SMGroup

The Iterative Process of Quantitative Modeling of Infection Dynamics in Renal Transplant Recipients

Banks HT[1]*, Everett RA[1] and Neha Murad[1]

[1]Department of Mathematics, Center for Research in Scientific Computation, North Carolina State University, USA

***Corresponding author:** Banks HT, Department of Mathematics, North Carolina State University, Raleigh, NC, USA, Tel: 27695-8212 919-515-5289, Fax: 919-515-1636, Email: htbanks@ncsu.edu

Published Date: December 29, 2016

ABSTRACT

Mathematical models play a significant role in providing a numerical and analytical perspective to biological models. When formulating these mathematical models to improve our understanding of biological processes, it is not always possible to find all parameter values in literature. In such cases, and in the presence of data, inverse problems are performed to estimate these unknown parameters. Statistical error models used during inverse problem formulations help quantify the uncertainty and variability that arises with using experimental data. This process of applying mathematical and statistical techniques for modeling physical processes is an iterative one and often leads to new insights with every new iteration. There is a relatively recent research effort in modeling the mechanisms of solid organ transplants, specifically kidney transplants. We present mathematical and statistical models to illustrate the iterative process of modeling for renal transplant recipients infected by BK virus. Using a second order difference-based method to eliminate statistical error model misspecification, we show how modified residuals from the inverse problem can be used to detect discrepancies in mathematical model formulation.

Keywords: Quantitative models; Inverse problem; Infection dynamics; Organ survival; Transplant patients; Immunosuppression; Iterative modeling process, Model misspecification, BKV, PVAN, HCMV

Abbreviations: BKV: BK Virus; CKD: Chronic Kidney Disease; ESRD: End Stage Renal Disease; GvHD: Graft-Versus-Host Disease; HCMV: Human Cytomegalovirus; HIV: Human Immunodeficiency Virus; PVAN: Polyomavirus type BK-associated Nephropathy; SOT: Solid Organ Transplants

INTRODUCTION

Mathematical and statistical models are important and beneficialtools that contribute to the understanding of biological phenomena. Models mathematically represent biological processesand allow for a systematic approach to investigate complex systems. They provide an opportunity to test hypotheses prior to biological experiments through computational simulations (a faster and cheaper method), which can in turn, aid in the design of experiments. Additionally, they require adegree of preciseness through mathematical expression that is often neglected [1].

In order to effectively use these tools, a collaborative effort and common language between mathematicians and biologists are necessary, as demonstrated by the following iterative modeling process [1] (Figure 1). The biologist first presents a research question about some biological process as well as knowledge about biological relationships and mechanisms, often depicted by a schematic or diagram. The mathematician then represents the biological hypothesis of the relationships using mathematical expressions, such as differential equations. Analytic or numerical analysis of the model produces results which are interpreted and compared to the biological system, possibly leading to a change in the understanding of the biological relationships. The process is continuously repeated, sometimes through multiple research efforts, using the resulting biological insights.

Models depend on parameters, such as initial conditions (the initial quantity for the variables) and rate parameters (e.g., growth rate), that are not always available in literature. When all the parameters are known (or assumed), the model is simulated and the resulting model solution is called a forward solution. When parameters are unknown, data can be used to estimate these free parameters, enhancing the utility of modeling. An inverse problem aims to estimate parameters such that the model solution emulates the data. In order to do this, the mathematician and biologist must quantify the uncertainty and variability in the given dataset through a statistical error model. Thus the analysis step in the iterative process can depend on both the mathematical and statistical error models. A new understanding of the system can then contribute to the redesign of the data collection process if necessary.

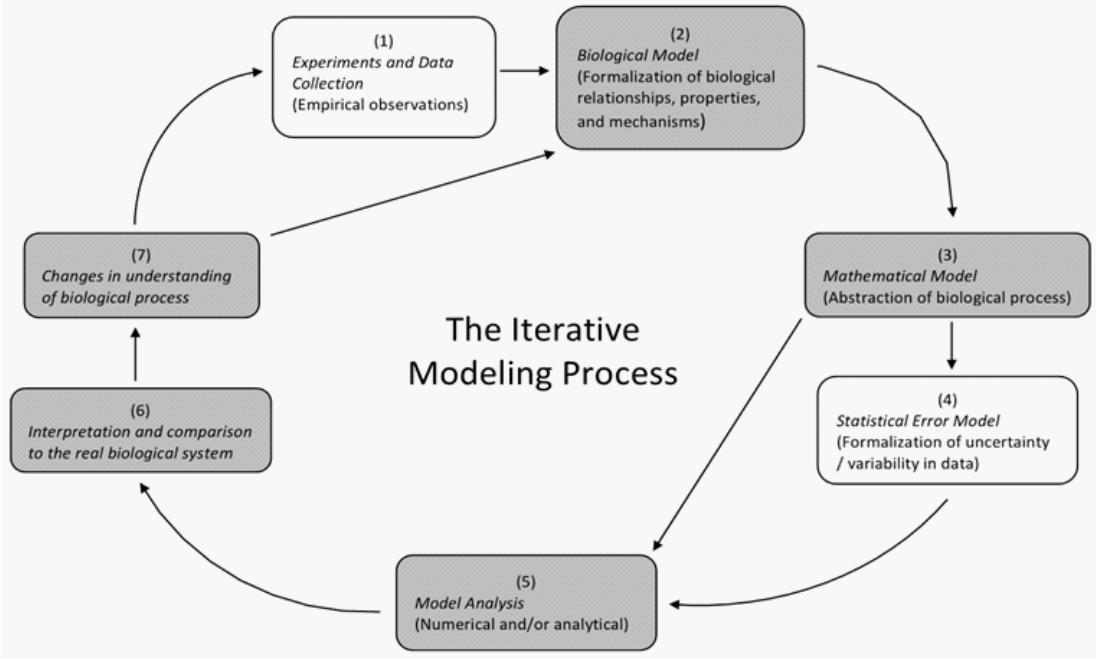

Figure 1: Schematic of the iterative modeling process [1]. White boxes indicate steps when data is available.

We emphasize that the inverse problem depends on both the mathematical model and the statistical error model. Often the uncertainty in the data is ignored and an incorrect statistical error model is assumed, which can lead to incorrect results. The statistical error model should accurately represent the uncertainty in the data [1-4]. The mathematical model, while being an approximation, should have biologically meaningful parameters and be theoretically consistent with the real biological system. Once the inverse problem is performed, the estimated parameter values should be biologically relevant. It is important to also provide associated confidence levels for the estimated values; a parameter value of 1 with a confidence level of ±1000conveys very little, whereas a parameter value of 1 with a confidence level of ±.001 is informative [1,4]. The model solution from the inverse problem should fit the data adequately, and can also be simulated for a longer period of time to produce future predictions. In addition, results from the inverse problem might point to shortcomings in the data collection process. Ultimately, good mathematical models and statistical error models should contribute to the understanding of the biological phenomena, despite the limitations. This idea is captured in the often quoted truth, "all models are incorrect, but some are more useful than others" [4,5].

