG}{dt} &= \lambda_G + \gamma_G GV - \mu_G G, \end{cases} \tag{6}$$

The Infection-free Equilibrium and Immunological Reproduction Number

Using the Van den Driessche and Watmough Approach, first we derive the immunological reproduction number of the within-host model of HIV and nutrition given by system (6). We arrange the equations in (6) so that the first 2 components of the ODE system correspond to the infected compartments, and the last 4 components correspond to the non-infected compartments. We re-write the system as,

$$\begin{aligned} x_m{}' &= f_m(x,y), \qquad m = 1,2 \\ y_n{}' &= g_n(x,y), \qquad n = 1,2,3,4, \end{aligned}$$

where $x_m = (T_i \;\; V)^T$ and $y_n = (T \;\; Z \;\; A \;\; G)^T$. Now define $F_1 = \rho_0 TV$, $F_2 = \pi T_i$, $V_1 = \delta T_i + \psi_{01} T_i Z$, and $V_2 = cV + \mu_V GV$ so that we have

$$\begin{aligned} x_m{}' &= F_m(x,y) - V_m(x,y), \qquad m = 1,2 \\ y_n{}' &= g_n(x,y), \qquad\qquad\quad n = 1,2,3,4 \end{aligned}$$

Now, the infection-free system $y' = g(0,y)$ has the unique equilibrium $\varepsilon_0 = (0,y_0)$, where $y_0 = (T_0, Z_0, A_0, G_0) = \left(\frac{r}{d}, \;\; \frac{\lambda_z}{\mu_z}, \;\; \frac{\lambda_A}{\mu_A}, \;\; \frac{\lambda_G}{\mu_G}\right)$. Observe that,

$$F := \left[\frac{\partial F_i(0,y_0)}{\partial x_j}\right] = \begin{bmatrix} 0 & \rho_0 T_0 \\ \pi & 0 \end{bmatrix} \quad \text{and} \quad V := \left[\frac{\partial V_i(0,y_0)}{\partial x_j}\right] = \begin{bmatrix} \delta + \psi_{01} Z_0 & 0 \\ 0 & c + \mu_V G_0 \end{bmatrix}$$

Hence the next-generation matrix is defined by

$$K := FV^{-1} = \begin{bmatrix} 0 & \frac{\rho_0 T_0}{c+\mu_V G_0} \\ \frac{\pi}{\delta + \psi_{01} Z_0} & 0 \end{bmatrix}.$$

This gives us the immunological reproduction number

$$\mathscr{R}_0 = \frac{\pi \rho_0 T_0}{(c+\mu_V G_0)(\delta + \psi_{01} Z_0)} = \frac{pr\rho \mu_G \mu_Z}{d(c\mu_G + \lambda_G \mu_V)(\delta \mu_Z + \psi_{01} \lambda_Z)} \tag{7}$$

Stability of the Infection-free Equilibrium

The infection-free equilibrium of the system (8) is given by $E_0 = (T_0, 0, 0, Z_0, A_0, G_0)$, where $T_0 = \frac{r}{d}$, $Z_0 = \frac{\lambda_Z}{\mu_Z}$, $A_0 = \frac{\lambda_A}{\mu_A}$ and $G_0 = \frac{\lambda_G}{\mu_G}$. In this subsection, we prove the following Theorem.

Theorem 1 *If $\mathscr{R}_0 < 1$, then the infection-free equilibrium is locally asymptotically stable. If $R_0 > 1$, it is unstable.*

Proof: The local stability of equilibria is determined by the eigenvalues of the Jacobian computed at this equilibrium. The Jacobian of the model (8) at the equilibrium $E_0 = (T_0, 0, 0, Z_0, A_0, G_0)$ is

$$J(E_0) = \begin{bmatrix} -d & 0 & -\frac{\rho_0 r}{d} & 0 & 0 & 0 \\ 0 & -\delta - \frac{\psi_{01}\lambda_Z}{\mu_Z} & \frac{\rho_0 r}{d} & 0 & 0 & 0 \\ 0 & \pi & -c - \frac{\mu_V \lambda_G}{\mu_G} & 0 & 0 & 0 \\ 0 & \frac{b\lambda_Z}{\mu_Z} & 0 & -\mu_Z & 0 & 0 \\ 0 & 0 & -\frac{\gamma_A \lambda_A}{\mu_A} & 0 & -\mu_A & 0 \\ 0 & 0 & \frac{\gamma_G \lambda_G}{\mu_G} & 0 & 0 & -\mu_G \end{bmatrix}$$

Hence the characteristic equation of the Jacobian is given by

$$|J(E_0) - kI| = \begin{vmatrix} -d-k & 0 & -\frac{\rho_0 r}{d} & 0 & 0 & 0 \\ 0 & -\delta - \frac{\psi_{01}\lambda_Z}{\mu_Z} - k & \frac{\rho_0 r}{d} & 0 & 0 & 0 \\ 0 & \pi & -c - \frac{\mu_V \lambda_G}{\mu_G} - k & 0 & 0 & 0 \\ 0 & \frac{b\lambda_Z}{\mu_Z} & 0 & -\mu_Z - k & 0 & 0 \\ 0 & 0 & -\frac{\gamma_A \lambda_A}{\mu_A} & 0 & -\mu_A - k & 0 \\ 0 & 0 & \frac{\gamma_G \lambda_G}{\mu_G} & 0 & 0 & -\mu_G - k \end{vmatrix} = 0$$

Clearly, $-d, -\mu_Z, -\mu_A$ and $-\mu_G$ are eigenvalues of the matrix $J(E_0)$ and the remaining eigenvalues are given by the eigenvalues of the following matrix J_1,

$$J_1 = \begin{vmatrix} -\delta - \frac{\psi_{01}\lambda_Z}{\mu_Z} & \frac{\rho_0 r}{d} \\ \pi & -c - \frac{\mu_V \lambda_G}{\mu_G} \end{vmatrix}$$

This leads to the following quadratic characteristic equation:

$$P(k) := k^2 + \left[\left(c + \frac{\mu_V \lambda_G}{\mu_G}\right) + \left(\delta + \frac{\psi_{01}\lambda_Z}{\mu_Z}\right)\right] k + (\delta + \frac{\psi_{01}\lambda_Z}{\mu_Z})(c + \frac{\mu_v \lambda_G}{\mu_G}) - \frac{\pi\rho_0 r}{d} = 0.$$

Now, it can be seen that this equation has one positive root if $\mathscr{R}_0 > 1$, and two negative real roots or two complex roots with negative real parts if $\mathscr{R}_0 < 1$.

