

## ABSTRACT

NIYONZIMA, YVONNE. Application of Mathematical Modeling in Toxicology and Human Immunodeficiency Virus. (Under the direction of Hien Tran).

1. Nonalcoholic fatty liver disease (NAFLD) impacts 25-30% of the US population. Hepatic steatosis, an increase of more than 5% of lipid content in the liver, is a biological manifestation of NAFLD. Hepatic steatosis can be initiated by lifestyle factors and exposure to some environmental chemicals such as carbon tetrachloride ( $\text{CCl}_4$ ). Past research has shown that  $\text{CCl}_4$  decreases the levels of hepatic very low-density lipoprotein (VLDL) while it increases the levels of sterol regulatory element binding protein 1c (SREBP-1c). The question that needs to be answered is the amount of  $\text{CCl}_4$  it takes to induce hepatic steatosis.

The dose-response, of  $\text{CCl}_4$  vs. hepatic triglycerides, is performed by incorporate the VLDL and SREBP-1c in a quantitative systems toxicology (QST) model describing carbon tetrachloride. This process involves several key components. The first step is the generation of a physiologically based pharmacokinetic (PBPK) model for  $\text{CCl}_4$ , to serves as a quantitative method to establish dose-response relationships for  $\text{CCl}_4$  toxicity regarding hepatic steatosis. The concentration of hepatic  $\text{CCl}_4$  is then fit into an exponential and a logistic equations that describe the relationship between  $\text{CCl}_4$  and VLDL and srebp-1c, respectively. The values from the exponential and logistic equations are substituted in the liver model. The last step is done to estimate the amount of hepatic triglycerides.

2. To prevent (decrease) children who have tested positive for human immunodeficiency virus (HIV) from dying, the world health organization recommends immediate initiation of antiretroviral therapy (ART). However, early initiation of ART in infants (< 3 months) leads to lack persistent HIV-1 immunoglobulin G (IgG) antibodies. Our project aims to answer if there is a difference in some pharmacokinetics (PK) metrics and to establish a mechanistic model that could be used to predict the antibodies in children who are on ART. We use longitudinal data of infant rhesus macaques that were infected after birth but were initiated on ART at different ages .

We will use a non-compartmental analysis approach to analyse the PK metrics (maximum observed concentration, time of the maximum observed concentration, and

the area under the curve) of antibodies (gp41 and gp120). The terminal half-life is also analyzed, but with a biphasic model. For the second part of this project, we devise a mechanistic model, by relying on the biology of antibody production, to fit our longitudinal data with both viral load and antibodies as output. The implications of the mechanistic model can be in answering when children should be started on ART while preserving the HIV-1 specific antibodies.

© Copyright 2022 by Yvonne Niyonzima

All Rights Reserved

# Application of Mathematical Modeling in Toxicology and Human Immunodeficiency Virus

by  
Yvonne Niyonzima

A dissertation submitted to the Graduate Faculty of  
North Carolina State University  
in partial fulfillment of the  
requirements for the Degree of  
Doctor of Philosophy

Applied Mathematics

Raleigh, North Carolina  
2022

APPROVED BY:

---

Cliburn Chan

---

Kevin Flores

---

Min Kang

---

Hien Tran  
Chair of Advisory Committee

## **DEDICATION**

To Adelphine Ndayishimiye Bindaba, my real life hidden figure.

## BIOGRAPHY

Yvonne Niyonzima grew up in Bujumbura, Burundi. When not in school, she enjoyed spending time outdoors playing, especially soccer. She wanted to become a medical doctor (she wanted to make sick kids who could not play outside feel better). She was not at all interested in mathematics (she incorrectly thought it was a waste of time). Fortunately, her mother (a former math teacher in secondary school) told her about some of the beauty of mathematics. The mother also informed her that strong mathematical scores make a great candidate for medical school. From that time on, she started to like mathematics, but only as a means to an end.

In July 2009, the family moved to Raleigh, NC. Upon graduating from Sanderson High School, Yvonne went to Wake Technical Community College as means to save money while improving her English language skills (she had been in America for less than two years). She earned an associate in science degree in two years before transferring to East Carolina University (ECU) to pursue a degree in Chemistry, with the goal of going to medical school. While at ECU, she discovered a passion for mathematics that she decided to add it on as a double major. In her senior year at ECU, she was conflicted because she was in love with math but had always wanted to attend medical school. While attending the Southeastern Conference for Undergraduate Women in Math, she met women who used math in other fields, including medicine. She decided that weekend that she would pursue a doctorate degree in applied mathematics with the hopes of using it in medical research. She knew her undergraduate bachelor of arts degree in mathematics was not enough to be a competitive applicant to a PhD program. That is why she took a year to go to Smith College to pursue a post-baccalaureate in math. There she took more math courses and partook in research. At the end of the year, she came to North Carolina State University to start her journey towards getting a PhD in Applied Mathematics.

Yvonne has several hobbies and passions. She loves going to concerts of christian contemporary artists. She is passionate about soccer. Soccer is her sport of choice because growing up, things seemed to stop during the world cup and African cup of Nations matches. She loved the ambiance and the joy in the air that everyone seemed to have. Yvonne is also passionate about mentoring young girls and encouraging them to pursue a math degree. She always thrive to explain to young kids the beauties and applications of mathematics that she did not get to learn until she was an adult. She likes learning about different cultures and trying their food.

## ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Hien Tran. When deciding on an advisor I was told Dr. Tran is pleasant professor who pushes his students to work and graduate on time. I noticed this first hand. I knew I was inadequate and was afraid at the start of this process. I cannot state how many times I showed up to his office without an appointment because I was lost. Dr. Tran did not even once hesitate to hold my hand while I learned to walk. He instilled confidence in me that when it came time to take off the safety wheels, I knew I was ready. Dr. Tran, thank you for taking me on as your student. You made this process enjoyable and less stressful.

Secondly, I would like to thank my committee members for agreeing to give their time to aid me in the completion of this degree. Dr. Tran, you made me into a researcher that I am today. Dr. Chan, you had confidence in me. You took me on as an intern the second time, when I did not think I deserved it. You were always available to help me with my research even after business hours. Dr. Flores, you sparked my interest in working with data. Dr. Kang, you are the pinnacle that those who can do can also teach. You never failed to capture my attention even when I was tired or uninterested. Your professionalism pushed me to be the best student because I wanted to make it worth your time and efforts.

I would also like to thank Dr. Hisham El-Masri and Dr. Stella Berendam. Thank you for helping me with the projects and being my unofficial advisors. To my friends, thank you very much for your love, support, and encouragement. Dr. Annabel Meade, sitting together on the first day of our graduate school orientation was like kismet. You truly were a godsend. Annabel and Nicholas Meyers, thank you for creating a virtual office so that we can do some work while pretending to be in office during the pandemic. Celia Schatt and Ellie Mainou, thank you for your encouragements and sharing your struggles as fellow graduate students with me. Our conversations always encouraged me because I knew I was not alone. Hang Nguyen and Wendy Guevara, thank you for your friendship and prayers. Talking to you two always make me feel better both in my studies and private life.

Last, but not least, I would like to thank from the bottom of heart everyone in my family. Alidor Nduwayo Bindaba, thank you for the support. Not only did you believe in me, you also gave me financial support whenever I needed it. Magdaleine Ntakandi, your pride (or bragging) and joy in my small accomplishment in school was highly motivating. Haila Bindaba, you are a star. A ma mère (femme chrétienne, femme noire, femme africaine, femme forte et femme belle), je t'aime beaucoup.

# TABLE OF CONTENTS

<b>List of Tables</b> . . . . .	<b>vii</b>
<b>List of Figures</b> . . . . .	<b>viii</b>
<b>Chapter 1 Model Calibration and Parameter Selection</b> . . . . .	<b>1</b>
1.1 Model Fitting . . . . .	1
1.2 Sensitivity Analysis . . . . .	3
1.2.1 Sensitivity Methods . . . . .	3
1.2.2 Sensitivity Rankings . . . . .	11
1.2.3 Global sensitivity . . . . .	11
1.3 K-means Clustering . . . . .	12
1.4 Identifiability Analysis . . . . .	15
1.5 Non-Linear Mixed Effect Model . . . . .	18
<b>Chapter 2 A Quantitative Systems Toxicology (QST) Model for Hepatic Steatosis Induction by Carbon Tetrachloride</b> . . . . .	<b>21</b>
2.1 Motivation . . . . .	21
2.2 Mathematical Modeling . . . . .	23
2.2.1 Physiologically-based pharmacokinetic (PBPK) . . . . .	23
2.2.2 Ingestion and Inhalation Models . . . . .	27
2.2.3 Parameter Impacted by Carbon Tetrachloride . . . . .	31
2.2.4 Liver model . . . . .	34
2.2.5 Complete model . . . . .	37
2.3 Model Validation in rats . . . . .	38
2.3.1 Liver model in rats . . . . .	39
2.3.2 Results in Rats . . . . .	41
<b>Chapter 3 QST Results</b> . . . . .	<b>45</b>
3.1 Simulation in the absence of carbon tetrachloride . . . . .	45
3.2 Ingestion or Inhalation of carbon tetrachloride and the hepatocytes impact . . . . .	46
3.2.1 Carbon tetrachloride v. hepatic triglyceride in the absence of high fat diet . . . . .	47
3.2.2 Carbon tetrachloride v. hepatic triglyceride in the presence of high fat diet . . . . .	47
3.2.3 More carbon tetrachloride v. hepatic triglyceride results . . . . .	48
3.3 Liver Model in the presence of Ethanol . . . . .	49
3.4 Sensitivity analysis of some parameters . . . . .	50
3.5 Our contributions . . . . .	53
<b>Chapter 4 Analysis of immunoglobulin G pharmacokinetics and pharmacodynamics (PK/PD) in infant Rhesus macaques</b> . . . . .	<b>55</b>