While the mathematical model represents unifying biological phenomena in population cohorts, parameters can vary within biological ranges for individuals. An inverse problem can be performed with data from an individual, resulting in estimated parameters and a model solution for that individual. These results can elucidate the biological process occurring in an individual

as well as inform treatment option decisions for that person. Often treatment guidelines are determined from clinical trials and applied to patients using a standardized approach. Personalized medicine attempts to find subgroups of similar patients who may need a different course of action or adaptively treat an individual patient based on their response to treatment [6].

One example, among many, of using mathematical models to enhance personalized treatment is the application of control theory to determine an optimal treatment strategy. The book [7] contains several biomedical applications of control theory, such as determining the optimal treatment regimen for cancer patients undergoing chemotherapy in order to minimize tumor density as well as treatment side effects. The authors also apply control theory to human immunodeficiency virus **(HIV)**-infected patients undergoing chemotherapy as well as to determine the optimal insulin injection level to better regulate blood glucose levels in diabetic patients. Further applications of control theory to HIV therapy strategies are presented in [8-10]. The authors in [11] present control problems related to cancer therapies, such as antiangiogenic treatments.

Another natural control problem involves designing an individualized adaptive immunosuppressant treatment schedule for patients undergoing voluntary immunosuppression. A balance between under-suppression and over-suppression of the immune system needs to be struck in order to inhibit the immune response while still preventing infection. The predictive power of models is ideal for determining the optimal time-varying amount of immunosuppression to retain this balance. As data is collected, the mathematical model can be updated and the treatment adjusted accordingly. Several diseases require voluntary immunosuppressant treatment. Allogeneic haematopoietic stem-cell transplant recipients receive immunosuppression therapy to diminish the risk of developing graft-versus-host disease **(GvHD)**, the attack of donor T cells on the host tissues [12]. Immunosuppression therapy is also one of the current treatment options for autoimmune diseases, which result from the body's immune system attacking its own body [13]. Organ transplant recipients require life-long immunosuppression to prevent the body from rejecting the transplant [14].

Here, we illustrate ideas and methodologies for using mathematical and statistical models to improve biological understanding in the context of kidney transplantation.

RECENT MODELING EFFORTS IN KIDNEY TRANSPLANTATION

Several research groups have recently contributed to developing mathematical models to study the mechanisms of the complex biological process involving solid organ transplants **(SOT)**, specifically kidney transplants.

PVAN Progression

Funk et al. [15] use simple mathematical equations to investigate the relationship between BKV replication and polyomavirus type BK-associated nephropathy **(PVAN)**. The authors assume the BK virus grows and decays exponentially and then present the corresponding equations to

calculate the viral doubling times and half-lives. The generation time and basic reproductive ratio **(R_0)** equations are also given. The authors then perform a retrospective mathematical analysis on15individual patient datasets with the given equations. Their results indicate rapid replication of the BK virus, elucidating the progressive nature of PVAN, contrary to the general perception in clinical practice. The authors also propose the use of R_0 as a measure of the efficiency of anti-viral interventions in future studies.

The authors of [16] extend the work in [15] and first perform statistical analysis on datasets from 223 kidney transplant patients to help understand the relationship between BKV in the plasma and urine. The authors present a differential equations model that considers four cell populations and two virus populations: uninfected kidney tubular epithelial cells, infected tubular epithelial cells, plasma virus, uninfected urothelial cells, infected urothelial cells, and urine virus. Antiviral interventions are represented as a time-dependent function that affects the growth of both viral populations. The basic reproductive ratio (R_0) for the kidney and urinary compartments are determined and used in parameter sensitivity analysis. The authors simulate the mathematical model to explore the dynamics of BKV for various scenarios and compare simulation results to 7 individual kidney transplant patient datasets. The scenario that best matches the clinical data assumes viral replication starts in the kidney, infects urothelial cells, and then a bidirectional viral flux increases replication in both compartments. The results provide insights into the relationship between the two compartments with respect to BKV replication and suggest further awareness into PVAN progression.

HCMV Infection

Kepler et al. [17] continue the iterative modeling process by incorporating the effect of the immune response on the viral infection in the presence of immunosuppression therapy. The authors specifically consider human cytomegalovirus **(HCMV)** infection in SOT recipients and present a model that describes the dynamics of the viral load, immune response, actively-infected cells, susceptible cells, and latently-infected cells. The authors show that the model can describe the three types of infection: primary, latent, and reactivated. Model simulations are given for varying amounts of immunosuppression. Due to the limited amount of corresponding data in literature, the authors do not compare their model to patient data. However, the authors state how inverse problems can be performed with individual datasets to estimate parameter values for individual patients, providing insights into the heterogeneity in disease progression.

Banks et al. [18] modify and extend the model in [17] to include the body's immune response to a donor kidney. Their model considers susceptible cells, infected cells, free HCMV, HCMV-specific CD8+T cells, allospecific CD8+T cells that target the donor kidney, and serum creatinine. Due to the lack of data, model validation is not provided; model simulations with varying antiviral and immunosuppressive drug efficacy are produced only to verify that results match clinical trends. The authors then use simulated data todemonstrate an optimal control problem to design an

adaptive anti-viral and immunosuppressant treatment schedule that balances over-suppression and under-suppression of the immune system.

BKV Infection

The authors in [14] adapt the model from [18] to consider renal transplant recipients infected with BKV. We now present this model to illustrate the iterative modeling process, motivated partially by the availability of data.

Steps 1 and 2: Data Collection and Biological Model

Recall that the first step in the iterative modeling process, is collection of data when available, followed by formulation of a biological model. Data was collected at Massachusetts General Hospital from a renal transplant patient TOS003 diagnosed with BKV in the first 3 months of kidney transplantation. Eight BK viral plasma load (DNA copies/mL) measurements and sixteen plasma creatinine level (mg/dL) measurements were collected. Figure 2 contains the plots of the dataset and Figure 3 contains the biological model depicted as a schematic.