■

0.1 Existence and Stability of Infection Equilibrium

We set the left hand side of the model (6) equals to 0 and obtain the following:

$$\begin{aligned} 0 &= r - \rho_0 T^* V^* - dT^*, \\ 0 &= \rho_0 T^* V^* - \delta T_i^* - \psi_{01} T_i^* Z^*, \\ 0 &= \pi T_i^* - cV^* - \mu_V G^* V^*, \\ 0 &= \lambda_z + bT_i^* Z^* - \mu_z Z^*, \\ 0 &= \lambda_A - \gamma_A A^* V^* - \mu_A A^*, \\ 0 &= \lambda_G + \gamma_G G^* V^* - \mu_G G^*. \end{aligned} \tag{8}$$

By the sixth equation of (8) we get that

$$G^* = \frac{\lambda_G}{\mu_G - \gamma_G V^*}. \tag{9}$$

Since we need system variables to stay positive we set $G^* > 0$ and obtain $0 < V^* < \frac{\mu_G}{\gamma_G}$.
Similarly, the fifth equation of (8) leads to

$$A^* = \frac{\lambda_A}{\gamma_A V^* + \mu_A}, \tag{10}$$

and the first equation of (8) gives

$$T^* = \frac{r}{d + \rho_0 V^*}. \tag{11}$$

Both A^* and T^* are trivially positive. Now, solving the third equation of (8) we get

$$T_i^* = \frac{V^*}{\pi}\left(c + \frac{\lambda_G \mu_V}{\mu_G - \gamma_G V^*}\right), \tag{12}$$

which is also positive since $0 < V^* < \frac{\mu_G}{\gamma_G}$. Next, we solve the fourth equation of (8) for Z^*. Substituting (12) we obtain,

$$Z^* = \frac{\lambda_Z}{\mu_Z - \frac{bV^*}{\pi}\left(c + \frac{\lambda_G \mu_V}{\mu_G - \gamma_G V^*}\right)}. \tag{13}$$

In order to obtain a positive solution for Z^*, we must satisfy

$$b\lambda_G \mu_V V^* < (\pi\mu_Z - bcV^*)(\mu_G - \gamma_G V^*). \tag{14}$$

The left hand side of the equation (14) is a straight line that goes through the origin while the right hand side is a parabola with roots $V_1{}^* = \frac{\mu_G}{\gamma_G}$ and $V_2{}^* = \frac{\pi\mu_Z}{bc}$. The parabola opens up as the coefficient of V^* is positive and meets the straight line $y = b\lambda_G \mu_V V^*$ at two distinct points, say V_1 and V_2, with $V_1 < V_2$. Then $V_1 < V_1{}^*$ and $V_1 < V_2{}^*$, for any $V_1{}^*, V_2{}^*$. Thus, $V_1 < \frac{\mu_G}{\gamma_G}$. As this satisfies our requirement to have positive values at infection equilibrium, $Z^* > 0$. Finally, substituting T^*, T_i^* and Z^* in the second equation of (8), $\delta T_i^* + \psi_{01} T_i^* Z^* = \rho_0 T^* V^*$, we get

$$\frac{\delta}{\pi}\left(c + \frac{\lambda_G \mu_V}{\mu_G - \gamma_G V^*}\right) + \frac{\psi_{01}}{\pi}\left(c + \frac{\lambda_G \mu_V}{\mu_G - \gamma_G V^*}\right)\left(\frac{\lambda_Z}{\mu_Z - \frac{bV^*}{\pi}\left(c + \frac{\lambda_G \mu_V}{\mu_G - \gamma_G V^*}\right)}\right) = \frac{\rho_0 r}{d + \rho_0 V^*}. \tag{15}$$

Let $y_1(V^*) = \frac{\delta}{\pi}\left(c + \frac{\lambda_G\mu_V}{\mu_G-\gamma_G V^*}\right) + \frac{\psi_{01}}{\pi}\left(c + \frac{\lambda_G\mu_V}{\mu_G-\gamma_G V^*}\right)\left(\frac{\lambda_Z}{\mu_Z - \frac{bV^*}{\pi}\left(c+\frac{\lambda_G\mu_V}{\mu_G-\gamma_G V^*}\right)}\right)$ and $y_2(V^*) = \frac{\rho_0 r}{d+\rho_0 V^*}$. Here, y_1 is increasing on $\left(0, \frac{\mu_G}{\gamma_G}\right)$, while y_2 is decreasing on $\left(0, \frac{\mu_G}{\gamma_G}\right)$ (see the following Remark).

Thus, if a solution exists, then it must be unique. We prove the existence of a solution as follows. Since $0 < V^* < \frac{\mu_G}{\gamma_G}$, we set $V^* = 0$ into $y_1(V^*)$ and $y_2(V^*)$ and obtain

$$y_1(0) = \frac{(c\mu_G + \lambda_G\mu_V)(\delta\mu_Z + \psi_{01}\lambda_Z)}{p\mu_G\mu_Z} \qquad \text{and} \qquad y_2(0) = \frac{\rho_0 r}{d}.$$

Now,

$$\mathscr{R}_0 = \frac{\pi r\rho_0\mu_G\mu_Z}{d(c\mu_G+\lambda_G\mu_V)(\delta\mu_Z+\Psi_{01}\lambda_Z)} > 1 \Longrightarrow \frac{(c\mu_G+\lambda_G\mu_V)(\delta\mu_Z+\psi_{01}\lambda_Z)}{\pi\mu_G\mu_Z} < \frac{\rho_0 r}{d}.$$

That is, $y_1(0) < y_2(0)$. Similarly, we set $V^* = \frac{\mu_G}{\gamma_G}$ to obtain

$$y_2\left(\frac{\mu_G}{\gamma_G}\right) = \frac{r\rho_0\gamma_G}{d\gamma_G+\rho_0\mu_G} \quad \text{and} \quad \lim_{V^* \longrightarrow \frac{\mu_G}{\gamma_G}} y_1(V^*) = \infty.$$

Thus, $y_2\left(\frac{\mu_G}{\gamma_G}\right) < y_1\left(\frac{\mu_G}{\gamma_G}\right)$. Since y_2 is decreasing on $\left(0, \frac{\mu_G}{\gamma_G}\right)$, y_1 is increasing on $\left(0, \frac{\mu_G}{\gamma_G}\right)$, $y_1(0) < y_2(0)$ and $y_2\left(\frac{\mu_G}{\gamma_G}\right) < y_1\left(\frac{\mu_G}{\gamma_G}\right)$, $y_1(V^*)$ and $y_2(V^*)$ meet at one point, V^*. When $\mathscr{R}_0 > 1$, the equation (15) has a unique root on $\left(0, \frac{\mu_G}{\gamma_G}\right)$. Hence, the equation (15) has a unique positive solution V^* and consequently the system (8) has a unique infection equilibrium $E_1 = (T^*, T_i^*, V^*, Z^*, A^*, G^*)$. Now, we have proved the following theorem.