<b>Chapter 5</b>	<b>Pharmacokinetic (PK) Analysis of gp41 and gp120 antibodies</b>	<b>58</b>
5.1	Non-Compartment Analysis	58
5.1.1	Area under the curve	59
5.1.2	Maximum Concentration	62
5.1.3	Time to reach maximum concentration	63
5.1.4	Terminal Half-life	64
5.2	Biphasic Model	65
5.2.1	Late Antiretroviral Therapy Initiation	65
5.2.2	Intermediate Antiretroviral Therapy Initiation	72
5.3	Comparison of the the three stages of ART administration	79
5.3.1	Statistical Test	79
5.3.2	Comparison Results	80
<b>Chapter 6</b>	<b>Modeling Viral Load and Antibodies</b>	<b>84</b>
6.1	Parameter Estimation	84
6.1.1	Model Selection	85
6.1.2	Sensitivity Analysis	88
6.1.3	K-means clustering	89
6.1.4	Identifiability	92
6.1.5	Parameter fitting of sensitive and identifiable parameters	95
6.2	Model Validation	97
6.3	Our contributions	100
<b>References</b>		<b>102</b>
<b>APPENDIX</b>		<b>114</b>
Appendix A	Quantitative System Toxicology	115
A.1	Physiological based pharmacokinetic (PBPK)	115
A.1.1	Michaelis-Menten Kinetics	115
A.1.2	Non-compartment PBPK equations	116
A.1.3	Common Compartments PBPK equations	117
A.1.4	PBPK Equations acronyms	118
A.1.5	PBPK parameters	119
A.2	Liver Model	123
A.2.1	Tables of Parameters that were changed	123
A.2.2	Plots of Liver compartments	125

## LIST OF TABLES

Table 3.1	Global Sensitivity values . . . . .	53
Table 5.1	Area under the curve for intermediate and Late data . . . . .	61
Table 5.2	Maximum concentration for Intermediate and Late data . . . . .	62
Table 5.3	Time to reach maximum Concentration for intermediate and Late data	63
Table 5.4	Half-life gp120 Late Data . . . . .	69
Table 5.5	tab: Half-life gp41 Late Data . . . . .	70
Table 5.6	The remaining parameters from the biphasic model for gp120: Late Data . . . . .	71
Table 5.7	The remaining parameters from the biphasic model for gp41: Late Data	72
Table 5.8	tab: Half-life gp41 Intermediate Data . . . . .	74
Table 5.9	Half-life gp120 Intermediate Data . . . . .	76
Table 5.10	tab: The remaining parameters from the biphasic model for gp120: Intermediate Data. . . . .	77
Table 5.11	tab: The remaining parameters from the biphasic model for gp41: Intermediate Data. . . . .	78
Table 5.12	Kruskal-Wallis Test for gp41 . . . . .	81
Table 5.13	Kruskal-Wallis Test for gp120 . . . . .	82
Table 6.1	Parameters definition and units of the HIV model . . . . .	88
Table 6.2	Parameters that are non-sensitive or non-identifiable. NA refer to sensitive and identifiable parameters . . . . .	94
Table 6.3	Sensitive and Identifiable parameters, gp120 . . . . .	95
Table 6.4	Sensitive and Identifiable parameters, gp41 . . . . .	96
Table 6.5	Model's residual errors . . . . .	97
Table A.1	Equations acronyms . . . . .	118
Table A.2	Percent of the organs' volume (in L) . . . . .	119
Table A.3	Percent of rate flow to the organs (in L/hr) . . . . .	120
Table A.4	Parameters used in the PBPK model for Carbon tetrachloride in both rats and humans . . . . .	121
Table A.5	Partition coefficient tissue:air . . . . .	122
Table A.6	Parameters used in the liver model and their corresponding values . .	123
Table A.7	Parameters used as initial conditions in the liver model and their corresponding values . . . . .	124

## LIST OF FIGURES

Figure 1.1	A function $y(q)$ and corresponding derivative $\frac{dy}{dq}(\hat{q})$ . . . . .	4
Figure 1.2	A solution to (1.13) with $k = 10, a = 1, p_0 = 0.05$ . . . . .	9
Figure 1.3	Plot of derivatives of P with respect to $a$ and $k$ . . . . .	10
Figure 1.4	Plot of local sensitivity values. . . . .	13
Figure 1.5	Plot of Identifiability clusters from local sensitivity. . . . .	15
Figure 2.1	A schematic example of a PBPK model . . . . .	25
Figure 2.2	A compartment of a PBPK model . . . . .	25
Figure 2.3	A Breakdown schematic of parameters input in the PBPK models . .	27
Figure 2.4	Schematic of PBPK model used in human. . . . .	29
Figure 2.5	Data fit after estimating $k_a$ and $me t_{V_{max}}$ . . . . .	31
Figure 2.6	Instantaneous effect of $CCl_4$ on VLDL . . . . .	33
Figure 2.7	Logistic approximation of srebp-1c. . . . .	34
Figure 2.8	A schematic of the liver model. The porto-central axis of the sinusoid is considered to be the repeating unit of the liver. . . . .	35
Figure 2.9	Some variables and conversions included in each hepatic compartment. . . . .	37
Figure 2.10	$CCl_4$ v Triglycerides Outline of Simulation. . . . .	38
Figure 2.11	Parameter estimation for rats . . . . .	41
Figure 2.12	Schematic of injection PBPK model used to validate rats results. . .	42
Figure 2.13	Plot of the number concentrations of hepatocytes triglycerides in rats in the presence of $CCl_4$ and high fat diet. . . . .	44
Figure 3.1	simulation of TGs under normal circumstances with or without food in humans. . . . .	46
Figure 3.2	$CCl_4$ vs. Hepatic triglycerides simulation . . . . .	47
Figure 3.3	$CCl_4$ and high fat diet vs. Hepatic triglycerides simulation . . . . .	48
Figure 3.4	Plot of the concentrations of hepatocytes triglycerides in rats and humans. . . . .	49
Figure 3.5	Triglycerides in. body as a response to ethanol consumption.. . . .	50
Figure 3.6	Plot of global sensitivity of some liver parameters . . . . .	52
Figure 5.1	Fit of antibodies using Biphasic equation . . . . .	66
Figure 5.2	(a) biphasic vs. (b) NCA gp41 for Late initiation ART data. . . . .	67
Figure 5.3	(a) biphasic vs. (b) NCA gp120 for Late initiation ART data. . . . .	67
Figure 5.4	Fit of gp41 antibodies using Biphasic equation for intermediate data	73
Figure 5.5	Fit of gp120 antibodies using Biphasic equation for intermediate data	75
Figure 5.6	Box plots of gp41 for the AUC, Cmax, half-life and Tmax. . . . .	80
Figure 5.7	Box plots of gp120 for the AUC, Cmax, half-life and Tmax. . . . .	83
Figure 6.1	Schematic of antibodies and viral load model . . . . .	87

Figure 6.2	Plot of sensitivity values. . . . .	90
Figure 6.3	Plot of sensitivity clusters: gp120. . . . .	91
Figure 6.4	Plot of sensitivity clusters: gp41. . . . .	92
Figure 6.5	Model fit of viral Load and gp120 Antibodies. . . . .	95
Figure 6.6	Model fit of viral Load and gp41 Antibodies. . . . .	96
Figure 6.7	Model fit of viral Load and gp41 Antibodies on 90% distribution. . . .	98
Figure 6.8	Model fit of viral Load and Antibodies on 90 % distribution. . . . .	99
Figure 6.9	Model fit of viral Load and Antibodies on 95 % distribution. . . . .	100
Figure A.1	Plot of the concentrations of hepatocytes triglycerides in rats and humans. . . . .	125

## CHAPTER

# 1

# MODEL CALIBRATION AND PARAMETER SELECTION

## 1.1 Model Fitting

Model fitting is an important part of all sciences that use quantitative measurements. Scientists often explore the relationships between models and measurements (1). In certain scenarios, there exists several model selection criteria, proposed over the years, that have become common procedures (2). The downside is that these procedures were designed for complete data, which is rare in actuality. Fitting the proposed model to data can lead to problems of over-fitting or under-fitting. Overfitting is the use of models or procedures that include more terms than are needed (1). Underfitting occurs when the model fails to capture the patterns in the data (3). An example is that a straight line under-fits a third-order polynomial. On the other hand, a polynomial over-fits a linear function. Occam's Razor, or the principle of parsimony, calls for using models and procedures that contain all that is necessary for the modeling but nothing more. An example, if the relationship can be captured by a linear function then using a quadratic violates the law parsimony (4).