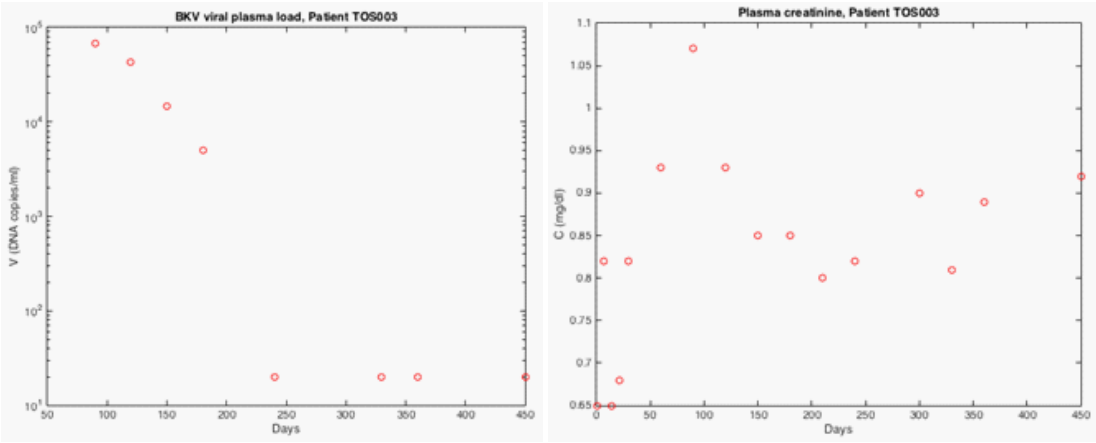

Figure 2: Patient TOS003 BKV viral plasma load and plasma creatinine levels [14].

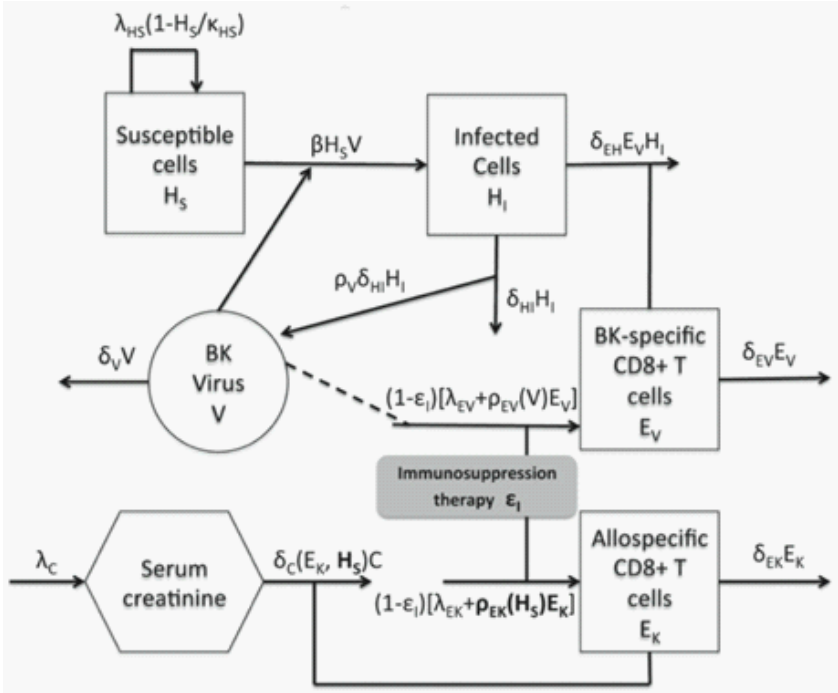

Figure 3: Model schematic [14].

Since no anti-viral treatment exists for BKV, the authors in [14] modify the model in [18] to only consider the efficiency of immunosuppressant treatment. Additionally, the authors model the effect of the susceptible cells on creatinine clearance andon the allospecific CD8+T cell population growth. As opposed to Funk et al. [14,22] does not consider the urothelial cells and only considers infection in tubular epithelial cells, in part due to the availability of data.

Step 3: Mathematical Model

Once the most salient biological relationships and mechanisms are formalized in the biological model, a mathematical model is built to represent this biological hypothesis. The authors in [14] consider six state variables to represent the different cell, virus, and creatinine populations given in the schematic (i.e., the squares, circle, and hexagon). The model consists of an ordinary differential equation for each state variable, which represents the rate of change of the given population, where

$$\begin{pmatrix}\text{rate of change}\\ \text{of population}\end{pmatrix} = \begin{pmatrix}\text{rate of}\\ \text{population gain}\end{pmatrix} - \begin{pmatrix}\text{rate of}\\ \text{population loss}\end{pmatrix}.$$

Several factors can contribute to the population gain or loss, such as proliferation, conversion from one cell type to another, and clearance from the body. Note that an arrow into a shape in the schematic indicates a gain in the population whereas an arrow leaving a shape indicates a loss in the population.

The mathematical model in [14] describe the concentrations of susceptible cells (H_S), infected cells (H_I), free BKV (V), BKV-specific CD8+T cells (E_V), allospecific CD8+T cells that target the kidney (E_K), and serum creatinine (C), by the following

$$\begin{aligned}
\dot{H}_S &= \underbrace{\lambda_{HS}\left(1-\frac{H_S}{\kappa_{HS}}\right)H_S}_{1} - \underbrace{\beta H_S V}_{2} \\
\dot{H}_I &= \underbrace{\beta H_S V}_{2} - \underbrace{\delta_{HI} H_I}_{3} - \underbrace{\delta_{EH} E_V H_I}_{4} \\
\dot{V} &= \underbrace{\rho_V \delta_{HI} H_I}_{3} - \underbrace{\delta_V V}_{5} - \underbrace{\beta H_S V}_{2} \\
\dot{E}_V &= \underbrace{(1-\epsilon_I)}_{10}\underbrace{[\lambda_{EV} + \rho_{EV}(V)E_V]}_{6} - \underbrace{\delta_{EV} E_V}_{7} \\
\dot{E}_K &= \underbrace{(1-\epsilon_I)}_{10}\underbrace{[\lambda_{EK} + \rho_{EK}(H_S)E_K]}_{8} - \underbrace{\delta_{EK} E_K}_{9} \\
\dot{C} &= \underbrace{\lambda_C}_{11} - \underbrace{\delta_C(E_K, H_S)C}_{12},
\end{aligned} \tag{1}$$

where

$$\rho_{EV}(V) = \frac{\bar{\rho}_{EV} V}{V + \kappa_V},$$

$$\rho_{EK}(H_S) = \frac{\bar{\rho}_{EK} H_S}{H_S + \kappa_{KH}},$$

$$\delta_C(E_K, H_S) = \frac{\delta_{C0}\kappa_{EK}}{E_K + \kappa_{EK}} \cdot \frac{H_S}{H_S + \kappa_{CH}},$$

and initial conditions (2)

$$(H_S(0), H_I(0), V(0), E_V(0), E_K(0), C(0)) = (H_{S0}, H_{I0}, V_0, E_{V0}, E_{K0}, C_0).$$

A description of the state variables and parameters are given in Tables 1 and 2 respectively. The individualterms in the mathematical model are explained by the following..

1. Susceptible cells proliferate logistically at maximum rate λ_{HS} with the carrying capacity κ_{HS}. (See [3] or [19] for information about modeling logistic growth).

