Analysis of HIV-1 Compartmental Model Parameters using Bayesian MCMC Estimation

Erwing Fabian Cardozo, Ryan Zurakowski, Nii Attoh-Okine

Abstract— In previous work Bayesian Markov-Chain Monte Carlo techniques were used to identify HIV dynamic parameters from patient data. This study used viral load data from HIV patients during interrupted cycles of antiretroviral drugs. The data were fit to a well-established single-compartment model of HIV dynamics. Experimental evidence supports the use of a new compartmental model that includes the re-circulation of T-cells between anatomical reservoirs. If the infection dynamics are indeed compartmentalized, it is not clear how this would affect the estimated values of the dynamic parameters. In this study, we identify parameters of the simple, one-compartment model using data generated by the complex spatial model, and investigate the bias in parameter values introduced by this method. Multiple instances of simulated noisy data was generated from the spatial model using parameter estimates from previous studies. Markov-Chain Monte Carlo methods were used to identify parameter values for the simple model from the simulated data, and the identified values were compared with the true values to determine the existence of bias. The maximum likelihood bias in the median estimates of the proliferation, death and infection rate parameters for target T-cells and virion production rates were on average smaller than one standard deviation of the reported parameter uncertainties for the simple model. The median estimates of the death rate of infected cells and the efficacy of the drug exhibited an average positive bias (the posteriors were larger than the priors) that was larger than one-standard deviation of the prior for all patients. Neglecting the spatial dynamics does not seem to significantly affect the estimation of the proliferation, death, and infection rate parameters for target T-Cells. Conversely, the values of the infected cell death rates and drug efficacies exhibited a consistent bias when estimated using the simplified model. Neglecting spatial dynamics will result in a consistent overestimation of the values of these parameters.

I. INTRODUCTION

Low-level viremic persistence is one of the main barriers to HIV eradication. We have previously presented a model describing how ongoing viral replication can be detected using 2-LTR circles [1]. When compared to the patient data from the raltegravir intensification trial reported in [2] the model shows that significant uncontrolled viremia must be occurring in anatomical sites poorly penetrated by the antiviral drugs in a subset of the patients on combination antiretroviral therapy (cART). We proposed a compartmental model of viral dynamics including transport of the cells and virus between compartments [3], [4]. This model was able to replicate the observed behavior in the clinical trial, and support the finding of persistent viremia associated with the 2-LTR measurements. The parameters used for HIV / T Cells proliferation, infection and clearance rates estimated from clinical data by the MCMC technique in [5]. However, it is not clear how the use of a simplified, non-compartmental model in [5] will affect the estimated values of these dynamic parameters. Identification of the full spatial model from the clinical data is not possible, due to the (technically infinite) degrees of freedom in the spatial model.

In this study, therefore, we analyze the bias introduced by the model mismatch by identifying the parameters of the simple model using simulated data generated from the compartmental model. The simulated data is corrupted by the measurement noise process described in [6]. The identified parameter values are then correlated to the parameter values used to generate the simulated data. We present the design for the creation of virtual clinical trials and a description of the models used, the comparison of the maximum likelihood estimate from the simulated data and the posterior distributions from [5], and a discussion of the results.

II. METHODS

A. Creation of Simulated Data

This study seeks to find a correlation between the parameters of a compartmental model presented in [3], [4] with the parameters of the simplified model presented in [5]. To do that we will use a compartmental model to create the simulated data using the maximum likelihood estimation obtained in Luo’s study. Then, we will use the MCMC Metropolis Hasting technique to obtain new posterior distributions of the same parameters fitting the simulated data to the simplified model, using the posterior distributions reported in Luo’s work as the priors.

a) Compartmental Model: The model used to create the simulated data was initially presented in [4]. This model is a compartmental discretization of a spatial PDE model of infection, with species moving between compartments in a diffusion-like manner. The equations model the follicle sites of the lymph nodes as remote compartments [3]. The model has the following configuration: one main compartment representing the blood, the HEVs and the lymphatic vessels and N secondary compartments representing all the lymph node paracortex/follicle sites in the human body. These N compartments have no connections between them but only with the main compartment. Each of the N secondary spherical compartments is assumed to be subdivided into n − 1 concentric spheres where only the most external one is connected with the blood compartment. Since all N compartments have the same geometrical configuration, we use...
an ODE model of \(4n\) equations with the first compartment having the form