To decrease the chances of over-fitting or under-fitting, it is crucial to do a variable selection of the proposed model as a means to select the "best" subset of predictor variables. The term "best", here, refers to a balance between the number of explanatory variables and goodness of fit (1). Various methods have been proposed to avoid model over-fitting and find the "best" model. The majority of these methods comprise one of the following strategies: 1) use validation data set(s) to evaluate the performance of the fitted model 2) penalize models with more parameters (5).

1. Cross validation tests the effectiveness of models. It can also be used as a re-sampling procedure to evaluate a model in case of a limited data. To perform cross validation, a sample/portion of the data is set aside in order to be used later to test/validate the model(6).
2. Penalizing additional parameters (predictors) can be achieved by either performing the traditional model selection (based on different criteria) or applying penalized regression models (7). Some criteria of traditional model selection:

(a) Selection based on: Adjusted R-squared, AIC and BIC

- i. R-squared is the percentage of outcome variable variation explained by the model, and describes how close the data are to the fitted regression.
- ii. AIC is "Akaike's Information Criterion", and BIC is "Schwartz' Bayesian Criterion." Both aim at achieving a compromise between model goodness of fit and model complexity. The only difference between AIC and BIC is the penalty term, where BIC is more stringent than AIC. The preferred models are those with minimum AIC/BIC.

(b) Best subset / forward / backward / step-wise selection

- i. In forward selection, the most significant variable (based on certain pre-set confidence level) is added to the model one at a time, until no additional variable meets the criterion.
- ii. Backward selection starts with the full model that includes all the variables of interest, and then drop non-significant variables one at a time, until all the variables left are significant.
- iii. Step-wise selection allows both adding and dropping variables to allow dropped variables to be reconsidered (5).

## 1.2 Sensitivity Analysis

The parameters in any model play a role in the output. However, the impact on the model's output differs among the parameters present. We are more interested in the parameters whose impacts are substantial. To know which of our parameters will produce substantial impact in the output, we perform sensitivity analysis.

Sensitivity analysis is the study of the relationship between the parameters and the output of the model. We look at how small changes in a model's parameter affect the model output. Sensitivity analysis is used to validate results in different areas of mathematics as well as the area where mathematics is applied (biology, physics, finance, etc). A parameter is said to be sensitive when a small change in its value leads to a large change in the output. This is crucial in understanding the relationship between parameters and output. Sensitivity analysis is also useful when estimating parameters of a model. Parameters that are less sensitive are normally fixed to values found in the literature, while those that are sensitive are then estimated using an inverse problem formation. For this reason, sensitivity analysis increases the confidence of the estimated parameters values. There are two types of sensitivity analysis, local and global.

### 1.2.1 Sensitivity Methods

Consider a function  $y(q)$ . Its derivative  $\frac{dy}{dq}$  defined as

$$\frac{dy}{dq}(\hat{q}) = \lim_{h \rightarrow 0} \frac{y(\hat{q} + h) - y(\hat{q})}{h}, \quad (1.1)$$

at a point  $(\hat{q})$  determines how much  $y$  changes as  $q$  changes near  $(\hat{q})$ . Large absolute values for  $\frac{dy}{dq}(\hat{q})$  indicate large changes in  $y$  near  $(\hat{q})$ , whereas small values of  $\frac{dy}{dq}(\hat{q})$  indicate small changes in  $y$  (8), as presented in Figure 1.1.

#### Direct Computation

We use a derivative-based approach for the sensitivity presented in this section. Consider the following ODE and output equations:

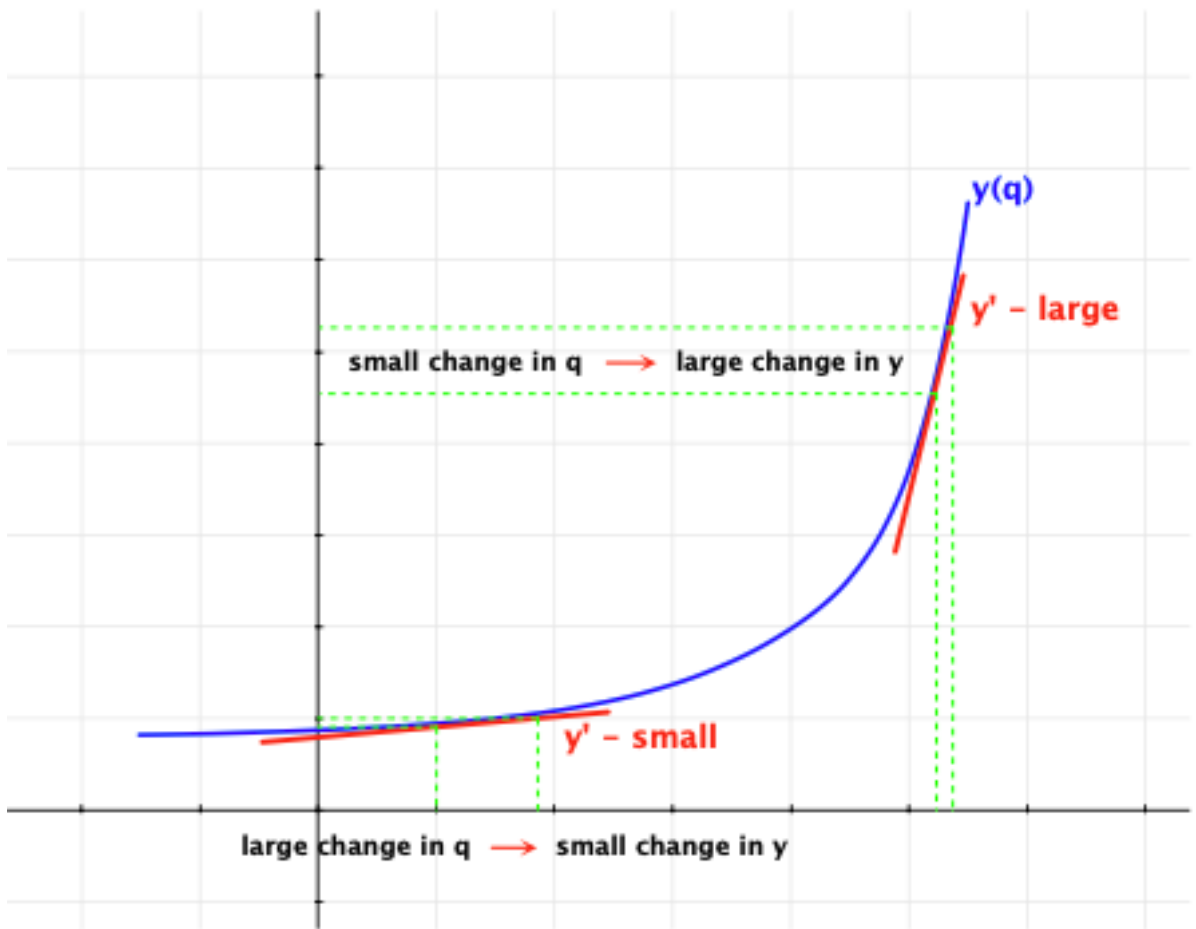


Figure 1.1: A function  $y(q)$  and corresponding derivative  $\frac{dy}{dq}(\hat{q})$



$$\frac{d\mathbf{x}}{dt} = f(t, \mathbf{x}; \mathbf{q}) \quad (1.2a)$$

$$y = g(t, \mathbf{x}; \mathbf{q}), \quad (1.2b)$$

where  $\mathbf{x}$  and  $\mathbf{q}$  are respectively vectors of the state variables and parameters in the model.

Since we are concerned about the sensitivities of our parameters with respect to the output  $y$ , we compute the partial derivatives of  $y$  with respect to  $\mathbf{q}$ ,  $\left(\frac{\partial y}{\partial q}\right)$ . In reality, this computation can be difficult especially when our output  $y$  does not have an analytical solution. However, we can apply automatic differentiation (AD) that is able to compute the desired derivatives directly by using elementary derivative rules. The partial derivative is then computed by differentiating both sides of the output of Eq (1.2) with respect to  $q$  to obtain:

$$\frac{\partial y}{\partial q} = \frac{\partial g}{\partial t} \frac{\partial t}{\partial q} + \frac{\partial g}{\partial x} \frac{\partial x}{\partial q} + \frac{\partial g}{\partial q} \frac{\partial q}{\partial q} \quad (1.3a)$$

$$= \frac{\partial g}{\partial x} \frac{\partial x}{\partial q} + \frac{\partial g}{\partial q}, \quad (1.3b)$$

where  $\frac{\partial t}{\partial q} = 0$  and  $\frac{\partial q}{\partial q} = 1$ . Given the functions  $f$  and  $g$ , the partial derivatives  $\frac{\partial g}{\partial x}$ ,  $\frac{\partial g}{\partial q}$ ,  $\frac{\partial f}{\partial x}$ , and  $\frac{\partial f}{\partial q}$  from  $g$  can be computed directly using elementary derivative formulas from calculus. These calculations can be burdensome to calculate by hand. Automatic differentiation numerically repeatedly implements the chain rule and basic arithmetic equations to compute the total derivative of a function with accuracy to working machine precision (9) (10) (11) (12).