2. The term $\beta H_S V$ represents the loss of susceptible cells due to infection, causing growth in the infected cell population and loss of free virions. For simplicity, we assume one virion infects one susceptible cell, creating one infected cell. The parameter β represents the infection rate of H_S by V.

3. Infected cells lyse at rate δ_{HI} due to the cytopathic effect of the BK virus, releasing ρ_V free virions.

4. The infected cell population can also decrease due to elimination by the BK-specific CD8+T cells at rate δ^{EH}.

5. The body naturally clears virionsfrom the blood at rate δ_V.

6. We assume both a virus-dependent and virus-independent growth rate of the BKV-specific CD8+T cell population. The parameter λ_{EV} represents the virus-independent source rate for E_V. The virus-dependent growth rate is represented by the term $\rho_{EV}(V)E_V$; The Michaelis-Menten type function $\rho_{EV}(V)$ is abounded monotonic increasing function, where ρ_{EV} represents the maximum proliferation rate and κ_V represents the half saturation constant. (See [20] or [19] for information on Michaelis-Menten kinetics). 7. The parameter δ_{EV} represents the death rate of the BKV-specific CD8+T cells.

8. We assume both a susceptible cell-dependent and susceptible cell-independent growth rate of the allospecific CD8+T cell population that targets the kidney. The parameter λ_{EK} represents the susceptible cell-independent source rate for E_K. The susceptible cell-dependent growth rate is represented by the term $\rho_{EK}(H_S)E_K$; the Michaelis-Menten type function $\rho_{EK}(H_S)$ is a bounded monotonic increasing function, where ρ^{-}_{EK} represents the maximum proliferation rate and κ_{KH} represents the half saturation constant.

9. The parameter δ_{EK} represents the death rate of the allospecific CD8+T cells that target the kidney.

10. The growth of the immune system inversely depends on the immunosuppressive treatment. This dependence is given by the term $(1-\epsilon_I)$, where$\epsilon_I \in [0,1]$ represents the efficiency of immunosuppressive drugs. We assume a drug efficiency of 0% ($\epsilon_I=0$) indicates that the treatment does not affect the immune system and the CD8+T cells grow normally; adrug efficiency of 100% ($\epsilon_I=1$) causes the immune system to decrease exponentially. The dosage, type, and concentration of drugs often change over the course of treatment. Therefore, for simplicity, we define ϵ_I by the following piecewise constant function

$$\epsilon_f(t) = \begin{cases} \epsilon_1 & t \in [0, 21] \\ \epsilon_2 & t \in (21, 60] \\ \epsilon_3 & t \in (60, 120] \\ \epsilon_4 & t \in (120, 450]. \end{cases}$$

11. The parameter λ_C represents the constant production rate of serum creatinine.

12. A damaged kidney is not able to filter waste from the blood as effectively, causing a build-up of creatinine. Thus, we assume the loss of serum creatinine depends on both the susceptible cells and the allospecific CD8+T cells that target the kidney (see equation (2)), with δ_{C0}representing the maximum clearance rate of serum creatinine. As the allospecific CD8+T cells that target the kidney increase (indicating damage to the kidney), the creatinine clearance rate δ_C decreases to 0, resulting in creatinine build-up. As the susceptible cell population increases (indicating a healthier kidney), the creatinine clearance rate δ_C increases, which in turn decreases the creatinine concentration in the body. The parameters κ_{EK} and κ_{CH} represent half saturation constants.

We note that while the mathematical model considers six variables, only two variables, virus load and creatinine, are measured. We refer to these two variables as observables.

Table 1: Description of state variables.

State	Description	Unit
H_s	Concentration of susceptible host cells	cells/mL
H_I	Concentration of infected host cells	cells/mL
V	Concentration of free BKV	copies/mL
E_V	Concentration of BKV-specific CD8+ T cells	cells/mL
E_K	Concentration of allospecific CD8+ T cells that target kidney	cells/mL
C	Concentration of serum creatinine	mg/dL

Table 2: Model Parameter Descriptions and Fixed Values (est. indicates a parameter to be estimated).

Parameter	Description	Unit	Value
λ_{HS}	Proliferation rate for H_s	1/day	0.030
κ_V	Saturation constant	copies/mL	180.676
κ_{HS}	Saturation constant	cells/mL	1025.888
λ_{EK}	Source rate of E_K	cells/(mL day)	0.002
β	Infection rate of H_s by V	mL/(copies day)	est.
δ_{EK}	Death rate of E_K	1/day	est.
δ_{HI}	Death rate of H_I by V	1/day	0.085
λ_C	Production rate for C	mg/(dL day)	0.007
ρ_V	# Virions produced by H_I before death	copies/cells	4292.398
δ_{C0}	Maximum clearance rate for C	1/day	0.014
δ_{EH}	Elimination rate of H_I by E_V	mL/(cells day)	0.002
κ_{EK}	Saturation constant	cells/mL	0.200
δ_V	Natural clearance rate of V	1/day	0.372
κ_{CH}	Saturation constant	cells/mL	10.000
λ_{EV}	Source rate of E_V	cells/(mL day)	0.001
$\bar{\rho}_{EK}$	Maximum proliferation rate for E_K	1/day	est.
δ_{EV}	Death rate of E_V	1/day	est.
κ_{KH}	Saturation constant	cells/mL	84.996
$\bar{\rho}_{EV}$	Maximum proliferation rate for E_V	1/day	est.
ϵ_I	Efficacy of immunosuppressive drugs		

Step 4: Statistical Error Model

We now present different statistical error models and techniques to determine the correct model that accurately reflects the uncertainty in the data.

The authors in [14] assume the simplest statistical error model where the variances of the error for each observable (viral load and creatinine) are equal and that the variances are constant over time. That is, the authors account for the uncertainty in the dataset by the following statistical error model

$$\begin{aligned} Y_i^1 &= \underbrace{f_1(t_i^1;\theta_0)}_{1} + \underbrace{\mathcal{E}_i^1}_{2}, \qquad i = 1,2,\ldots,n_1, \\ Y_j^2 &= \underbrace{f_2(t_j^2;\theta_0)}_{1} + \underbrace{\mathcal{E}_j^2}_{2}, \qquad j = 1,2,\ldots,n_2. \end{aligned} \tag{3}$$

Term 1 represents the model solution for the observables assuming a "true" or nominal parameter set θ_0.The existence of this "true" parameter set is a standard assumption in statistical

models [10]. The functions f_1 and f_2 represent the model solution for viral load and creatinine at times t_i^1 and t_j^2, respectively.