Theorem 2 *There exists a unique infection equilibrium $E_1 = (T^*, T_i^*, V^*, Z^*, A^*, G^*)$ of the system (8) when $\mathscr{R}_0 > 1$.*

Remark: Here, we prove that the function

$$y_1(V^*) = \frac{\delta}{p}\left(c + \frac{\lambda_G\mu_V}{\mu_G-\gamma_G V^*}\right) + \frac{\psi_{01}}{p}\left(c + \frac{\lambda_G\mu_V}{\mu_G-\gamma_G V^*}\right)\left(\frac{\lambda_Z}{\mu_Z - \frac{bV^*}{p}\left(c+\frac{\lambda_G\mu_V}{\mu_G-\gamma_G V^*}\right)}\right)$$

is increasing on $\left(0, \frac{\mu_G}{\gamma_G}\right)$, while the function

$$y_2(V^*) = \frac{\rho_0 r}{d+\rho_0 V^*}$$

is decreasing on $\left(0, \frac{\mu_G}{\gamma_G}\right)$, which was an integral part in proving Theorem (2). Observe that,

$${y_2}'(V^*) = -\frac{r{\rho_0}^2}{(d+\rho_0 V^*)^2} < 0 \quad \text{on} \quad \left(0, \frac{\mu_G}{\gamma_G}\right) \Longrightarrow y_2 \quad \text{is decreasing on} \quad \left(0, \frac{\mu_G}{\gamma_G}\right).$$

Now, write $y_1(V^*) = f(V^*) + g(V^*)h(V^*)$, where $f(V^*) = \frac{\delta}{p}\left(c + \frac{\lambda_G\mu_V}{\mu_G-\gamma_G V^*}\right)$, $g(V^*) = \frac{\psi_{01}}{p}\left(c + \frac{\lambda_G\mu_V}{\mu_G-\gamma_G V^*}\right)$ and $h(V^*) = \frac{\lambda_Z}{\mu_Z - \frac{bV^*}{p}\left(c+\frac{\lambda_G\mu_V}{\mu_G-\gamma_G V^*}\right)}$. Then,

$$\begin{aligned} f'(V^*) &= \frac{\delta}{p}\left[\frac{\lambda_G\mu_V\gamma_G}{(\mu_G-\gamma_G V^*)^2}\right], \\ g'(V^*) &= \frac{\psi_{01}}{p}\left[\frac{\lambda_G\mu_V\gamma_G}{(\mu_G-\gamma_G V^*)^2}\right], \\ \text{and} \quad h'(V^*) &= \frac{bp\lambda_Z\left[c(\mu_G-\gamma_G V^*)^2 + \lambda_G\mu_G\mu_V\right]}{\left[p\mu_Z(\mu_G-\gamma_G V^*) - bV^*(c\mu_G - c\gamma_G V^* + \lambda_G\mu_V)\right]^2}. \end{aligned}$$

It is easy to see that all three derivatives f', g' and h' are positive on $\left(0, \frac{\mu_G}{\gamma_G}\right)$. Furthermore, the functions g and h themselves are positive on $\left(0, \frac{\mu_G}{\gamma_G}\right)$. Therefore,

$$y_1(V^*) = f'(V^*) + g(V^*)h'(V^*) + h(V^*)g'(V^*) > 0 \quad \text{on} \quad \left(0, \frac{\mu_G}{\gamma_G}\right),$$

implying that y_1 is increasing.

Theorem 3 *If $\mathscr{R}_0 > 1$ the unique endemic equilibrium is locally asymptotically stable.*

Proof 1 *To determine the local stability of the unique infection equilibrium we compute the Jacobian of the system. We arrange the variables as (T, T_i, V, Z, A, G) and with this ordering the Jacobian takes the form*

$$J = \begin{pmatrix} -\rho_0 V^* - d & 0 & -\rho_0 T^* & 0 & 0 & 0 \\ \rho_0 V^* & -\delta - \psi_{01} Z^* & \rho_0 T^* & -\psi_{01} T_i & 0 & 0 \\ 0 & \pi & -c - \mu_v G^* & 0 & 0 & -\mu_V V^* \\ 0 & bZ^* & 0 & bT_i^* - \mu_z & 0 & 0 \\ 0 & 0 & -\gamma_A A^* & 0 & -\gamma_A V^* - \mu_A & 0 \\ 0 & 0 & \gamma_G G^* & 0 & 0 & \gamma_G V^* - \mu_G \end{pmatrix}$$

This Jacobian clearly has one negative eigenvalue $\lambda_1 = -\gamma_A V^* - \mu_A$. The remaining eigenvalues are the solution of the following matrix equation

$$0 = \begin{vmatrix} -\rho_0 V^* - d - \lambda & 0 & -\rho_0 T^* & 0 & 0 \\ \rho_0 V^* & -\delta - \psi_{01} Z^* - \lambda & \rho_0 T^* & -\psi_{01} T_i & 0 \\ 0 & \pi & -c - \mu_V G^* - \lambda & 0 & -\mu_V V^* \\ 0 & bZ^* & 0 & bT_i^* - \mu_z - \lambda & 0 \\ 0 & 0 & \gamma_G G^* & 0 & \gamma_G V^* - \mu_G - \lambda \end{vmatrix}$$

We expand the determinant by the 3rd column. The resulting x determinants can easily be expanded, and we obtain the characteristic equation:

$$\begin{gathered} -\rho_0 T^* \rho_0 V p(\lambda + \mu_G - \gamma_G V^*)(\lambda + \mu_z - bT_i^*) \\ +\rho_0 T^* \pi(\lambda + d + \rho_0 V^*)(\lambda + \mu_G - \gamma_G V^*)(\lambda + \mu_z - bT_i^*) \\ -(\lambda + c + \mu_V G)(\lambda + \mu_G - \gamma_G V^*)(\lambda + d + \rho_0 V^*)[(\lambda + \mu_z - bT_i^*)(\lambda + \delta + \psi_{01} Z^*) + bZ^* \psi_{01} T_i^*] \\ -\lambda_G G^* \mu_V V^* (\lambda + d + \rho_0 V^*)[(\lambda + \mu_z - bT_i^*)(\lambda + \delta + \psi_{01} Z^*) + bZ^* \psi_{01} T_i^*] = 0 \end{gathered}$$

Combining the first two terms and the last two terms and moving the negative term to the right hand side of the equation, we obtain the following characteristic equation.

$$\begin{gathered} \rho_0 T^* \pi(\lambda + d)(\lambda + \mu_G - \gamma_G G^*)(\lambda + \mu_z - bT_i^*) \\ = (\lambda + d + \rho_0 V^*)[(\lambda + c + \mu_V G^*)(\lambda + \mu_G - \gamma_G V^*) + \gamma_G G^* \mu_V V^*][(\lambda + \mu_z - bT_i^*)(\lambda + \delta + \psi_{01} Z^*) + bZ^* \psi_{01} T_i^*] \end{gathered}$$

In what follows we shall prove that this equation does not have roots with non-negative real parts. To do that we rewrite the equation in the form $\mathscr{H}(\lambda) = \mathscr{G}(\lambda)$ where $\mathscr{H}(\lambda) = \rho_0 T^* p$ and

$$\mathscr{G}(\lambda) = \frac{\lambda + d + \rho_0 V^*}{\lambda + d} \left[\lambda + c + \mu_V G^* + \frac{\gamma_G G^* \mu_V V^*}{\lambda + \mu_G - \gamma_G V^*}\right] \left[\lambda + \delta + \psi_{01} Z^* + \frac{bZ^* \psi_{01} T_i^*}{\lambda + \mu_z - bT_i^*}\right]$$