Using automatic differentiation, the partial derivative  $\frac{\partial x}{\partial q}$  is obtained:

$$\frac{\partial}{\partial \mathbf{q}} \left( \frac{d\mathbf{x}}{dt} \right) = \frac{d}{dt} \left( \frac{\partial \mathbf{x}}{\partial \mathbf{q}} \right) = \frac{\partial}{\partial q} f(t, \mathbf{x}; q) = \frac{\partial f}{\partial x} \frac{\partial x}{\partial q} + \frac{\partial f}{\partial q}. \quad (1.4)$$

This is a differential equation for the sensitivity variable  $\frac{\partial x}{\partial q}$

with

$$\left. \frac{\partial x}{\partial \mathbf{q}} \right|_{t=0} = \frac{\partial x_0}{\partial \mathbf{q}}.$$

The disadvantage of AD is computational cost and CPU memory.

In addition to AD, we can compute the derivatives via Finite Difference (FD) method

and a one-step complex method.

### Finite Differences

Finite Difference (FD) method estimates  $\frac{dy}{dq_i}$ , where  $q_i$  is the  $i^{th}$  parameter in the vector  $\mathbf{q}$ , the Taylor series,

$$f(x+h) = f(x) + h \frac{df}{dx} + \frac{h^2}{2!} \frac{d^2f}{dx^2} + \dots + \frac{d^n f}{dx^n} + \dots \quad (1.5)$$

The derivative  $\frac{df}{dx}$  can be approximated by the first two terms as follows:

$$\frac{df}{dx} = \frac{f(x+h) - f(x)}{h} + \mathcal{O}(h) \quad (1.6)$$

where  $h$  is the step-size and  $\mathcal{O}(h)$  is the truncation error. Equation (1.6) is referred to as the forward difference. We can also estimate the backward difference as

$$\frac{df}{dx} = \frac{f(x) - f(x-h)}{h} + \mathcal{O}(h) \quad (1.7)$$

A combination of the forward and backward gives the central finite difference,

$$\frac{df}{dx} = \frac{f(x+h) - f(x-h)}{2h} + \mathcal{O}(h^2) \quad (1.8)$$

The step size,  $h$ , should be chosen in order to minimize the truncation error and subtractive cancellation error due to finite precision arithmetic (10). The central difference possesses the advantage of being accurate on an order of  $h^2$ , instead of the order  $h$  for the forward and backward differences (8). A good option for the step size in the backwards and forward differences is  $h = \sqrt{\text{macheps}} \cdot q_i$  and for central difference  $h = \sqrt[3]{\text{macheps}} \cdot q_i$  (13). These values are given in terms of macheps, the relative error in computing  $f(x)$ . These values minimize the round-off and machine error in the calculations.

### Complex Step Method

The one-step complex method uses complex variables to estimate derivatives. The method originated with the work of Lyness and Moler (14) and Lyness (15). It has been shown to be extremely accurate, robust while maintaining a reasonable computational cost. The potential of this technique has been recognized and used for sensitivity analysis in computational fluid dynamics. The method is also popular in aerodynamic optimization. Complex

step method is known to be accurate and computationally inexpensive for small systems (16). The method uses complex functions to calculate derivatives (17). It uses the fact that complex functions are the extensions of their real counterparts (10).

Consider a complex function:  $f = u + i v$  of the complex variable  $z = x + i y$  where  $x$  is the real part  $Re(z)$  and  $y$  is the imaginary part  $Im(z)$  where

$$f(z) = u(x + i y) + i v(x + i y)$$

is an analytic complex functions with  $u$  and  $v$  being the real and imaginary parts of  $f$ , respectively.

Because  $f$  is analytic, the Cauchy-Riemann equations are satisfied. That is, the components satisfy

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \quad (1.9)$$

Using the forward difference approximation, we rewrite (1.9) as,

$$\frac{\partial u}{\partial x} \approx \frac{v(x + i(y + h)) - v(x + i y)}{h}, \quad (1.10)$$

where  $h$  is the step size and a real number.

If we restrict the function to the real axis, then

$$y = 0 \quad (1.11a)$$

$$u(x) = f(x) \quad (1.11b)$$

$$v(x) = 0 \quad (1.11c)$$

This lead the previous Equation (1.9) to become

$$\frac{\partial f}{\partial x} = \frac{\partial u}{\partial x} \approx \frac{v(x + i h) - v(x)}{h} = \frac{v(x + i h)}{h} = \frac{Im(f(x + i h))}{h}$$

The approximation of the first derivative of  $f$  at a given parameter  $x$  is therefore:

$$\frac{df}{dx} \approx \frac{Im(f(x + i h))}{h}, \quad (1.12)$$

where  $Im$  represent the imaginary part of the output and  $h$  is a small positive perturbation

of  $x$ .  $\frac{df}{dx}$  is called the complex step derivative approximation.

The complex step method's error can be determined by utilizing the Taylor series expansion with an imaginary step  $ih$ . Since the function  $f$  is analytic and real in real variables, the Taylor series about a real point  $x$  is;

$$f(x + ih) = f(x) + ih \frac{df}{dx} - \frac{h^2}{2!} \frac{d^2f}{dx^2} - i \frac{h^3}{3!} \frac{d^3f}{dx^3} + \dots$$

Considering the imaginary parts of the left and right hand side of the previous equation and dividing by the step side  $h$ , we get:

$$\frac{df}{dx} = \frac{\text{Im}(f(x + ih))}{h} + \frac{h^2}{3!} \frac{d^3f}{dx^3}.$$

Hence we have  $\mathcal{O}(h^2)$  error. Because there is no subtractive cancellation,  $h$  can be reduced to very small values to achieve higher accuracy in the derivative (10).

One of the advantages of the complex-step method is not having to face the dilemma of taking small steps  $h$  (18). In this method, there is no subtraction operation that leads to subtractive cancellation errors. The lack of cancellation error is an advantage over finite-differences method. It also has an implementation advantage over automatic differentiation. It is normally quicker and uses less memory because it does not have to evaluate multiple functions. The method can be implemented in MATLAB by evaluating  $f(x + ih)$  and recovering the imaginary component of the output over  $h$ .

### Comparison of Finite Difference and Complex Step Method

For our models, we use one-step complex method because it allows us to take  $h$  to be as small as we want and still give accurate results. We consider a logistic model example to motivate the reason we picked this method over the finite difference method. Consider the Verhulst-Pearl logistic equation studied in (16)(19)

$$\frac{dp}{dt} = a p \left(1 - \frac{p}{k}\right), \quad (1.13a)$$

$$p(0) = p_0, \quad (1.13b)$$

where  $a$  is the intrinsic growth rate and  $k$  is the carrying capacity. The exact solution of the differential equation (1.13) is given by

$$p(t) = \frac{k p_0}{p_0 + (k - p_0) e^{-a(t-t_0)}} \quad (1.14)$$

The model's parameters are  $r, k, p_0$  that can be fixed to desired values and to obtain the solution to (1.13) as shown in Figure 1.2 .