\[
\begin{align*}
\dot{x}_1 &= \lambda - dx_1 - \beta x_1 v_1 (1 - \eta_{RTI} u_{RTI} \theta_1) \\
&+ N \frac{D_{1,2}}{l_1} \frac{A_{1,2}}{V_1} (x_2 - x_1) \\
\dot{y}_1 &= \beta x_1 v_1 (1 - \eta_{RTI} u_{RTI} \theta_1) - a y_1 + y_e \\
&+ N \frac{D_{1,2}}{l_1} \frac{A_{1,2}}{V_1} (y_2 - y_1) \\
\dot{v}_1 &= \gamma (1 - \eta_{PI} u_{PI} \varphi_1) y_1 - \omega v_1 \\
&+ N \frac{D_{1,2}}{l_1} \frac{A_{1,2}}{V_1} (v_2 - v_1),
\end{align*}
\]

where \(A_{1,2}\) represents the surface area of the paracortex/follicle site in the lymph node, \(V_1\) represents the volume of the main compartment and \(\frac{D_{1,2}}{l_1} = \frac{D_{2,1}}{l_2}\) and \(\frac{D_{1,2}}{l_1}\) represents the effective diffusivity of T-Cells and HIV virions between the main compartment and the paracortex/follicle site. Furthermore, for each concentric layer \(s = 2, \ldots, n\), of each of the \(N\) compartments have the form

\[
\begin{align*}
\dot{x}_s &= \lambda - dx_s - \beta x_s v_s (1 - \eta_{RTI} u_{RTI} \theta_s) (1 - \eta_{II} u_{II} \xi_s) \\
&+ \sum_{i \neq s} \frac{D_{i,s}}{l_s} \frac{A_{i,s}}{V_s} (x_i - x_s) \\
\dot{y}_s &= \beta x_s v_s (1 - \eta_{RTI} u_{RTI} \theta_s) (1 - \eta_{II} u_{II} \xi_s) \\
&- a y_s + y_e + \sum_{i \neq s} \frac{D_{i,s}}{l_s} \frac{A_{i,s}}{V_s} (y_i - y_s) \\
\dot{v}_s &= \gamma (1 - \eta_{PI} u_{PI} \varphi_s) y_s - \omega v_s \\
&+ \sum_{i \neq s} \frac{D_{i,s}}{l_s} \frac{A_{i,s}}{V_s} (v_i - v_s),
\end{align*}
\]

where \(A_{i,s}\) represents the surface area between each layer in the sphere, \(V_s\) represents the volume of the layer, and \(\frac{D_{i,s}}{l_s} = \frac{D_{i,s}}{l_i}\) and \(\frac{D_{i,s}}{l_s}\) represents the effective diffusivity of T-Cells and HIV virions between layers.

In each compartment CD4+ T target cells \(x_s\), are produced at a rate \(\lambda\), decay at a rate \(d x_s\), and are infected at a rate \(\beta x_s v_s\). The infection is inhibited by reverse transcriptase inhibitors (RTI) \(u_{RTI}\) and integrase inhibitors (II) \(u_{II}\) with maximum effectiveness of \(\eta_{RTI}\) and \(\eta_{II}\) respectively. We hypothesize that the efficacy of the drug depends on the domain; therefore, we include a spatial dependence drug penetration distribution \(\theta_s\) and \(\xi_s\) for RTI and II respectively. Moreover, exogenous sources, including the activation of latent infected T-cells, contribute to create actively infected cells at a rate \(y_e\). The infected cells \(y_s\) disappear at a rate \(a y_s\) and produce virions at a rate \(\gamma y_s\). This viral production is interrupted by protease inhibitors \(u_{PI}\) with maximum efficiency \(\eta_{PI}\). The activity of this drug is affected spatially by a penetration distribution \(\varphi_s\). Finally, viral load exponentially decays with a rate \(\omega v_s\). The parameters related to HIV/T-Cells infection, proliferation and clearance, and drug efficiency are drawn from the posterior distributions proposed in [5]. The parameters related to the diffusion term and drug penetration were selected according to the assumptions in [3].