Term 2 represents the measurement error that causes the measurements to differ from the model solution with the "true" parameter set. We assume the $n_1 \times 1$ and $n_1 \times 2$ random vectors ε^1 and ε^2 respectively, are independent and identically distributed with mean zero and $Var(\varepsilon_i^1)=\sigma_{01}^2$ and $Var(\varepsilon_j^2)=\sigma_{02}^2$. The authors in [14] assume $\sigma_{01}^2=\sigma_{02}^2$.

In [3], we build on the works of [14] and consider what we believe to be a more realistic statistical error model; we assume the variances in observation errors are not equal and allow for the errors to depend on the size of the observed quantity. That is, we assume the following statistical error model,

$$\begin{aligned} Y_i^1 &= \underbrace{f_1(t_i^1;\boldsymbol{\theta}_0)}_{1} + \underbrace{f_1(t_i^1;\boldsymbol{\theta}_0)^{\gamma_1}\mathcal{E}_i^1}_{2}, \qquad i=1,2,\ldots,n_1, \\ Y_j^2 &= \underbrace{f_2(t_j^2;\boldsymbol{\theta}_0)}_{1} + \underbrace{f_2(t_j^2;\boldsymbol{\theta}_0)^{\gamma_2}\mathcal{E}_j^2}_{2}, \qquad j=1,2,\ldots,n_2, \end{aligned} \tag{4}$$

for $\gamma_k \geq 0$, k=1,2. Term1 remains the same as before, but term 2 now depends on the size of the model solution of the observables; as the size of the solution increases, it makes sense for the measurement error to also increase. Note that if both $\gamma_1=0$ and $\gamma_2=0$, the two statistical error models are equivalent.

To determine the correct γ_k value, we follow [2] and apply a second order difference-based method directly to the data to calculate the modified pseudo errors for some chosen $\gamma_k \geq 0$ value, k=1,2. The plots of the modified pseudo errors vs. time should produce a random scatter plot around the horizontal-axis. If an undesired non-random shape is present (e.g., a megaphone shape), then the process is repeated with a different γ_k value until the desired scatter plot is produced. (See [2] and [3] for further details).

Another method that is often used to determine the correct statistical error model requires performing an inverse problem with a chosen $\gamma_k \geq 0$ value and then calculating the modified residuals

$$\mathcal{M}_{kl} \equiv \frac{y_l^k - f_k(t_l^k;\hat{\theta})}{f_k(t_l^k;\hat{\theta})^{\gamma_k}}, \tag{5}$$

where y_l^k represents the data point for observable k at time t_l^k, l=1,... ,n_k and $\hat{\theta}$ represents the estimated parameter set determined from the inverse problem. Similar to modified pseudo errors, the plot of modified residuals vs. time and modified residual vs model [4,1] should be a randomly scattered band around the horizontal axis. If a non-randomshape is present, the process is repeated with a different γ_k value.

Note that modified pseudo errors are not equivalent to modified residuals. Modified pseudo errors are calculated directly from the data to account for the uncertainty in the data and are independent of the mathematical model. Modified residuals depend on both the data and the model solution (see equation (5) above) and represent the normalized difference between the data and the model. Although the two methods are similar, the method using modified residuals can require performing multiple inverse problems, which is computationally more expensive, and more notably, tacitly assumes a correct mathematical model.

Step 5: Model analysis

The authors of [14] numerically analyze the model by performing an inverse problem with the data to estimate parameters. That is, the authors seek to find a parameter set that minimizes the distance between the collected data and mathematical model. However, the model has a large number of parameters (29 parameters) and thus not all the parameters can be reliably estimated with such a small dataset. Although there is no set number of data points needed per estimated parameter, a general rule of thumb is about 25 to 30 data points per parameter. However, we emphasize that this rule of thumb is highly model-dependent and certainly depends on the problem. Clearly, estimating 29 parameters confidently with 24 data points is not possible. Thus, the authors of [14] implement an iterative procedure to determine the most sensitive parameters. An inverse problem is then performed to estimate the 5 most sensitive parameters and the resulting model solutions fit the data well.

In [3], we numerically analyze the model by performing an inverse problem assuming the more accurate statistical model. Following [14], we estimate only the 5 most sensitive parameters and fix the remaining parameters to biologically relevant values (shown in Table 2) when possible. The model solutions fit the data well and are given in Figure 4. The remaining state variables that do not have corresponding data are plotted in Figure 5. We note that although the model fit is similar to that in [14], we have more confidence that our estimated parameter values and model solution are accurate since a more realistic statistical error model is considered.

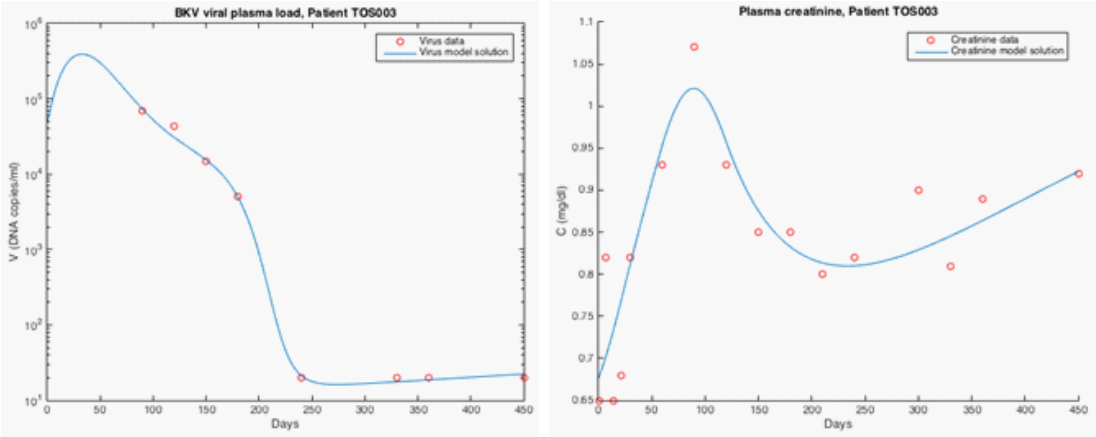

Figure 4: BK virus (V) (left) and Creatinine (C) (right) model solutions from inverse problem for $\gamma = [0.5,0]$. Estimated parameters $[\beta,\bar{\rho}_{EV},\delta_{EV},\delta_{EK},\bar{\rho}_{EK}]=[8.690 \times 10^{-8}, 0.233, 0.098, 0.120, 0.198]$.