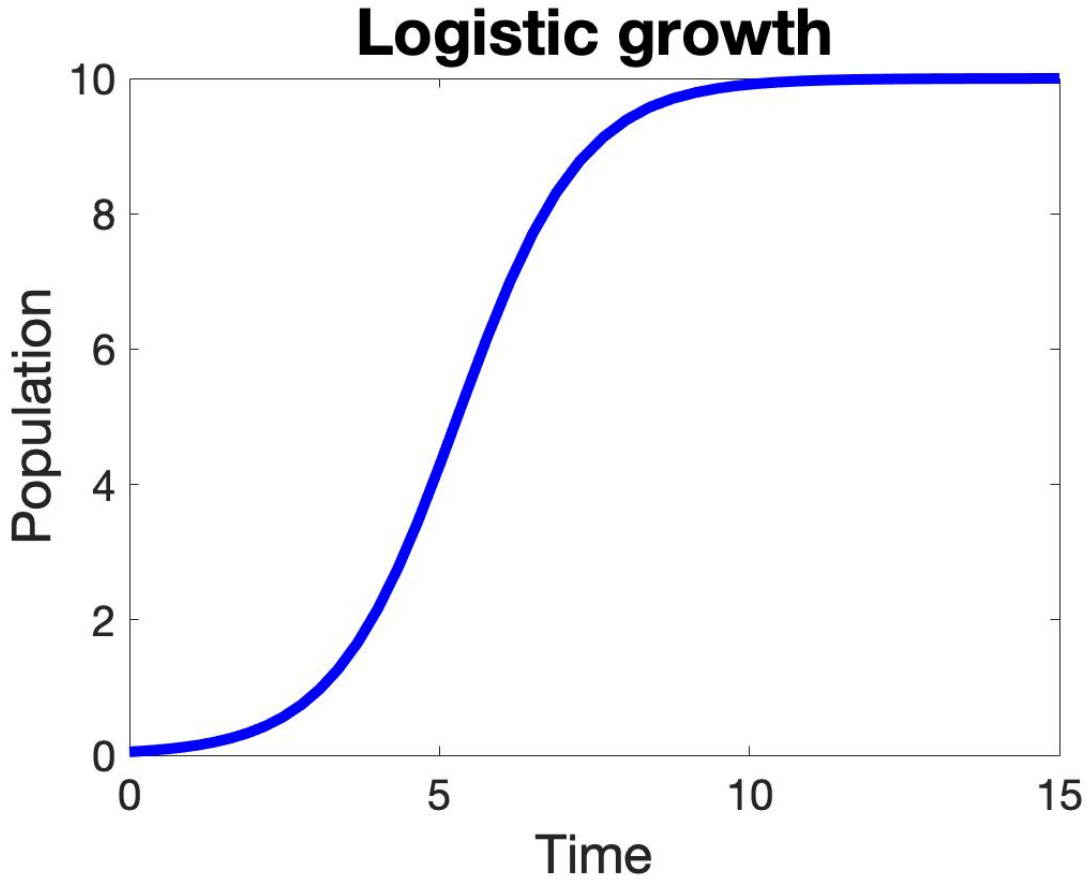
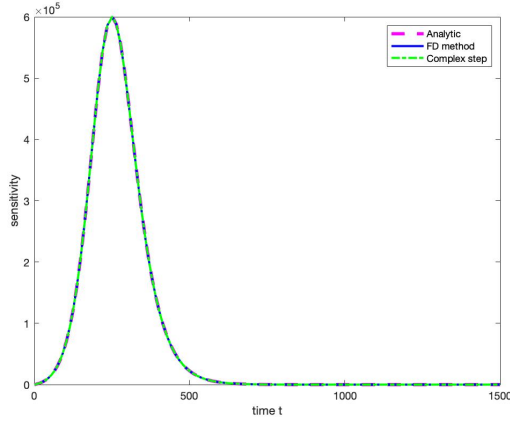


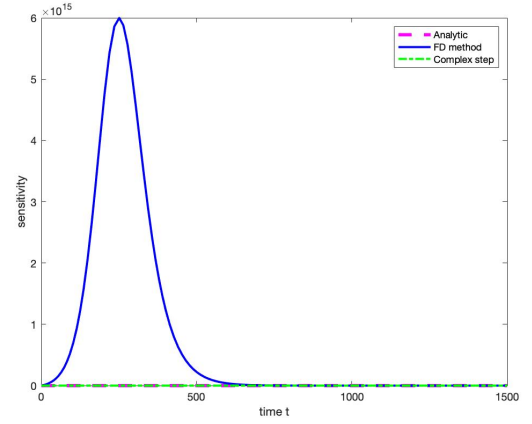
Figure 1.2: A solution to (1.13) with  $k = 10, a = 1, p_0 = 0.05$

Using the analytical solution of the logistic model, we can compute the sensitivity variables ( $p_a = \frac{dp}{da}$  and  $p_k = \frac{dp}{dk}$ ) and compare them to the results of the approximated first derivatives obtained via FD methods and one-step complex methods. In Figure 1.3, the

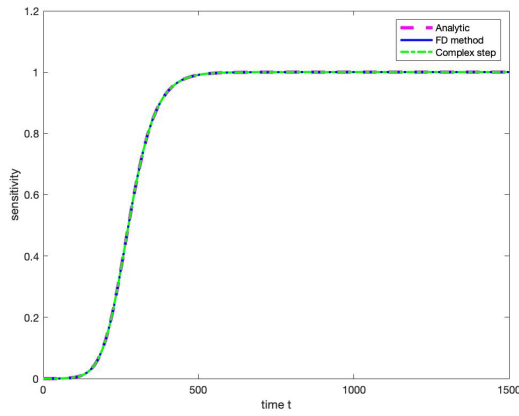
magenta dash line represent the analytical solution, the solid blue line is for finite difference and the dotted green line is from complex step method. The figures on top and bottom are the sensitivity of  $p$  with respect to  $a$  and  $k$ , respectively. On the left side, the three methods appear to be the same when using  $h = 10^{-5}$ . However, when we make the step size  $h$  smaller,  $h = 10^{-15}$ , the finite difference method becomes inaccurate (Figure 1.3b, 1.3d). With machine accuracy  $1 * 10^{-323}$ , we can make  $h = 10^{-320}$  and the complex step method is still going to be accurate (10).



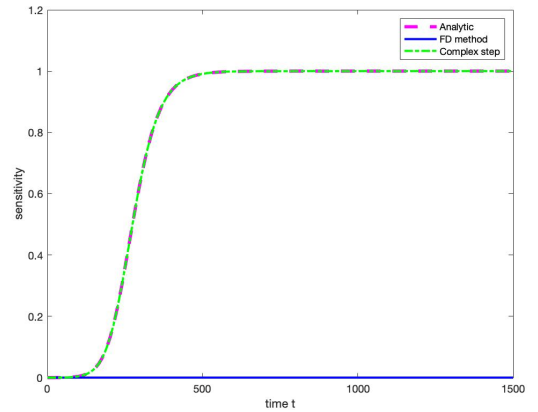
(a) Plot of  $\frac{dp}{da}$  with  $h = 10^{-5}$  for FD and  $h = 10^{-16}$  for one-step complex



(b) Plot of  $\frac{dp}{da}$  with  $h = 10^{-15}$  for both methods



(c) Plot of  $\frac{dp}{dk}$  with  $h = 10^{-5}$  for FD and  $h = 10^{-16}$  for one-step complex



(d) Plot of  $\frac{dp}{dk}$  with  $h = 10^{-15}$  for both methods

Figure 1.3: Plot of derivatives of P with respect to  $a$  and  $k$

### 1.2.2 Sensitivity Rankings

After the calculation of the sensitivity coefficients, we compare the sensitivity with respect to different parameters, to determine the most (or least) influential model's parameters. Because a model's parameters are likely to be in different units, we non-dimensionalize the parameters by multiplying by the parameter value,

$$q_j \frac{\partial y_i}{\partial q_i}. \quad (1.15)$$

We finally take the the  $L^2$  norm of (1.15) to give the normalized sensitivity rankings,

$$C_i(q) = \left\| q_i \frac{\partial y_i}{\partial q_j} \right\| = \sqrt{\int_{t_0}^t \left| q_i \frac{\partial y_i}{\partial q_i} \right|^2 dt} = q_i \sqrt{\int_{t_0}^t \left| \frac{\partial y_i}{\partial q_i} \right|^2 dt} \quad (1.16)$$

Upon computing the ranking values, we rank these parameter sensitivity values from largest to smallest. For simplicity, it is considered that the larger the coefficients, then the more sensible is the system output with respect to that parameter.

Parameters whose  $C_i(q)$  are large are considered to be sensitive. When looking at several parameters, determining which  $C_i(q)$  is large becomes relative. Rather than arbitrary determining which  $C_i(q)$  is large, we use k-means clustering in deciding sensitive parameters.

### 1.2.3 Global sensitivity

In biological problems such as system toxicology or human immunodeficiency virus, a model's parameter is likely to be different in different individuals. This is because the parameters is determined by different factors such as: height, weight, race, comorbidities, lifestyle, etc. Usually a specific parameter is presented with a mean and standard deviation (or variance to represent the population). When doing the sensitivity analysis, we want to take into account these difference and ensure that we are performing the calculations across the population instead of a person. This is why it is good policy to carry out the calculations of the global sensitivity.