Taking absolute value of $\mathscr{G}(\lambda)$, we notice that

$$\left|\frac{\lambda + d + \rho_0 V^*}{\lambda + d}\right| > 1$$

We consider more carefully the absolute value of one of the terms in [.]. The other terms can be handled in a similar way. We have

$$\left|\lambda + \delta + \psi_{01} Z^* + \frac{bZ^* \psi_{01} T_i^*}{\lambda + \mu_z - bT_i^*}\right| = \left|\lambda + \delta + \psi_{01} Z^* \frac{\lambda + \mu_z}{\lambda + \mu_z - bT_i^*}\right|$$

Let $\lambda = x + yi$ with $x \geq 0$. Then

$$\left|\lambda + \delta + \psi_{01} Z^* \frac{\lambda + \mu_z}{\lambda + \mu_z - bT_i^*}\right| = \left|x + yi + \delta + \psi_{01} Z^* \frac{x + yi + \mu_z}{x + yi + \mu_z - bT_i^*}\right|$$

$$= \left|x + yi + \delta + \psi_{01} Z^* \frac{(x + \mu_x)(x + \mu_z - bT_i^*) + y^2 - bT_i^* yi}{(x + \mu_z - bT_i^*)^2 + y^2}\right|$$

Parameter	Definition	Estimate
r	Recruitment rate of target cells	47513.3 cells/(mL$\times$ day)
d	Clearance rate of target cells	0.170469 1/day
ρ_0	Rate at which target cells are infected	5.94738×10^{-9} mL/(vRNA $\times$ day)
δ	Clearance rate of infected target cells	0.0315779 1/day
ψ_{01}	Rate at which immune cells attack infected cells	0.200186 mL/(cells $\times$ day)
π	Viral shedding rate of infected cells	6112.02 vRNA/(cells $\times$ day)
c	Clearance rate of viral cells	8.10898 1/day
μ_v	Viral clearance by the immunoglobulins	2.48331×10^{-9} dL/(mg $\times$ day)
b	Antigen driven proliferation rate of immune cells	4.04578×10^{-7} mL/(cells $\times$ day)
μ_z	Clearance rate of immune cells	0.161802 1/day
λ_A	Production rate of albumin	28.5878 mg/(dL $\times$ day)
γ_A	Serum albumin recruitment	2.02926×10^{-8} mL/(vRNA $\times$ day)
μ_A	Serum albumin use and clearance rate of albumin	8.74127 1/day
λ_G	Globulin recruitment rate	0.0436947mg/(dL $\times$ day)
γ_G	Virus driven increase in globulin	1.55765×10^{-12} mL/(vRNA $\times$ day)
μ_G	Clearance rate of globulin cells	0.0163711 1/day

Table 4. Definitions of the within-host HIV protein model (6) parameters and their estimated values. The immune cell production rate, λ_z, is fixed in order to get a structurally identifiable model. $\lambda_z = 1$ cells/(mL $\times$ day)[33]. The estimated values are obtained by minimizing the objective function (3).

$$\geq x+\delta+\psi_{01}Z^*\frac{(x+\mu_x)(x+\mu_z-bT_i^*)+y^2}{(x+\mu_z-bT_i^*)^2+y^2} \geq \delta+\psi_{01}Z^*$$

Similarly, if $\lambda = x+yi$ *with* $x \geq 0$*, we have*

$$\left|\lambda+c+\mu_V G^*+\frac{\gamma_G G^*\mu_V V^*}{\lambda+\mu_G-\gamma_G V^*}\right| \geq c+\mu_V G^*$$

Hence $|\mathscr{G}(\lambda)| > (\delta+\psi_{01}Z^*)(c+\mu_V G^*)$*. We notiqce that from the equilibrium equations for* T_i^* *and* V^* *we have that*

$$\rho_0 T^* V^* \pi T_i^* = (\delta+\psi_{01}Z^*)(c+\mu_V G^*)T_i^* V^*$$

Hence

$$\rho_0 T^* \pi = (\delta+\psi_{01}Z^*)(c+\mu_V G^*).$$

From the above discussion, we have that for $\lambda = x+yi$ *with* $x \geq 0$ $|\mathscr{G}(\lambda)| > |\mathscr{H}(\lambda)|$ *for all* λ *with non-negative real part. Hence the equation* $\mathscr{H}(\lambda) = \mathscr{G}(\lambda)$ *does not have roots with non-negative real parts. Hence, the infection equilibrium is locally asymptotically stable. This completes the proof.*

Parameter Estimation, Identifiability Analysis and Numerical Simulations

It is a standard practice to estimate the model parameters by fitting the model to the experimental data. However, before fitting any model to data, there is an essential step, called "identifiability analysis" that must be addressed. Identifiability analysis is the process of determining whether model parameters can be derived from a given data set. Basically, it addresses two critically important questions regarding parameter estimation problem. (i) Is the within-host model structured to reveal its parameters from the given noise-free observations? (ii) How accurate are the parameters obtained by fitting the within-host model to noisy data? Here we use the term observation to refer to continuous smooth output and the data to refer measurements of observations at the discrete time points. Clearly, observations are noise-free and data is contaminated with noise due to measurement errors. Identifiability analysis is a crucial step in parameter estimation problem and is performed in two steps, each step answering one of the critical questions. Structural identifiability analysis is the first step and addresses the the first question. It investigates if the within host model is structured to reveal its parameters from the observations. Structural identifiability is a theoretical

feature of the within-host model, therefore is a prior identifiability analysis and should be performed prior to estimating the parameters from the experimental data. The second step in identifiability analysis is practical identifiability which considers the actual experimental data with noise and should be performed after estimating the parameters. We will perform both structural and practical identifiability analysis for the selected within-host model (6).

Structural identifiability of the selected within-host model (6)

In this study, the observations of the parameter estimation problem are the within-host model's state variables. Namely, the target cells $T(t,p)$, virus $V(t,p)$, albumin $A(t,p)$, and globulin $G(t,p)$. Using the experimental data, the objective is to estimate the parameters of the model which is denoted by $p=(r,d,\rho_0,\delta,\psi_{01},\pi,c,\mu_v,\lambda_z,b,\mu_z,\lambda_A,\gamma_A,\mu_A,\lambda_G,\gamma_G,\mu_G)$. Before estimating the parameters, we need to study the structural identifiability of the model, which is defined as below.

Definition 1 *Let p and $\hat{p}$ be two distinct parameter sets of the within-host model (6). If*

$$T(t,p)=T(t,\hat{p}),\quad V(t,p)=V(t,\hat{p}),\quad A(t,p)=A(t,\hat{p}),\quad \textit{and}\quad G(t,p)=G(t,\hat{p})\quad \textit{implies}\quad p=\hat{p}$$

then we say that the within-host model (6) is structurally identifiable.