b) Uncertainty in HIV measurements: To model the uncertainty of the measurements we use the following assumptions. We assume that the measurements are sampled from the virus in the first compartment of the model in equation 1. The original study data were HIV-1 RNA PCR measurements; we assume the uncertainty related to that technique which has a log-normal distribution with variance that increases as the expected number of virions number decrease. Finally, we assume that each measurement is independent from patient to patient, and form time to time. This yield to the model for patient \(i\) virus measurements, as proposed in [1]:

\[
m_i(t_{i,k}) = \max\{v_1(t_{i,k}, \mathbf{p}) + e_{i,k}, 50\}.
\]

In this model \(m_i(t_{i,k})\) represents the simulated measurement of viral load in compartment 1 for patient \(i\) at time \(t_{i,k}\) and \(e_{i,k}\) is the log-normally distributed zero-mean sample variance, where \(\sigma(v)\) is defined as the derivation in [6] for HIV-1 PCR measurements:

\[
\sigma(v) = 10^{-0.21 - 0.24 \times \log_{10}(v)}.
\]

c) Study design: We created simulated data with the same treatment interruption scheme as in [7], but with viral load measured every 2 days. For each patient we use the Maximum Likelihood estimated values obtained in [5] to generate the data. Figure 1 shows the virus load generated with the parameters estimated using data of patient 2. In the figure, the graph in blue presents the simulation of \(v_1\) in equation 1, and in red the sampled data with the inclusion of the error modeled in equation 3.

![Fig. 1](image)

B. Model to fit

Using the simulated data, we want to estimate the parameters of the following model,

\[
\begin{align*}
\dot{x} &= \lambda - dx - \beta xv(1 - \eta u) \\
\dot{y} &= \beta xv(1 - \eta u) - a y + y_e \\
\dot{v} &= \gamma y - \omega v,
\end{align*}
\]

as identified in [5]. The model presents the same dynamics as the model in equations 1 and 2 but without the diffusion terms (See Table 1 for a description of the parameters). According to [5] is possible to identify the parameters \(\lambda, \beta, a, \gamma, \eta\) and \(y_e\). The parameter \(d\) is also identifiable as
TABLE I
PARAMETER DEFINITIONS FOR HIV DYNAMICS FROM [5]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda )</td>
<td>Proliferation rate of CD4+ T cells.</td>
<td>( \text{cells mL}^{-1} \text{day}^{-1} )</td>
</tr>
<tr>
<td>( d )</td>
<td>Death rate of CD4+ T cells.</td>
<td>( \text{day}^{-1} )</td>
</tr>
<tr>
<td>( \beta )</td>
<td>Infection rate.</td>
<td>( \text{copies cell}^{-1} \text{day}^{-1} )</td>
</tr>
<tr>
<td>( a )</td>
<td>Death rate of active infected cells.</td>
<td>( \text{day}^{-1} )</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>Rate of free virus production.</td>
<td>( \text{copies mL}^{-1} \text{day}^{-1} )</td>
</tr>
<tr>
<td>( \omega )</td>
<td>Clearance rate for the free virus.</td>
<td>( \text{day}^{-1} )</td>
</tr>
<tr>
<td>( \eta )</td>
<td>Drug efficacy.</td>
<td>%</td>
</tr>
<tr>
<td>( y_c )</td>
<td>Reservoir activation rate.</td>
<td>( \log_{10} \left( \frac{\text{cells}}{\text{mL} \times \text{day}} \right) )</td>
</tr>
</tbody>
</table>

For the prior distribution of the parameters in the second stage, we choose informative distributions drawn from the posteriors estimated in [5]. This is the information that we want to compare with the distributions to estimate using MCMC. Thus, for the parameter vector, we define the priors for patient \( i \) as

\[
p_i \sim \mathcal{LN}(\mu_i, \Sigma_i). \tag{7}
\]

Here \( \mu_i \) and \( \Sigma_i \) represent the parameter mean and standard deviation for each patient \( i \) as reported in [5] and presented in table II. Commonly, in order to estimate \( \mu_i \) and \( \Sigma_i \) it is defined a third stage with hyperprior distributions [8]–[11]. Here we use known values for them estimated previously in [5] as informative priors. That is why we do not use a third stage. Thus, with the definition of the likelihood and the priors from equations (6) and (7), recalling Bayes’ theorem, we assume that the posterior distribution of the parameters given the complete set of data for all patients in all time \( M \), would have the form,

\[
\Pr(\lambda, \beta, a, \gamma, \eta, y_c | M) \propto \prod_{i=1}^{N} \mathcal{L}(p_i|m_{ik}) \times \prod_{i=1}^{N} \mathcal{LN}(\mu_i, \Sigma_i). \tag{8}
\]

Notice that there is no analytical solution for equation 6 and that computation of posterior distribution would involve solving multidimensional integrals. Here, we compute the posterior distribution with MCMC using Metropolis Hasting algorithm to update in each iteration the vector \( p_i \) until the Markov chain converges.