The value  $C_i(q)$  in equation (1.16) refers to the local sensitivity. We use the same techniques to carry out global sensitivity. To perform the global sensitivity, the parameter is assumed to follow some type of distribution  $F(\mathbf{q})$  on the population level. One can then perform the expected value of the sensitivity by using the equation,

$$\mathbb{E}_{\mathbf{q}} C_i = \int_Q C_i(\mathbf{q}) dF(\mathbf{q}). \quad (1.17)$$

However due to the high dimensional space of  $Q$ , and the consequence of its inconvenient numerical integration, it is preferable to use Monte Carlo approximation. A sequence of independent parameter values  $\{q_j\}_{j=1}^M$  is generated from the distribution  $F(\mathbf{q})$  and the expected value with a sample mean is approximated,

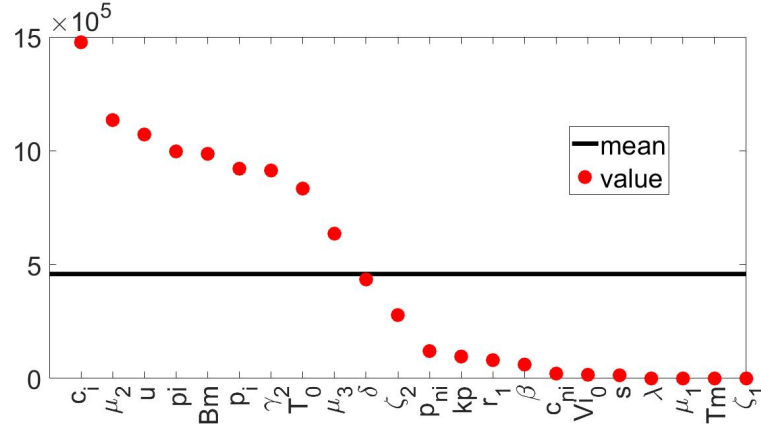
$$\overline{C_i} = \frac{1}{M} \sum_{j=1}^M C_i(\mathbf{q}_j). \quad (1.18)$$

Because each sample mean is a Monte Carlo estimate, the combination of Law of Large numbers and central limit theorem indicate that for a sufficiently large  $M$ , these samples are representatives of the entire population and convergence is asymptotically guaranteed to behave as a normal distribution (20).

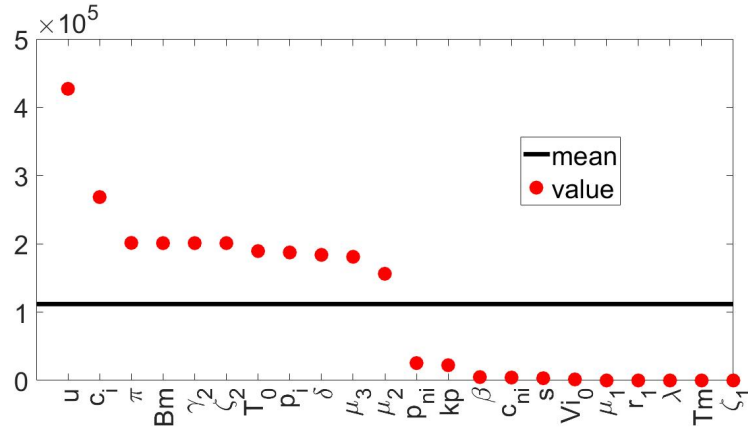
### 1.3 K-means Clustering

As mentioned, parameters with large coefficients  $C_i(q)$  are considered to be sensitive. However, it can be tricky to decide what is considered a large value in a reproducible manner. One method that is utilized is drawing a line that represent the mean of all the values (black line in Figure 1.4). The parameters above the lines are then considered sensitive whereas the parameters below are non-sensitive. Figure 1.4 represent the local sensitivity rankings of a human immunodeficiency virus (HIV) model presented later. At times, drawing the line to determine sensitive values does not work. In this case we use clustering to divide the parameters into sensitive and non-sensitive clusters. Clustering is useful when there are two sensitivity values that appear to be almost equal, but one values is above the mean line while the other is below.





(a) gp120.



(b) gp41.

Figure 1.4: Plot of local sensitivity values.

To determine the clustering of the data, one can use the naive way of considering all the possible clustering and select the one that is most sensible according to a criterion or rationale (21). An example of a naive approach is to choose the clustering that optimizes a pre-selected criterion, which quantifies the requirement for more “similar” vectors to be in the same cluster and less “similar” vectors to be in different clusters. This naive technique is problematic because the number of all possible clustering is huge, even for a moderate number of patterns. A solution to this situation is the development of clustering algorithms that consider only a small fraction of the possible clustering. The considered clustering depend on the specific algorithmic procedure (21).

There are several developed clustering algorithms. We are interested in the k-means

clustering. K-means clustering is the most known clustering algorithm of unsupervised learning problem.

In k-means clustering, a vector of parameters  $\theta_j$ , also called *cluster representatives*, correspond to points in the  $l$ -dimensional space. The aim is to determine the values of  $\theta_j$  in order to characterize the clustering structure of the data set  $X$ . We assume that each *cluster representative* uniquely belongs to a single cluster. k-means assumes that the number of clusters,  $k$ , is known. Its objective is to position the points  $\theta_j, j = 1, \dots, k$ , into areas that are dense in points of  $X$  (clusters). The k-means algorithm is of iterative nature. It starts with some initial estimates  $\theta_1(0), \dots, \theta_k(0)$ , for the parameter vectors  $\theta_1, \dots, \theta_k$ .

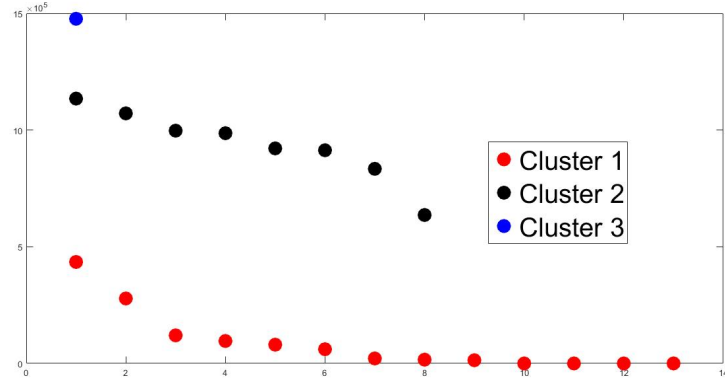
The k-means algorithm follows two steps:

1. the vectors  $x_i$  that are close to each  $\phi_j(t-1)$  are identified
2. the new (updated) value of  $\theta_j$ ,  $\theta_j(t)$ , is computed as the mean of the data vectors that lie closer to  $\theta_j(t-1)$
3. The algorithm terminates when no changes occur in  $\theta_j$ 's between two consecutive iterations.

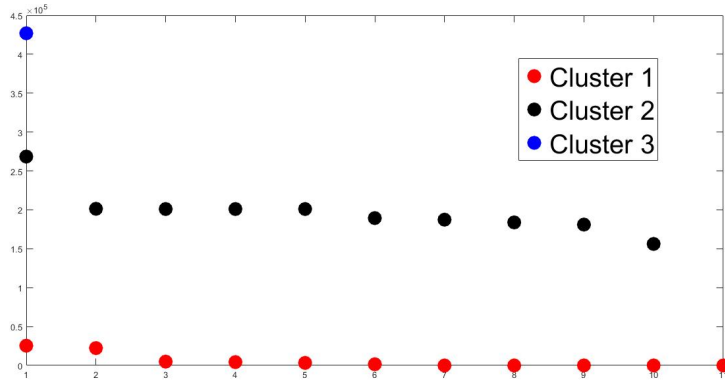
In reality, the algorithm seeks to minimize the following cost function  $J(\theta, U)$ . Equation (1.19) implies that k-means minimizes the sum of the squared Euclidean distances of each data vector from its closest parameter vector (21).

$$\begin{aligned}
 J(\theta, U) &= \sum_i^N \sum_j^k u_{ij} \|x_i - \theta_j\|^2 \\
 \theta &= [\theta_1^T, \dots, \theta_k^T]^T \\
 u_{ij} &= \begin{cases} 1 & x_i \text{ is closest to } \theta_j \\ 0 & \text{otherwise} \end{cases}
 \end{aligned} \tag{1.19}$$

Utilizing the sensitivity coefficients from Figure 1.4 of an HIV model, we can use K-means clustering with three clusters (ultra-sensitive, sensitive, and non-sensitive). Figure 1.5 shows the clustering. In this case, any parameters corresponding to red dot values will be considered non-sensitive.



(a) gp120.



(b) gp41.

Figure 1.5: Plot of Identifiability clusters from local sensitivity.

## 1.4 Identifiability Analysis

After performing the sensitivity analysis, we proceed to parameter selection, which is significant for the model calibration. The parameters that were found to be sensitive are further evaluated to determine which ones can be uniquely identified. A model is considered identifiable if and only if there is an unique input-output behavior for every parameter set. For a model that has more parameters than what can be uniquely determined under ideal

experimental conditions then it is unidentifiable (22). For an input-output map,

$$y = f(q), \quad q = [q_1, \dots, q_p].$$

The parameter  $q$  is considered identifiable at  $q^*$  if for any  $q^* \in \mathbb{Q}$ ,  $f(q) = f(q^*)$  implies  $q = q^*$ . When this condition holds for all  $q^* \in I(q)$ , we say that the parameter  $q$  is identifiable in the space  $I(q)$ .

In certain case, it is obvious that parameters are not identifiable. This is the case of when parameters form product or quotient. As an example, consider the following ODE,

$$\frac{d y}{d t} = a b y(t).$$

On the case of the ODE above, neither " $a$ " nor " $b$ " is identifiable because there are infinitely many numbers that can yield to the same product. However in many cases, it is not apparent whether or not the parameters are identifiable. In the latter instance, there are various methods to perform the identifiability of parameters. The focus is given to the sensitivity-based method for local identifiability analysis. It is important to note that, the structure of the model and the data available can affect whether a parameter is identifiable.