Simply put, the definition states that when the model parameters are different so are the observations. Therefore, two identical observations are only possible if the parameters are the same. There are several methods to study structural identifiability of compartmental ODE models, for an in depth review we refer[34]. We use DAISY (Differential Algebra for Identifiability of SYstems) software to determine the identifiability of the within host model (6)[35]. DAISY gives the following parameter correlations;

$$\begin{aligned} &r=\hat{r},\quad d=\hat{d},\quad \rho_0=\hat{\rho}_0,\quad \delta=\hat{\delta},\quad \pi=\hat{\pi},\quad c=\hat{c},\quad \mu_v=\hat{\mu}_v,\quad b=\hat{b},\quad \mu_z=\hat{\mu}_z,\quad \lambda_A=\hat{\lambda}_A,\\ &\gamma_A=\hat{\gamma}_A,\quad \mu_A=\hat{\mu}_A,\quad \lambda_G=\hat{\lambda}_G,\quad \gamma_G=\hat{\gamma}_G,\quad \mu_G=\hat{\mu}_G,\quad \psi_{01}\lambda_z=\hat{\psi}_{01}\hat{\lambda}_z \end{aligned} \tag{16}$$

The DAISY result (16) states that the parameters, $(r,d,\rho_0,\delta,\pi,c,\mu_v,b,\mu_z,\lambda_A,\ \gamma_A,\mu_A,\lambda_G,\gamma_G,\mu_G)$ can be identified from the observations of the target cells, $T(t,p)$, viral load, $V(t,p)$, albumin $A(t,p)$,and globulin levels $G(t,p)$. But, we can not identify ψ_{01} or λ_z uniquely. We can only identify their product $\psi_{01}\lambda_z$. Since the recruitment of immune cells, λ_z and the rate at which the immune cells attach the infected cells ψ_{01} are correlated, infinitely many parameters will yield to identical observations of the target cells, $T(t,p)$, viral load, $V(t,p)$, albumin $A(t,p)$, and globulin levels $G(t,p)$ as long as the the product $\psi_{01}\lambda_z$ is constant. Therefore the within-host model (6) is not structured to reveal its parameters from the given observations. However, since we know the parameter correlation that leads to unidentifiability, we can fix either the recruitment of immune cells, λ_z, or the rate at which the immune cells attach the infected cells, ψ_{01}, to obtain a structurally identifiable model. We summarize the structural identifiability analysis in the following proposition.

Proposition 1 *The within-host HIV protein model (6) is not structured to identify all its parameters from the observations of target cells, $T(t)$, viral load, $V(t)$, albumin, $A(t)$, and globulin, $G(t)$. If the parameters λ_z or ψ_{01} are fixed, then the within-host model (6) becomes structurally (globally) identifiable.*

First, we fix the immune cell production rate to $\lambda_z=1$ cells/(mL $\times$ day)[33] to obtain a structurally identifiable model. Then, we estimate the parameters of the within-host model (6) using the experimental data. The estimated parameters are presented in Table 4 and the model predictions are shown in Figure 2.

Practical identifiability of the selected within-host model (6)

A model which is structurally identifiable, might not be identifiable in practice when noisy data is considered in determining the parameters. In structural identifiability analyses, experimental data is not considered at all. Our analysis shows that within host model is structurally identifiable (6) and this feature of the model is based on observations, which are smooth continuous functions of the model's state variables. Simply put, structural identifiability analysis assumes infinitely many noise-free data points, whereas practical identifiability uses noisy finitely many points. Therefore, we continue with the practical identifiability analysis of the within-host model (6) using Monte Carlo simulations. The Monte Carlo simulations performed are outlined below[36,37].

1. Set the estimated parameters given in Table 4 as the true parameters, p_0. Solve the within-host model (6) numerically with the true parameters.

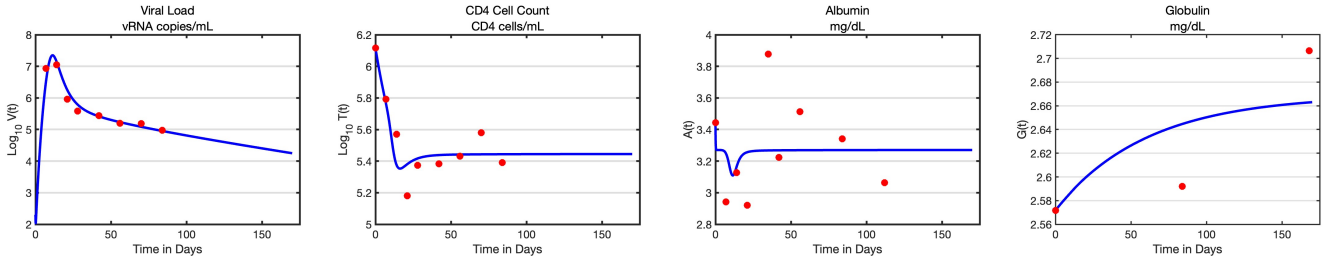


Figure 2. Fitting results: Simulation of the within-host model (6) using the estimated values given Table 4 is shown in blue curve. Red circles are the data given in Table 2. Initial conditions are $T(0) = 1304000$, $T_i(0) = 0$, $V(0) = 200$, $Z(0) = 10$, $A(0) = 3.44$, $G(0) = 2.571$.