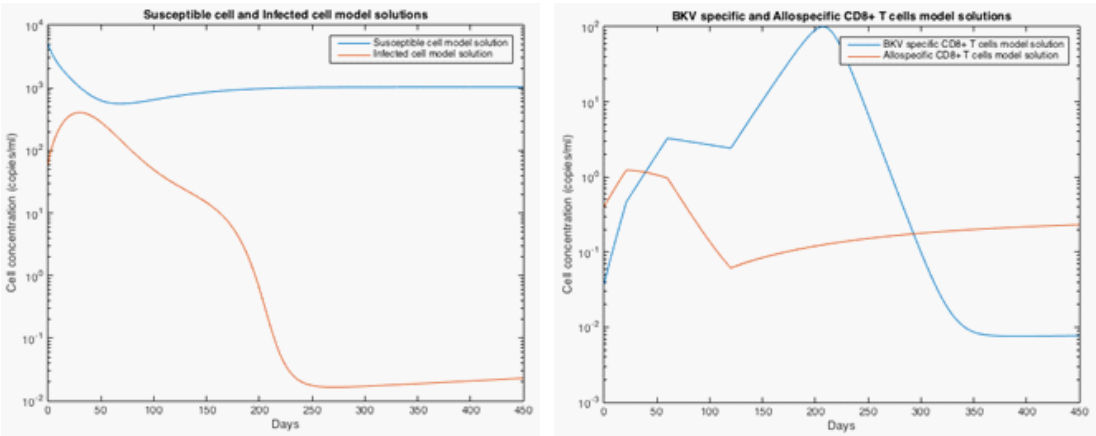

Figure 5: Model solutions (left) for susceptible (blue) and infected cells (red) and model solutions (right) for BKV specific immune cells (blue) and allospecific immune cells (red).

Steps 6 and 7: Comparison to the real biological system and changes in understanding

To further verify that the model is accurately describing the biological process, we follow [3] and calculate the modified residuals in equation (5) in order to detect mathematical model misspecification. By applying the difference-based method directly to the data, we already account for measurement error. Thus, mathematical model discrepancy remains as the primary possible source for error, indicated by a non-random shape in the modified residual plots. Figure 6 contains the plots of the modified residuals *vs* time for viral load and creatinine, which appear to form a random band around the horizontal axis. This suggests that our mathematical model correctly accounts for the biological dynamics observed in the data.

However, in this case it is difficult to determine the accuracy of the mathematical model due to the limited amount of data. An advantage of mathematical modeling is the ability to detect these inadequacies in data collection and then contribute to the redesign of the collection process [21-25].

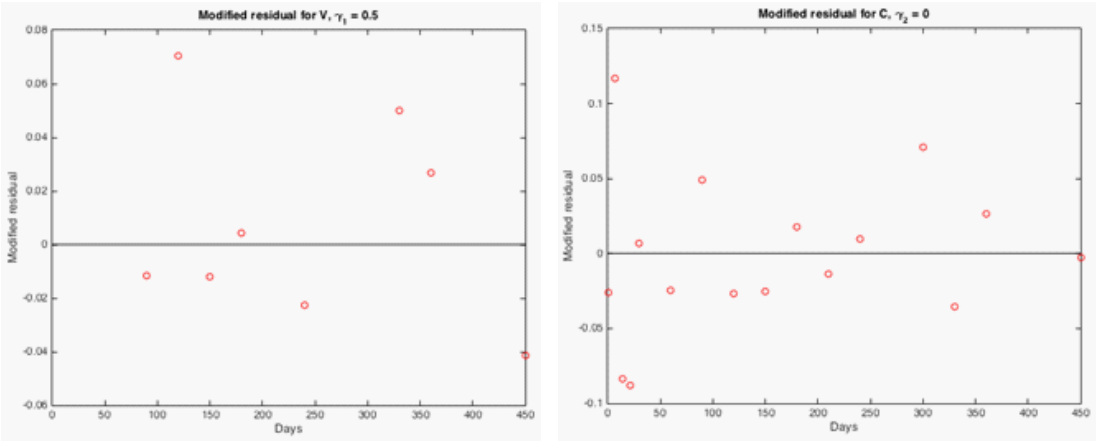

Figure 6: BK virus (left) and Creatinine (right) modified residuals vs. time for γ=[0.5,0].

Although we do not have data to compare to the solutions of the remaining four variables, the dynamics seem biologically realistic (Figure 5). The infected cells behave similar to the free BK virus cell population while the susceptible cell population shows a corresponding decrease and then eventually reaches carrying capacity. The dynamics of the immune cells follow from the dynamics of the viral load and susceptible cells. The allospecific CD8+ T cells rise and then decay to trace amounts as susceptible cells reach carrying capacity. Meanwhile the BK specific CD8+ T cells mirror the dynamics of the BK virus, growing in population as more virions are produced and decaying steadily to trace amounts as BKV concentrations decrease below a minimal threshold.

Simulated Data

Since the results above are difficult to conclude with conviction due to sparseness in the data, we now demonstrate detecting mathematical model discrepancy using simulated data [3]. This is a procedure often adopted by modelers when data is scarce. We create simulated data by simulating model (1) with the given parameters in Table 2 and [β, ρ^{-}_{EV}, δ_{EV}, δ_{EK}, ρ^{-}_{EK}]=[8.570×10^{-8}, 0.251, 0.109, 0.101, 0.164] to produce the forward solution and then adding random noise to the solution at each time point; this is given by

$$V_i = f_1(t_i^1;\theta_0) + f_1(t_i^1;\theta_0)^{\gamma_1}\epsilon_i^1,$$
$$C_j = f_2(t_j^2;\theta_0) + f_2(t_j^2;\theta_0)^{\gamma_2}\epsilon_j^2, \quad (6)$$

where $\varepsilon^1 \sim N(0,0.3)$, $\varepsilon^2 \sim N(0,0.03)$, and $\gamma=[0.5,0]$. Figure 7 contains the plots of the simulated data. The advantage of simulated data is that we know the "true" biological process that is represented by the data (i.e., the mathematical model) and the "true" measurement noise (i.e., the statistical model).

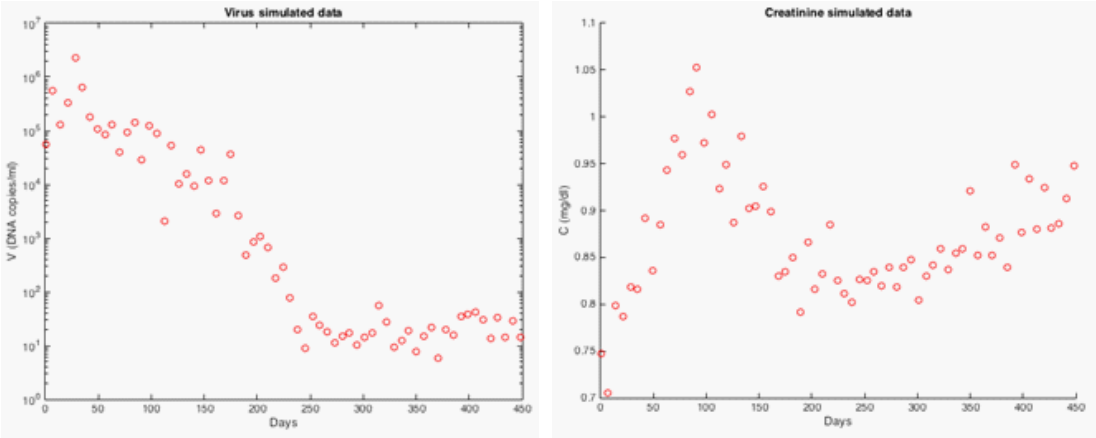

Figure 7: Simulated data for BK virus (left) and Creatinine (right).