To motivate this technique of sensitivity-based method for local identifiability analysis, the first order Taylor expansion is considered (20). In the inverse least-squares formulation for parameter estimation (Equation (1.20)),  $y_d$  represents the data,  $y_m$  represent the model output and  $q$  is the parameter.

$$J(q) = \sum_{i=1}^N (y_d(t_i) - y_m(t_i, q))^2. \quad (1.20)$$

We assume that  $J(q^*) = 0$  where  $q^*$  is the local optimized parameter. The assumption that  $J(q^*) = 0$  implies that  $y_d(t_i) - y_m(t_i, q^*) = 0$ .

We estimate  $y_m(t, q)$  using the first two terms of the Taylor series expansion around  $q = q^*$  gives:

$$y_m(t, q) \approx y_m(t, q^*) + \frac{\partial y_m}{\partial q}(t, q^*)(q - q^*)$$

Substituting this approximation for  $y_m(t, q)$  in equation (1.20) we obtain,

$$J(q) = \sum_{i=1}^N \left( y_d(t_i) - y_m(t, q^*) - \frac{\partial y_m}{\partial q}(t, q^*)(q - q^*) \right)^2 \quad (1.21)$$

since,

$$y_d(t_i) - y_m(t, q^*) = 0,$$

(1.21) becomes equation (1.22)

$$J(q) = \sum_{i=1}^N \left( \frac{\partial y_m}{\partial q}(t, q^*)(q - q^*) \right)^2 \quad (1.22)$$

Letting  $\Delta q = q - q^*$ ,  $S_{ij} = \frac{\partial y_m}{\partial q_j}(t_i, q^*)$  and  $S$  be the  $n$  by  $m$  matrix of sensitivity coefficients given by

$$S = \begin{bmatrix} \frac{\partial y(t_1)}{\partial q_1} & \frac{\partial y(t_1)}{\partial q_2} & \cdots & \frac{\partial y(t_1)}{\partial q_m} \\ \frac{\partial y(t_2)}{\partial q_1} & \frac{\partial y(t_2)}{\partial q_2} & \cdots & \frac{\partial y(t_2)}{\partial q_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y(t_n)}{\partial q_1} & \frac{\partial y(t_n)}{\partial q_2} & \cdots & \frac{\partial y(t_n)}{\partial q_m} \end{bmatrix}$$

then  $J(q)$  can be written as

$$J(q) = (S\Delta q)^T (S\Delta q) = \Delta q^T S^T S \Delta q \quad (1.23)$$

Here,  $S^T S$  is a square matrix known as the Fischer information matrix.

Suppose that  $\Delta q$  is an eigenvector of  $S^T S$ , then:  $S^T S \Delta q = \lambda \Delta q$ . Substituting  $S^T S \Delta q = \lambda \Delta q$  into (1.23), we obtain

$$J(q) = \Delta q^T \lambda \Delta q = \lambda \Delta q^T \Delta q = \lambda \|\Delta q\|^2. \quad (1.24)$$

If we assume that  $\lambda = 0$  then  $J(q) = J(q^* + \Delta q) = 0$ . In addition, if  $\Delta q$  is an eigenvector, then for an arbitrary  $k$ ,  $k\Delta q$  is also an eigenvector. Hence  $q^* + k\Delta q$  is also a minimizer. We say that  $q^*$  is a local minimizer and an non-identifiable parameter.

It is now clear that identifiability depends on the rank of the Fisher information matrix  $S^T S$ . In implementation, because  $S^T S$  is often *nearly* singular, the smallest eigenvalues of  $S^T S$  pose problems for identifiability. To solve the problems of small eigenvalues (threshold of  $< 10^{-4}$ ), the algorithm below, described in (23), is used in determining parameters that are identifiable.

### Algorithm for identifiability

1. Form the matrix  $S^T S$  where  $S$  is the matrix of sensitivity coefficient.
2. Compute eigenvalues of  $S^T S$  and arrange them in ascending order of their absolute values:

$$|\lambda_1| < |\lambda_2| < \dots < |\lambda_m|$$

.

3. If  $|\lambda_1|$  is less than a threshold of  $10^{-4}$ , there is an unidentifiable parameter.
4. The largest component of the the eigenvector  $V_1$ , eigenvector corresponding to  $\lambda_1$ , corresponds to the least identifiable parameter. Remove the corresponding column from  $S$ .
5. Repeat step 1-4 until there are no longer eigenvalues that are smaller than the threshold.

## 1.5 Non-Linear Mixed Effect Model

Mixed-effects modelling is the most used method for the analysis of population pharmacokinetic/pharmacodynamic (PK/PD) data (24). Nonlinear mixed effects models (NLMEM) are statistical models used to analyze repeated longitudinal measured data. NLMEM for repeated measures can be thought of as a hierarchical model involving a modeling framework of both fixed-effects associated with the population parameters and random-effects incorporating uncertainty associated with variability of the data for unexplained inter-individual and intra-individual variability.

- The inter-individual variability accounts for factors that facilitate heterogeneity of subjects within the population. It is the variation in dynamic parameters across the population (25),
- The intra-individual describing the difference between the individual predicted values and the observations. It also account for errors due to measurement collection, dosing, and sampling, numerical approximation of the model equations, etc. (25)

. The intra-individual (residual: first-stage model) variability describing the difference between the individual predicted values and the observations is modelled as,

$$y_{ij} = f(\phi_i, x_{ij}) + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}(0; \sigma^2), \quad i = 1 \dots N; j = 1 \dots n_i \quad (1.25)$$

where  $y_{ij}$  is the  $j^{th}$  response of the  $i^{th}$  individual,  $f(\cdot)$  is a nonlinear observation function of an individual-specific parameter vector  $\phi_i$ ,  $x_{ij}$  is state dynamics,  $N$  is the number of individuals, and  $n_i$  is the total number of measurements for individual  $i$ . The residual error terms  $\epsilon_{ij}$  are assumed to be independently and identically distributed (iid) with mean zero and variance  $\sigma^2$ .

The state dynamics  $x$  quantify the behavior of the system (25). State dynamics are most commonly modeled using ordinary differential equations (ODEs) described below,

$$\frac{dx_i}{dt} = g(t, x_i, u_i; \phi_i), \quad x_{i0} = x_i(0, \phi_i), \quad t \geq t_0 \quad (1.26)$$

where  $x$  is an  $M$ -dimensional dependent variable vector,  $x_0$  is the initial conditions,  $g(\cdot)$  is the model,  $t$  is an independent variable,  $u$  is an exogenous input, and  $\phi_i$  is the  $O$ -dimensional parameter vector.

In the inter-individual (second-stage model) hierarchy, the model is associated with the parameters of the different individuals. It is described as,

$$\phi_i = A_{ij}\beta + B_{ij}b_i \quad (1.27)$$

where  $A_{ij}$  and  $B_{ij}$  are respective matrices for the fixed-effects vector  $\beta$  and random-effects vector  $b_i$ . The inter-individual variability is modelled by the random-effects vector  $b_i$  which consists of  $k$  zero-mean variables assumed to be i.i.d. with variance-covariance matrix  $\psi$ . The residual error terms  $\epsilon_{ij}$  and  $b_i$  are assumed independent for all  $i$  and  $j$ . The parameters in the mixed-effects model described by (1.25) and (1.27) are estimated either by maximum likelihood (ML) or by restricted maximum likelihood (REML) based on the marginal density of  $y$  (24).

We use MONOLiX ("MOdèles NON LInéaires à effets miXtes" or Non-linear mixed-effects models in English), a platform of reference for model based drug development (26). MONOLiX is an easy, fast and powerful tool for parameter estimation in non-linear mixed effect models, model diagnosis and assessment, and advanced graphical representation. It uses maximum likelihood to estimate parameters.

With the assumption of the data being continuous, a scalar outcomes  $y_{ij} \in \mathbb{R}^+$  is as-

sumed. In MONOLIX (27), this outcome is given by general model,

$$y_{ij} = f(t_{ij}, \Phi_i) + g(t_{ij}, \Phi_i, \xi) \epsilon_{ij},$$

$i = 1 \dots N$  and  $j = 1 \dots n$ ,  $\Phi_i$  is the parameter vector of the model  $f$  for individual  $i$ . In this case,  $f$  is the conditional mean and  $g$  is standard deviation of  $y_{ij}$ . The residual errors  $\epsilon_{ij} \sim \mathcal{N}(0, 1)$ . MONOLIX only considers the function  $g$  to be a function of the structural model  $f$ , i.e.  $g(t_{ij}, \Phi_i, \xi) = g(f(t_{ij}, \Phi_i), \xi)$  leading to,

$$y_{ij} = f(t_{ij}, \Phi_i) + g(f(t_{ij}, \Phi_i), \xi) \epsilon_{ij}. \quad (1.28)$$

There are several error models available in MONOLIX. We considered the **combined1** in our work. The model defined as,

$$y = f + (a + b * f^c) \epsilon \quad (1.29)$$

The function  $g$  in (1.28) is  $g = (a + b * f)$ , and the parameters  $\xi$  and  $c$  are  $\xi = (a, b)$   $c = 1$ .