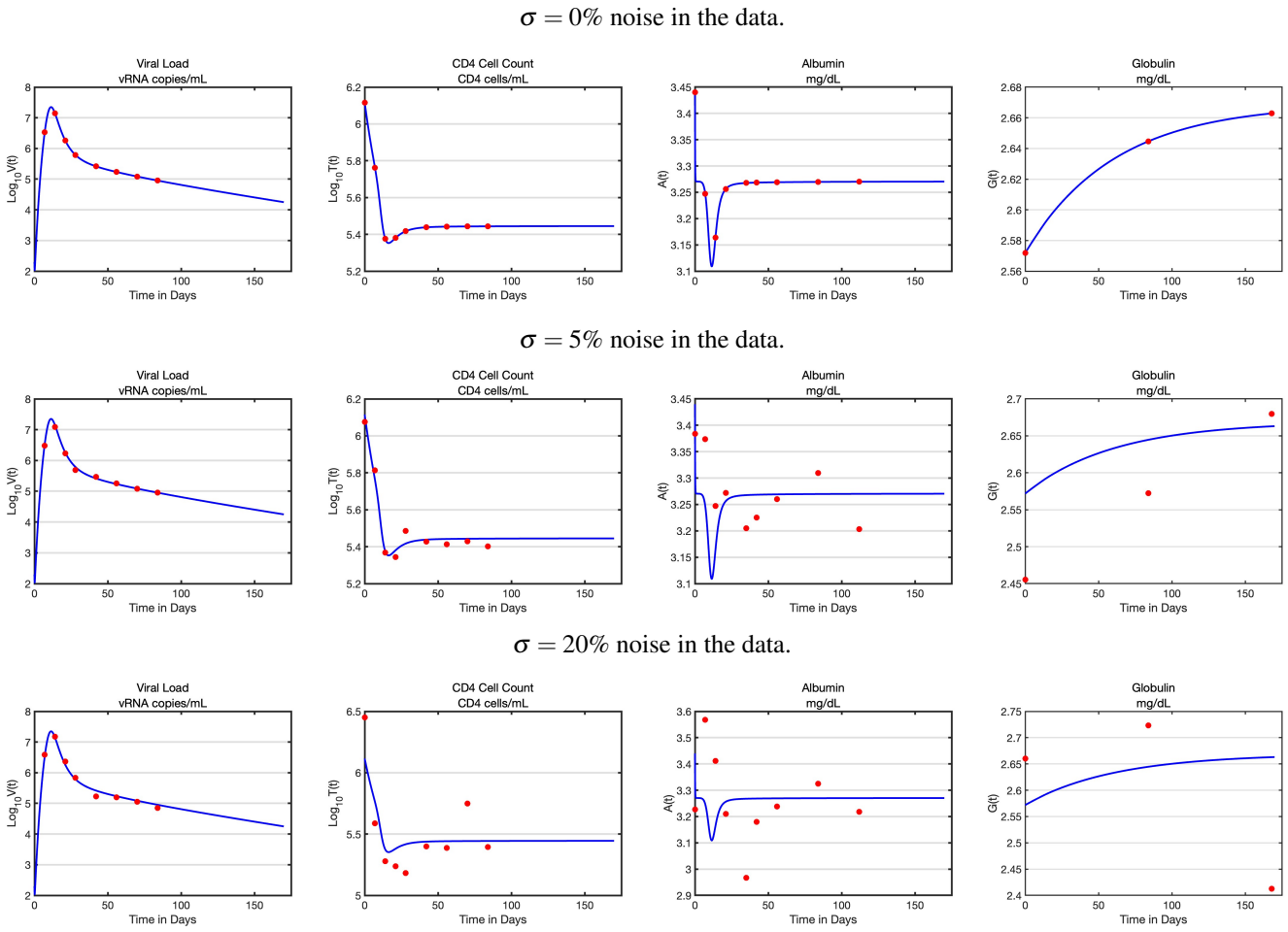


Figure 3. Monte Carlo Simulations (MCS): A single set of viral load, CD4 cell count, albumin and globulin data generated for the MCS is presented. Blue curve is the model predictions with the true parameters, and red circles are the synthetic data. The synthetic data are generated by taking the estimated parameters given in Table 4 as true parameters and generating random data with normal distribution whose mean are the model predictions and the standard deviation is $\sigma = 0$ (Top Row), $\sigma = 5$ (Middle Row), and $\sigma = 20$ (Bottom Row).

2. Generate M = 1000 synthetic datasets from a normal distribution whose mean is the model predictions at the discrete data points. Namely, generate synthetic data using the true parameters,

$$\begin{aligned} Y_1^i &= T(t_i,p_0)+E_i \quad i=1,\cdots,n_1 \qquad Y_2^j = V(t_j,p_0)+E_j \quad j=1,\cdots,n_2 \\ Y_3^k &= A(t_k,p_0)+E_k \quad k=1\cdots,n_3 \qquad Y_4^l = G(t_l,p_0)+E_l \quad l=1,\cdots,n_4 \end{aligned} \tag{17}$$

where each measurement error E_i, E_j, E_k, E_l is normally distributed with zero mean and standard deviation σ. Figure 3 shows a single synthetic data generated for $\sigma = 0\%, 5\%, 20\%$.

3. Fit the within-host model (6) to data generated in step 2 and obtain estimated parameters p_j for each data set $j = 1, 2, \ldots M$.

4. Calculate the average relative estimation error (ARE) for each parameter by

$$\text{ARE}(p^{(k)}) = 100\% \frac{1}{M} \sum_{j=1}^{M} \frac{|p_0^{(k)} - p_j^{(k)}|}{p_0^{(k)}} \tag{18}$$

where $p_0^{(k)}$ is the k^{kt} parameter in p_0 and $p_j^{(k)}$ is the k^{th} parameter in the set p_j.

5. Repeat steps 1 through 4 with higher noise levels. That is, at each iteration take σ = 0%, 1%, 5%, 10%, 20%.

The practical identifiability performed with MCS is based on generating synthetic data using the estimated parameters, therefore it is a local result for the estimated parameters. Since, the true parameter set that generated the synthetic data is known, we can compute the average estimation error of each parameter. AREs describes how noise in the data affects the parameters. The AREs gradually increase as the noise level in data increases. The sensitivity of this increase determines the practical identifiability of the parameters. A practically unidentifiable parameter is very sensitive to noise in the data. In general, the AREs of those parameters will increase rapidly than the corresponding noise level. We introduce the following definition of practical identifiability.

Definition 2 *Let p be the parameter set of the within-host model (6). If the ARE of the k^{th} parameter in p is bounded by the measurement error, then we say that the parameter p^k is practically identifiable. Specifically, if*

$$ARE(p^{(k)}) \leq c\sigma \quad with \quad 0 < c < 10$$

then we say that the parameter p^k is practically identifiable.

Examining the AREs presented in Table 5, and using Definition 2 we conclude that only the parameters, $r, d, \rho_0, \pi, c, \psi_{01}$ and μ_z are practically identifiable. All the other parameters have significantly high AREs and therefore are practically unidentifiable. AREs of parameters increase as the noise level in data increases, but the parameters μ_v and γ_G have already have very high AREs when there is only 1% noise in data. It seems like noise in the data is not the only source of high AREs for s μ_v and γ_G. Even though the within-host model (6) is structurally identifiable, when λ_z is fixed, we see that most of its parameters are practically unidentifiable. We suspect that the practical unidentifiably might be due to lack of enough data points. To test this hypothesis, we rerun the MCS with high frequency data points, where 10 measurements is collected each day for 170 days, which results in total of 1700 data points for each observations; viral load, target cells, albumin and globulin. AREs of the parameters with high frequency data is presented in Table 6. Based on the results in Table 6, we conclude that when there is enough data points, all the parameters of the within-host model (6) except μ_v and γ_G are practically identifiable.
For the next numerical simulation, we studied the effect of varying albumin and globulin intake on the viral load. Both albumin and globulin intake were varied by two orders of magnitude. We select λ_G at random from a uniform distribution $[0.005, 0.5]$ and λ_A at random from a uniform distribution $[3, 300]$. The simulation results are shown in Figure 4. Clearly, albumin and globulin intakes have significant impact on the albumin and globulin levels. However, their impact on the viral load is minimal.

Discussion

Nutrition plays important role in the immune system and the immune system is of essence for protecting us from infectious diseases. Thus there is an intrinsic link between nutrition and infection diseases. Despite that extensive modeling has been done on infectious diseases and their interaction with the immune system, and significant work has been done on modeling nutrition and the dynamics of its components within-host, virtually no models have been introduced covering the interplay of infectious diseases and nutrition. In this paper we consider a within host HIV model and the interaction of the virus with the

Table 5. Monte Carlo Simulation (MCS) Results: *AREs* of the parameters of the within-host model (6) is presented for each noise level. Synthetic data is generated at the discrete time points of actual data used for parameter estimation (see Figure 3).