To demonstrate how mathematical model misspecification can be detected, we perform an inverse problem with a different mathematical model. We simplify the model (1) by removing the nonlinearity from the susceptible cell population growth dynamics. That is, we use model (1) but with the following susceptible cell population differential equation

$$\dot{H}_S = \lambda_{HS} - \delta_{HS} H_S - \beta H_S V.$$

Figures 8 and 9 contain the model solution and corresponding modified residuals respectively. Although the model solution produces a reasonable fit to the data, the modified residuals display a strong non-random pattern. This indicates mathematical model error and demonstrates that the simpler model is unable to capture the dynamics in the data. If these results were derived using clinical data, another iteration of modeling would be deemed necessary.

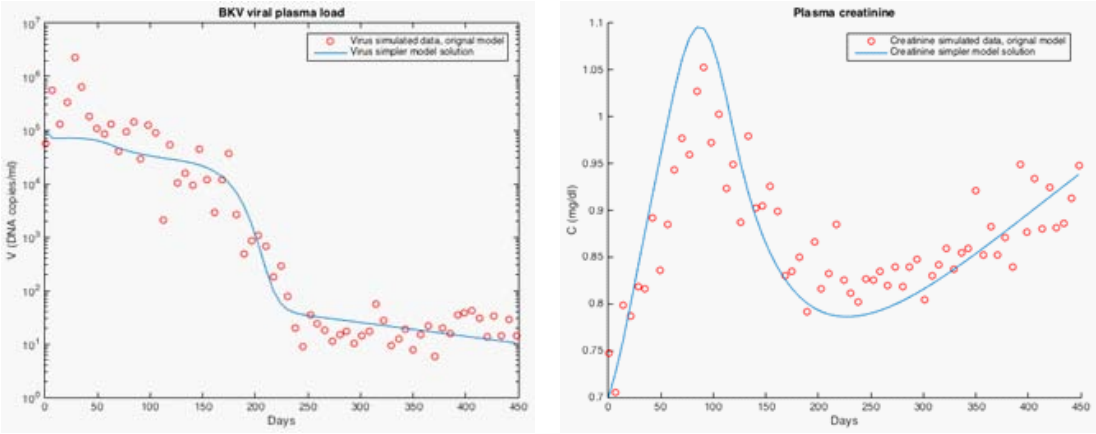

Figure 8: Simpler model solutions and simulated data created from the original model (1) for γ=[0.5,0]. Estimated parameters [β, ρ^{-}_{EV}, δ_{EV}, δ_{EK}, δ_{HS}, ρ^{-}_{EK}] = [1.791 $\times 10^{-8}$, 0.399, 0.195, 0.204, 2.511 $\times 10^{-5}$, 0.322].

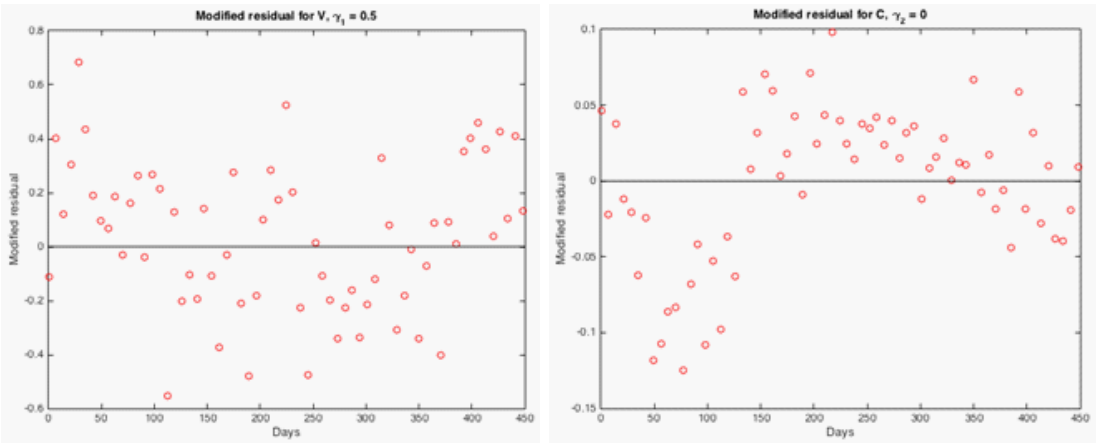

Figure 9: BK virus (left) and Creatinine (right) modified residuals vs. time for γ=[0.5,0].

In [3], we also investigate mathematical model misspecification when a more complex model is assumed than that represented by the data. In a similar finding, the model solution from the inverse problem also fits the data well, but the modified residuals display a strong non-random pattern. Thus, this method is able to detect mathematical model discrepancy when either a simpler or more complex model is assumed compared to the data dynamics.

DISCUSSION

We discuss the important roles of mathematical and statistical models which contribute in solidifying our understanding of biological processes and quantifying the uncertainty that arises from data. We emphasize the equal importance of both these modeling processes in the context of inverse problems and parameter estimation. After summarizing the recent research efforts in

modeling organ transplants, particularly kidneys, with respect to PVAN progression and HCMV infection, we then delve into explaining and illustrating the iterative process of modeling through our current efforts in understanding the dynamics of BKV infection in renal transplant patients.

Using a differential equations model to mathematically describe the biological relationship of BKV infection and the donor kidney in the presence of an immunosuppressant regime, we numerically analyze the results which demonstrate that the model fits the data well and the modified residuals appear random. This indicates that together with an accurate statistical model, the mathematical model captures the dynamics of the real system observed in the data. Next we show that if one chooses a correct statistical model using difference-based techniques and eliminates the possibility of statistical model misspecification, then modified residuals can be used to capture discrepancies, if any, between the biological process observed and the mathematical model. However, due to the limited amount of data, the results are difficult to conclude with a high level of confidence. Therefore, we demonstrate this method of detecting mathematical model mismatch using a dense simulated dataset.

The authors in [14] partially validate the BKV model by fitting the model solution from the inverse problem to a sparse clinical dataset from one individual and producing reasonable results. However, with a denser dataset and longitudinal data from more individual patients, one would be able to validate the mathematical model with more confidence.