III. RESULTS

A Markov-Chain Monte-Carlo method was used to fit the system in the equation (5) to the simulated clinical trials data for four patients similar to the Autovac study with parameters \( \lambda, d, \beta, a, \gamma \) and \( y_c \) described in table I. The objective is to compare the Maximum Likelihood (ML) estimate for the simulated data with the previous posterior distribution of each patient obtained in [5].

The chain was run for 1,000,000 iterations, and the first 200,000 were discarded to allow for chain convergence. Figure 3 presents the simulations of the model with the maximum likelihood estimate for the patients, and figure 4 depicts the likelihood values when the chain converges. The entire posterior distributions are depicted in figures 5 to 9, and the analysis for each parameter is discussed below.

d) Proliferation and clearance rates of Target CD4+ T Cells (\( \lambda, \beta \)).: With respect to the parameters related to target cells proliferation and death, for all patients the bias of the estimates was almost zero and smaller than one standard deviation of the prior distribution as presented in Table II. The values of the maximum likelihood for \( \lambda \) were (41.02 – 755.8) \( \text{cells mL}^{-1} \text{day}^{-1} \), inside the range of the maximum likelihood estimated (35 – 760) \( \text{cells mL}^{-1} \text{day}^{-1} \). Therefore, we can initially suggest that the prior distributions used for \( \lambda \) and for \( d \) are good estimates for the compartmental model.
Fig. 3. Simulation of simplified model with the ML estimate compared with simulated data.

Fig. 4. Likelihood estimation for the last 800,000 iteration in the Metropolis-Hasting algorithm for four patients.

Fig. 5. Posterior distributions of the parameters for Patient 1.

Fig. 6. Posterior distributions of the parameters for Patient 2.

Fig. 7. Posterior distributions of the parameters for Patient 3.

Fig. 8. Posterior distributions of the parameters for Patient 4.

Fig. 9. Posterior distributions of the parameters for Patient 5.
e) Density Dependent Infection Rate ($\beta$).: For the infection rate, as $\lambda$ and $d$, the maximum likelihood and the mean of the posterior distribution were very close to the priors, except for patient 2., with a bias smaller to one-standard deviation, even much closer to zero than the standard deviation. For patient 2, notice that the bias was between one and two standard deviations (of the prior); furthermore, the maximum likelihood estimate was inside the Maximum likelihood inter-patient range in [5] from $2 \times 10^{-6}$ to $6 \times 10^{-6}$ \( \frac{mL}{\text{viruses/day}} \). Thus, we can conclude that the prior distribution used is still a reliable parameter estimation for this parameter in the compartmental model. Moreover, the results of this parameter and the two previous ones, we may suggest that the dynamics of target cells is not significantly changed by the inclusion of diffusion terms.

f) Free virus production by infected cells rate ($\gamma$).: For the virions production rate is evident that the bias of the mean is zero, and the bias of the maximum likelihood estimate is very close to zero. Hence, we can suggest that given that for the compartmental model we assumed that free virions cannot penetrate the sanctuary site directly but through infected cells, the dynamics of virions is also not significantly affected by the inclusion of diffusion terms.

g) Infected cells clearance rate ($\alpha$).: For infected cells death rate, we found that, except for patient 3, the maximum likelihood estimate was inside the inter-patient range (0.18 - 2.3) \( \frac{1}{\text{day}} \) estimated in [5]. For patient 3 the ML value was 3.7 \( \frac{1}{\text{day}} \), significantly grater than the mentioned range, and for the most accepted range (0.7 - 1.3) \( \frac{1}{\text{day}} \). Further, for the same patient, the bias was around 5-fold the standard deviation of the prior. Even for patient 1 and 2, the bias of the ML estimates were almost equal or greater than one-standard deviation. In fact only for patient 4, the bias was significantly low. Notice also that in general the ML estimates values were greater than the mean of the prior distribution. therefore, as an initial conclusion we suggest that we can not use the prior distribution for the compartmental model.