Parameters	$\sigma = 1\%$	$\sigma = 5\%$	$\sigma = 10\%$	$\sigma = 20\%$
r	2.1%	9.9%	21.6%	82.5%
d	2.0%	9.6%	20.4%	83.1%
ρ_0	5.2%	20.6%	36.0%	87.4%
δ	227.9%	959.9%	2026.8%	7003.1%
π	12.4%	44.2%	84.2%	169.5%
c	15.5%	52.1%	78.2%	123.4%
ψ_{01}	10.7%	28.6%	47.7%	85.3%
b	55.3%	485.9%	3801.5%	9706.8%
μ_z	6.6%	24.7%	42.7%	75.7%
μ_v	17376%	14194%	29714%	18721%
λ_A	31.1%	108.2%	243.8%	977.2%
γ_A	33.9%	148.3%	424.6%	2409.5%
μ_A	31.1%	108.0%	243.2%	952.6%
λ_G	104.9%	491.2%	768.1%	2910.8%
γ_G	6275.8%	11938%	19139%	22374%
μ_G	105.7%	494.3%	773.3%	3075%

Table 6. Monte Carlo Simulation (MCS) Results: *AREs* of the parameters of the within-host model (6) is presented for each noise level. Synthetic data is generated with high frequency data, meaning that there is 10 data points for each day for 170 days.

Parameters	$\sigma = 1\%$	$\sigma = 5\%$	$\sigma = 10\%$	$\sigma = 20\%$
r	0.1%	0.7%	1.4%	2.8%
d	0.1%	0.6%	1.3%	2.7%
ρ_0	0.2%	1.4%	2.9%	4.6%
δ	8.3%	41.8%	83%	124.2%
π	0.3%	1.7%	3.4%	5.7%
c	0.1%	0.4%	1%	2%
ψ_{01}	0.1%	0.5%	1.1%	2%
b	2.5%	14.3%	28.2%	55.7%
μ_z	0.1%	0.6%	1.3%	2.3%
μ_v	6152.6%	14112%	26711%	18392%
λ_A	1.4%	8.7%	19%	32.8%
γ_A	1.6%	9.9%	21.6%	39.6%
μ_A	1.4%	8.7%	19%	32.8%
λ_G	1.9%	11.5%	23.1%	50.3%
γ_G	111.5%	435.2%	803.3%	1441.8%
μ_G	1.9%	11.6%	23.4%	50.9%

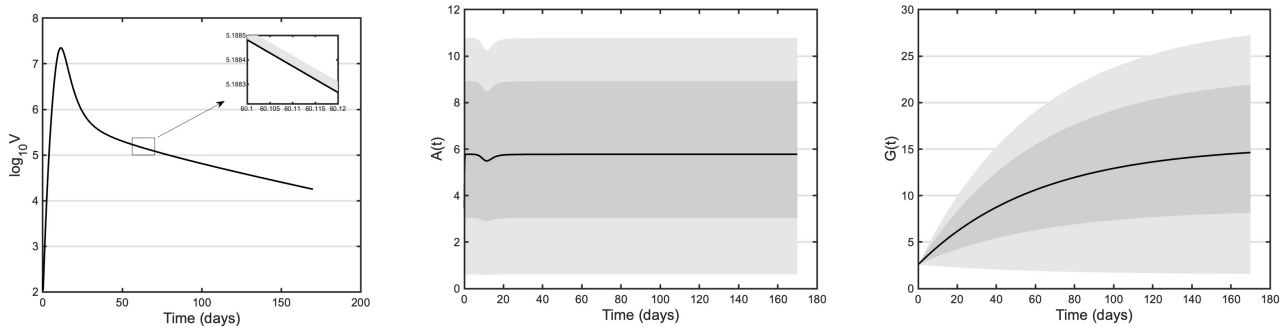


Figure 4. Impact of albumin and globulin intake: The albumin and globulin intakes are varied by two magnitudes of order around their fitted value. λ_G is selected randomly from a uniform distribution $[0.005, 0.5]$ and λ_A from a uniform distribution $[3, 300]$. The grey shaded areas represent the 95% of the resulting simulations.

protein in the body, namely the albumin and the globulins. We introduce a global model that incorporates target healthy and infected cells, viral load, immune cell response, and the concentrations of albumin and globulin. Because the impact of the two components of protein on the infection of healthy cells and the killing of the virus by the immune system are not known, we hypothesize a generic functional form for the terms. Further we assume a linear dependence of CD4 and CD8 cells recruitment on the globulins. We use data on CD4 cells, viral load, albumin and globulin concentrations in rhesus monkeys to determine the specific form of dependence of the incidence term and infected T cells clearance term on the protein concentrations, as well as the form of the recruitment terms of target cells and immune cells. We specify 17 submodels of the general model, fit all 17 models to the data, and use the Akaike Information Criterion (AIC) to select the best model. We find out that the simplest model fits the data best and solely explains the data. According to the best fitted model albumin and globulins do not impact the new infections term, nor do they impact the clearance of infected T cells by the immune system term. Further the recruitment of CD4 and CD8 cells also does not depend on globulins. The best fitted model outperforms the remaining 16 models significantly and has a relative likelihood of 0.996, therefore explaining the data by itself[23].

We further analyze the best fitted model. We find out that it has a unique infection-free equilibrium, and we compute the reproduction number of the infection $\mathscr{R}_0$. The infection-free state is locally stable if the reproduction number is smaller than one. If the reproduction number is greater than one, the infection-free equilibrium is unstable and there is a unique infection equilibrium. The unique infection equilibrium is locally stable whenever it exits.

We fit the best fitted model to the data and determine its parameters. Before that we address the question whether the parameter estimation problem is well posed and we study the structural identifiability of the model relative to the given data sets. We find out that all parameters of the model are structurally identifiable except λ_Z and ψ_{01} but their product is identifiable. Hence, we fix the immune cell production rate λ_Z at a value obtained from the literature. Fixing the value of λ_Z makes all parameters of the model structurally identifiable. Further we perform practical identifiability of the model using Monte Carlo Simulations. We find that with the given data only seven of the 16 parameters have Average Relative Estimation Errors (AREs) within an order of magnitude of the noise errors and are practically identifiable. We surmise that the reason for the lack of practical identifiability of most of the parameters is the low quality of the data. To test this hypothesis we perform Monte Carlo Simulations with synthetic data which are sampled a lot more frequently. We find out that with the high quality data only the parameters μ_v and γ_G are not practically identifiable. We conclude that collecting more consistent and more frequently sampled data for this problem can lead to more reliably estimated parameters.

Availability of Data and Materials

The datasets generated and/or analyzed during the current study are available in the github repository, https://github.com/NecibeTuncer/HIV_and_Nutrtion.

References

1. of Health, N. I. HIV and Nutrition and Food Safety. https://https://hivinfo.nih.gov/understanding-hiv/fact-sheets/hiv-and-nutrition-and-food-safety. Accessed: 2023-01-12.

2. Scrimshaw, N. S. & SanGiovanni, J. P. Synergism of nutrition, infection and immunity: An overview. *The Am. J. Clin. Nutr.* **66**, 464S–477S (1997).