In the future we plan to develop an adaptive optimal control problem to determine the optimal level of immunosuppressant treatment for individual patients to balance over-suppression and under-suppression. The authors in [18] present this control problem in the context of HCMV-infected kidney transplant recipients. However, they consider a model that is not validated with patient data and present results for hypothetical patients. Unlike HCMV infection, the lack of an approved antiviral therapy for symptomatic BKV nephropathy makes the task of carefully monitoring the immunosuppressant dosage an even more imperative one.

Candidate-gene and genome-wide association studies have assessed the genetic connection for common renal diseases, including Chronic Kidney Disease **(CKD)**. These studies were conducted on individuals with a diverse array of ethnic and racial backgrounds, which makes a good case for studying genetic polymorphism with respect to CKD or ESRD **(End Stage Renal Disease)** [26,27]. Even though certain diseases, such as breast cancer, have diagnosis and determination of risk factors based on genotyping [28], there is no such definitive DNA-based diagnostics and risk prediction available yet for patients suffering from CKD. When such genetic markers do become easily determinable, a natural next step after genetic composition-based diagnosis for renal transplant recipients (renal pharmacogenomics) would be the regulation and control of these identified biomarkers. The application of control theory to the process of gene regulation as seen in [29] could play a crucial role in shaping the future of renal pharmacogenomics.

AKNOWLEDGEMENTS

This research was supported in part by the National Institute on Alcohol Abuse and Alcoholism under subcontract 500693NCSU from the North well Health, Manhasset, NY and in part by the Air Force Office of Scientific Research under grant number AFOSR FA9550-15-1-0298.

References

1. HT Banks, HT Tran. Mathematical and Experimental Modeling of Physical and Biological Processes, Taylor/Francis-Chapman/Hall-CRC Press, Boca Raton, FL. 2009.
2. HT Banks, J Catenacci, S Hu. Use of difference-based methods to explore statistical and mathematical model discrepancy in inverse problems, Journal of Inverse and Ill-Posed Problems. 2016; 24: 413-433.
3. HT Banks, RA Everett, S Hu, N Murad, HT Tran. Mathematical and statistical model misspecifications in modeling immune response in renal transplant recipients, CRSC-TR16-14, Center for Research in Scientific Computation, North Carolina State University, Raleigh, 2016; Inverse Problems in Science and Engineering, submitted. 2016.
4. HT Banks, S Hu, WC Thompson. Modeling and Inverse Problems in the Presence of Uncertainty, Taylor/Francis-Chapman/Hall-CRC Press, Boca Raton, FL. 2014.
5. GEP Box, NR Draper. Empirical Model-Building and Response Surfaces, Wiley Series in Probability and Mathematical Statistics. 1987.
6. M Prague, D Commenges, R Thiébaut. Dynamical models of biomarkers and clinical progression for personalized medicine: the HIV context, Advanced Drug Delivery Reviews. 2013; 65: 954-965.
7. S Lenhart, JT Workman. Optimal Control Applied to Biological Models, Chapman&Hall/CRC/Taylor & Francis Group, Boca Raton, FL. 2007.
8. BM Adams, HT Banks, H Kwon, HT Tran. Dynamic multidrug therapies for HIV: optimal and STI control approaches, Mathematical Biosciences and Engineering. 2004; 1: 223-241.
9. HT Banks, H Kwon, JA Toivanen, HT Tran. A state-dependent Riccati equation-based estimator approach for HIV feedback control, Optimal control Application and Methods. 2006; 27: 93-121.
10. TS Jang, J Kim, H Kwon, J Lee. Hybrid on-off controls for an HIV model based on a linear control problem,Journal of the Korean Mathematical Society. 2015; 52: 469-487.
11. H Schättler, U Ledzewicz. Optimal Control for Mathematical Models of Cancer Therapies: An Application of Geometric Methods, Springer, New York. 2015.
12. WD Shlomchik. Graft-versus-host disease, Nature Reviews Immunology. 2007; 7: 340-352.
13. L Chatenoud. Precision medicine for autoimmune disease, Nature Biotechnology. 2016; 34: 930-932.
14. HT Banks, S Hu, K Link, ES Rosenberg, S Mitsuma, et al. Rosario, Modeling immune response to BK virus infection and donor kidney in renal transplant recipients, Inverse Problems in Science and Engineering. 2016; 24: 127-152.
15. GA Funk, J Steiger, HH Hirsch. Rapid dynamics of polyomavirus type BK in renal transplant recipients, The Journal of Infectious Diseases. 2006; 190: 80-87.
16. GA Funk, R Gosert, P Comoli, F Ginevri, HH Hirsch. Polyomavirus BK replication dynamics in vivo and in silico to predict cytopathology and viral clearance in kidney transplants, American Journal of Transplantation. 2008; 8: 2368-2377.
17. GM Kepler, HT Banks, M Davidian, ES Rosenberg. A model for HCMV infection in immunosuppressed patients, Mathematical and Computational Modeling. 2009; 49: 1653-1663.
18. HT Banks, S Hu, T Jang, HD Kwon. Modeling and optimal control of immune response of renal transplant recipients, Journal of Biological Dynamics. 2012; 6: 539-567.
19. LE Keshet. Mathematical Models in Biology, Siam Classics in Applied Mathematics Series. 2004.
20. HT Banks. Modeling and Control in the Biomedical Sciences. Springer. 1975.
21. HT Banks, JE Banks, R Everett, J Stark. An adaptive feedback methodology for determining information content in stable population studies, Mathematical Biosciences and Engineering. 2015; 13: 653-671.
22. JE Banks, HT Banks, K Rinnovatore, CM Jackson. Optimal sampling frequency and timing of threatened tropical bird 2 populations: a modeling approach, Ecological Modeling. 2015; 303: 70-77.

23. HT Banks, KL Rehm. Experimental design for distributed parameter vector Systems, Applied Mathematics Letters. 2013; 26: 10-14.

24. HT Banks, KL Rehm. Experimental design for vector output systems, Inverse Problems in Science and Engineering. 2014; 22: 557-590.

25. HT Banks, KL Rehm. Parameter estimation in distributed systems: optimal design, Eurasian Journal of Mathematical and Computer Applications. 2014; 2: 70-80.

26. F Kronenberg. Emerging risk factors and markers of chronic kidney disease progression. Nature Reviews Nephrology. 2009; 5: 677-689.

27. K Luttropp, P Stenvinkel, JJ Carrero, R Pecoits-Filho, B Lindholm, et al. Understanding the role of genetic polymorphisms in chronic kidney disease, Pediatric Nephrology. 2008; 23: 1941-1949.

28. M Robson, K Offit. Management of an inherited predisposition to breast cancer,New England Journal of Medicine. 2007; 357: 154-162.

29. FL Baldissera, JER Curyand, J Raisch. A supervisory control theory approach to control gene regulatory networks, IEEE Transactions on Automatic Control. 2016; 61: 16-33.