h) Efficacy of the drug ($\eta$).: For this parameter, for patient one and three the maximum likelihood estimate decreased. For the other three patients, however, the values increased significantly, approaching values of 100% drug efficacy. Also, the median of the maximum likelihood of all patients was 0.94, significantly greater than the median of the prior of 0.75. Than means that, despite the fact that the ML estimates were smaller than one-standard deviation of the priors, the efficacy of the drug parameter for the compartmental model might be in a range smaller than the used in the simplified model. The results of the estimate of $\eta$ and $\alpha$ might be biologically interpreted as including the contribution not only of the drug-induced/natural death of infected cells but also the contribution of circulation of the cells to sanctuary sites as represented by the compartmental model.

i) Contribution of the reservoir to actively infected cells ($\eta_r$).: The reservoir contribution rate was the only parameter that the bias in the maximum likelihood estimate for the posterior distribution was bigger than one-standard deviation of the prior for all patients. Despite the bias of the estimate was in the range of 1.1786 to 3.8942, the ML inter-patient values were very homogenous in the range (0.0164 - 0.0196) \( \frac{\text{cells}}{\mu L/day} \). This result is in contrast against the estimates previously done in [5] where they were very heterogenous among patients. Therefore, this bias suggest that the parameter representing the contribution of the reservoirs in the simplified model might include also the contribution of recirculating infected cells returning to the blood.

IV. DISCUSSION

The main aim of this report was to quantify the relationship of a general simplified HIV infection model parameters with the parameters of a compartmental model that represents mechanisms of T-Cell infection by HIV virus and their recirculation between compartments.

The compartmental and simplified model share the same distributions for parameters related to free virus and target CD4+T Cells. Moreover, the most affected parameters are the ones related to the infected cells. In general the estimates of parameters $\alpha$, $\eta$ and $\gamma$, when fit to a simplified model, are biased with respect to their values in a more complicated spatial model.

REFERENCES


<table>
<thead>
<tr>
<th>Patient</th>
<th>Parameter</th>
<th>Units</th>
<th>Mean Prior</th>
<th>Mean Posterior</th>
<th>ML Posterior</th>
<th>Standard Deviation Prior</th>
<th>Bias Mean</th>
<th>Bias ML</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\log_{10}(\lambda)$</td>
<td></td>
<td>1.77</td>
<td>1.77</td>
<td>2.7797</td>
<td>0.07</td>
<td>0</td>
<td>0.0007</td>
</tr>
<tr>
<td>2</td>
<td>$\log_{10}(\mu)$</td>
<td></td>
<td>$-0.6879$</td>
<td>$-0.6879$</td>
<td>$-0.6872$</td>
<td>0.07</td>
<td>0</td>
<td>0.0007</td>
</tr>
<tr>
<td>3</td>
<td>$\log_{10}(\beta)$</td>
<td></td>
<td>$-5.2600$</td>
<td>$-5.2600$</td>
<td>$-5.2602$</td>
<td>0.09</td>
<td>0.0002</td>
<td>0.0002</td>
</tr>
<tr>
<td>4</td>
<td>$\log_{10}(\gamma)$</td>
<td></td>
<td>0.2300</td>
<td>0.1652</td>
<td>0.1605</td>
<td>0.08</td>
<td>0.0648</td>
<td>0.0695</td>
</tr>
<tr>
<td>5</td>
<td>$\log_{10}(\sigma)$</td>
<td></td>
<td>$3.4500$</td>
<td>$3.4500$</td>
<td>$3.4496$</td>
<td>0.1200</td>
<td>0</td>
<td>0.0004</td>
</tr>
<tr>
<td></td>
<td>$\log_{10}(\delta)$</td>
<td></td>
<td>$-0.1400$</td>
<td>$-0.1814$</td>
<td>$-0.1890$</td>
<td>0.18</td>
<td>0.0414</td>
<td>0.0490</td>
</tr>
<tr>
<td></td>
<td>$\log_{10}(\epsilon)$</td>
<td></td>
<td>$-5.67$</td>
<td>$-1.7725$</td>
<td>$-1.7758$</td>
<td>3.8942</td>
<td>3.9975</td>
<td>0.30</td>
</tr>
</tbody>
</table>

**Comparison of the mean and maximum likelihood of the prior and posterior distributions for 5 patients.**

**Table II**