3. Colecraft, E. HIV/AIDS: nutritional implications and impact on human development. *Proc. Nutr. Soc.* **67**, 109–113 (2008).
4. Baisel, W. Nutrition and immune function: Overview. *The J. Nutr.* **126**, 2611S–2615S (1996).
5. Thuppal, S. V., Jun, S., Cowan, A. & Bailey, R. L. The nutritional status of HIV-infected US adults. *Curr. Dev. Nutr.* **1**, 1–6 (2017).
6. Schenker, N. *et al.* Multiple imputation of missing dual-energy X-ray absorptiometry data in the National Health and Nutrition Examination Survey. *Stat. Med.* **30**, 260–276 (2011).
7. Caudill, S. P. Use of pooled samples from the national health and nutrition examination survey. *Stat. Med.* **31**, 3269–3277 (2012).
8. Vasconcellos, M. T. L. d., Silva, P. L. d. N. & Anjos, L. A. d. Sample design for the nutrition, physical activity and health survey (PNAFS), Niterói, Rio de Janeiro, Brazil. *Estadistica* **65**, 83–98 (2013).
9. Ma, J., Chan, W., Tsai, C.-L., Xiong, M. & Tilley, B. C. Analysis of transtheoretical model of health behavioral changes in a nutrition intervention study—a continuous time Markov chain model with Bayesian approach. *Stat. Med.* **34**, 3577–3589 (2015).
10. Novotny, J. A., Green, M. H. & Boston, R. C. (eds.) *Mathematical modeling in nutrition and the health sciences*, vol. 537 of *Advances in Experimental Medicine and Biology* (Kluwer Academic/Plenum Publishers, New York, 2003). Papers from the 7th Conference held at Pennsylvania State University, University Park, PA, July 29–August 1, 2000.
11. Juillet, B., Salomon, J., Tomé, D. & Fouillet, H. Development and calibration of a modeling tool for the analysis of clinical data in human nutrition. In *CEMRACS 2004—mathematics and applications to biology and medicine*, vol. 14 of *ESAIM Proc.*, 124–155 (EDP Sci., Les Ulis, 2005).
12. Drew, D. A. A mathematical model for nutrient metabolic chemistry. *Appl. Math.* **9**, 647–671 (2018).
13. Chudtong, M. & De Gaetano, A. A mathematical model of food intake. *Math. Biosci. Eng.* **18**, 1238–1279 (2021).
14. Shah, N. H., Thakkar, F. A. & Yeolekar, B. M. Optimal control model for poor nutrition in life cycle. *Dyn. Contin. Discret. Impuls. Syst. Ser. B Appl. Algorithms* **24**, 387–400 (2017).
15. Baleanu, D., Jajarmi, A., Bonyah, E. & Hajipour, M. New aspects of poor nutrition in the life cycle within the fractional calculus. *Adv. Differ. Equ.* Paper No. 230, 14 (2018).
16. Mathematical nutrition models. https://https://www.nutritionmodels.com. Accessed: 2023-01-11.
17. Coburn, S. & Townsend, D. W. *Mathematical Modeling in Experimental Nutrition: Vitamins, Proteins, Methods* (Academic Press, San Diego, 1996).
18. Nowak, M. & May, R. *Virus dynamics: mathematical principles of immunology and virology* (Oxford University Press, New York, USA, 2000).
19. de Boer, R. & Perelson, A. Target cell limited and immune control models of HIV infection: A comparison. *J. Theor. Biol.* **190**, 201–214 (1998).
20. Baral, S., Antia, R. & Dixit, N. M. A dynamical motif comprising the interactions between antigens and CD8 T cells may underlie the outcomes of viral infections. *PNAS* **116**, 17393–17398 (2019).
21. SelfDecode. Albumin/Globulin Ratio: High and Low Ratios + Normal Range. https://labs.selfdecode.com/blog/albumin-globulin-ratio/#Normal_AlbuminGlobulin_Ratio. Accessed: 2021-08-04.
22. Haefner, J. W. *Modeling biological systems: principles and applications* (Springer, New York, 2005), second edn. With 1 CD-ROM (Windows and UNIX).
23. Burnham, K. P. & Anderson, D. R. *Model Selection and Multi model Inference: A Practical Information-Theoretic Approach* (Springer, 2002).
24. Healthline. Total Protein Test. https://www.healthline.com/health/total-protein#results. Accessed: 2021-07-18.
25. YS, S. *et al.* Serum protein electrophoresis: any role in monitoring for antiretroviral therapy? *Afr. Heal. Sci.* **10**, 138–143 (2010).
26. Mehta, S. H., Astemborski, J., Sterling, T. R., Thomas, D. L. & Vlahov, D. Serum albumin as a prognostic indicator for HIV disease progression. *AIDS Res. Hum. Retroviruses* **22**, 14–21 (2006).
27. Ronit, A. *et al.* Serum albumin as a prognostic marker for serious non-AIDS endpoints in the strategic timing of antiretroviral treatment (START) study. *The J. Infect. Dis.* **217**, 405–412 (2017).

28. Conway, J. M. & Perelson, A. S. Post-treatment control of HIV infection. *PNAS* **12**, 5467–5472 (2015).

29. Duggal, S., Chugh, T. D. & Duggal, A. K. HIV and malnutrition: effects on immune system. *Clin. Dev. Immunol.* **2012** (2012).

30. Vaidya, N. K., Ribeiro, R. M., Perelson, A. S. & Kumar, A. Modeling the effects of morphine on simian immunodeficiency virus dynamics. *PLOS Comput. Biol.* **12**, DOI: 10.1371/journal.pcbi.1005127 (2016).

31. Graham, S. M., Holte, S., Kimata, J. T., Wener, M. H. & Overbaugh, J. A Decrease in Albumin in Early SIV Infection Is Related to Viral Pathogenicity. *AIDS Res. Hum. Retroviruses* **25**, 433–440 (2009).

32. Sun, J. *et al.* The DTI changes and peripheral blood test results corroborate the early brain damage of SIV-infected rhesus. *Radiol. Infect. Dis.* **6**, 8–14 (2019).

33. Conway, J. M. & Perelson, A. S. Post-treatment control of hiv infection. *Proc. Natl. Acad. Sci.* **112**, 5467–5472, DOI: 10.1073/pnas.1419162112 (2015). https://www.pnas.org/doi/pdf/10.1073/pnas.1419162112.

34. Miao, H., Xia, X., Perelson, A. S. & Wu, H. On identifiability of nonlinear ode models and applications in viral dynamics. *SIAM Rev.* **53**, 3–39, DOI: 10.1137/090757009 (2011).

35. Bellu, G., Saccomani, M., Audoly, S. & D'Angio, L. Daisy: a new software tool to test global identifiability of biological and physiological systems. *Comput. Methods Programs Biomed.* **88**, 52–61 (2007).

36. Tuncer, N., Mohanakumar, C., Swanson, S. & Martcheva, M. Efficacy of control measures in the control of Ebola, Liberia 2014-2015. *J. Biol. Dyn.* **12**, 913–937 (2018).

37. Tuncer, N. & Martcheva, M. Determining reliable parameter estimates for within-host and within-vector models of zika virus. *J. Biol. Dyn.* **15**, 430–454, DOI: 10.1080/17513758.2021.1970261 (2021).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SuplementaryFile.pdf