## CHAPTER

## 2

# A QUANTITATIVE SYSTEMS TOXICOLOGY (QST) MODEL FOR HEPATIC STEATOSIS INDUCTION BY CARBON TETRACHLORIDE

### **2.1 Motivation**

Fatty liver disease impacts close to 25-30% of the US population. There are two main types of liver diseases: alcoholic liver disease (ALD) and nonalcoholic fatty liver disease (NAFLD). ALD is the liver disease caused by consuming excess amount of alcohol. Ethanol's toxins perturb almost all aspects of hepatic lipid metabolism (28). NAFLD, on the other hand, is the build-up of excess fats in liver cells when other causes have been ruled out (29).

Hepatic steatosis is a biological manifestation of NAFLD and is defined by an increase of more than 5% of lipid content in the liver. Hepatic steatosis can lead to liver failure and altered xenobiotic metabolism (30). If not reversed, it can progress to irreversible stages of

liver disease including fibrosis, cirrhosis, hepatocellular carcinoma, and death (31).

Considering the large percent of the population that suffers from fatty liver diseases, it is paramount to the public health to identify the causes of fatty liver diseases. It had been determined that a variety of exposures, such as high-calorie high-fat diets, inactivity, therapeutic drugs, and environmental chemicals, can induce NAFLD (31). While some of these causes may not be prevented (i.e. therapeutic drugs) (32), for the sake of public health, it is crucial to prevent hepatotoxicity and liver disease progression that can be precipitated by chemical exposure. Identification of such chemicals is important in order to minimize its use (or ban them if possible) by the public.

Carbon tetrachloride,  $\text{CCl}_4$ , is one of the chemicals that has been shown to lead to liver cirrhosis and to liver cancer following prolonged exposure to it (33). Carbon tetrachloride was the first of the five chlorinated solvents to come into general use. Production of commercial quantities in Europe began in approximately 1900 or earlier, and in the United States between 1905 and 1908. Due to the fact that acute unintentional or intentional intoxication by  $\text{CCl}_4$ , while may rarely occur, are potentially life-threatening, it is used under regulatory surveillance to ensure safety at the working place and to protect the workers' health (34). This is because  $\text{CCl}_4$  is an efficient but highly toxic solvent, used in households and commercially in the industry. Carbon tetrachloride was banned in consumer products in 1970, in aerosol products in 1978, and in grain fumigation in 1985 (35).

In humans as well as in experimental animals,  $\text{CCl}_4$  is readily absorbed by all routes of exposure (inhalation, oral and dermal). Post absorption, it is widely found in fat but also in blood, muscle, liver, brain and other tissues and organs (36).  $\text{CCl}_4$  is a manufactured chemical that does not occur naturally (37). Carbon tetrachloride is a clear liquid that evaporates very easily. Most carbon tetrachloride that escapes to the environment is therefore found as a gas. Carbon tetrachloride does not easily burn. Carbon tetrachloride has a sweet odor, and most people can begin to smell it in air when the concentration reaches 10 parts per million parts of air (ppm) (38). The primary effects of carbon tetrachloride in humans are on the liver, kidneys, and central nervous system (CNS). Human symptoms of acute (short-term) inhalation and oral exposures to carbon tetrachloride include headache, weakness, lethargy, nausea, and vomiting (39).

Our main goal for this project is to determine the amount of carbon tetrachloride it takes to raise the levels of triglycerides in the liver to a point of hepatocytes. We accomplish our goal by establishing a dose response of carbon tetrachloride (in the air or water) and the amount of triglycerides in the liver. We carry out the dose-response task by incorporating the presence of carbon tetrachloride in the human liver model from (29). Our other goal is

to use sensitivity analysis to determine the parameters that play a significant role in the production of hepatic triglycerides.

## **2.2 Mathematical Modeling**

Traditional toxicity testing approaches are time and resource intensive. The objective of dose-response between carbon tetrachloride and amount of triglycerides in the liver is established via modeling and simulations. Modeling and simulations were used due to the fact that they are highly valued in scientific discovery as they provide insights that are often impractical or impossible to discover through real-world experimental and theoretical analysis alone (40). This is on par with this work as it is unethical and inhuman to perform experiments involving carbon tetrachloride, which has been established to be harmful, on humans.

To reach the goal of estimating the amount of carbon tetrachloride it takes for humans to develop hepatic steatosis, the model follows three steps:

1. Use of PBPK model to determine amount of carbon tetrachloride in the liver,
2. the impact that the presence of  $\text{CCl}_4$  in the liver have on biological parameters, and
3. the implementation of the biological parameters and their impact due to  $\text{CCl}_4$  on the liver model.

### **2.2.1 Physiologically-based pharmacokinetic (PBPK)**

Pharmacokinetics (PK) is defined as the movement of drugs through the body. Its properties, such as, absorption, distribution, metabolism, and excretion (ADME), are studied (41). These studies are due to the substantial impact ADME can have on achieving the desired therapeutic concentration of a drug. While a particular substance's behavior are usually described for a general population using physiologically-based pharmacokinetic (PBPK) models, the difference in individuals' responses due the difference among humans (age, gender, race, weight, comorbidities, etc) can also be taken into account (42).

PBPK is a mathematical modeling approach aimed to realistically describe the physiology of a biological system. PBPK models are used to toxicologically predict the relevant

internal doses (43). They also convey any non-linear relationship between internal and external applied dose or exposure. Due to the models relevance to toxicological applications, the role of PBPK modeling have increased over the past forty years. Some of the several applications of PBPK models include: the use of PBPK modeling to test hypothesis regarding the effects of metabolic variation, the appropriate dose metric from mode of action (MOA) information, and environmental exposure from sampling of parent and/or metabolite(s) in tissue and biological fluids.

PBPK models are constructed with compartments corresponding to the different physiological organs of the body. Each compartment describes a tissue volume and blood flow rate that is specific to the species of interest (i.e: rats, humans, hamsters, etc,) tissue-partition coefficient, and permeability. Each tissue is defined with assumptions of either perfusion-rate-limited or permeability-rate-limited (44). Perfusion-rate-limited kinetics usually exists in small lipophilic molecules where the blood flow to the tissue is the limiting process of the absorption. Permeability-rate-limited kinetics takes place in more hydrophilic and larger molecules, such that the permeability across the cell membrane becomes the limiting process of absorption (44). If applicable, the chemical/drug is moved from one organ through another via blood flow among organs.

The organism various parameters are used as direct input in the model. These parameters represent the knowledge available a priori on the anatomy and physiology (45). In some cases such parameters can be measured in vivo or in vitro, but most often they are estimated from the physio-chemical properties of the drug. The parameters come into two properties, organism and drug properties.

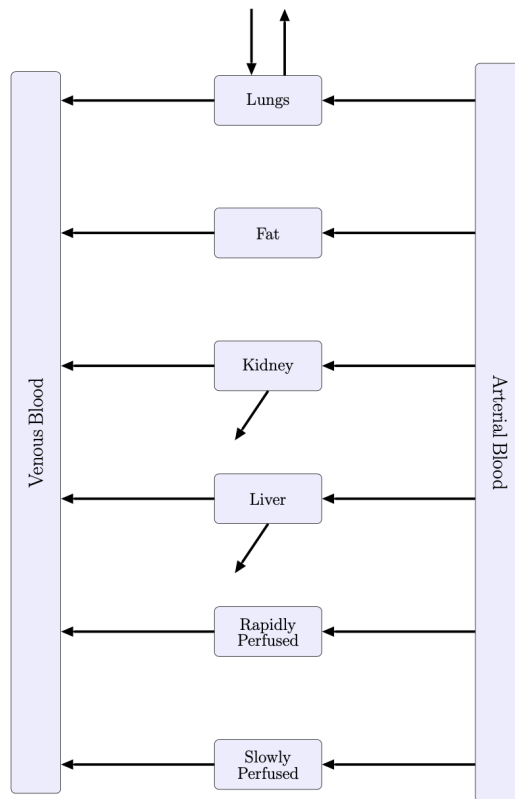


Figure 2.1: A schematic example of a PBPK model

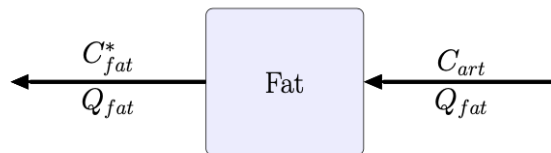


Figure 2.2: A compartment of a PBPK model

A schematic of a six compartment PBPK model is shown in Figure 2.1. In the figure, the change in the concentration for each compartment,  $X$ , is described as

$$\frac{dX}{dt} = Q * (C_{in} - C_{out}), \quad (2.1)$$

where  $X$  is the amount in the compartment,  $Q$  is the blood flow through the compartment,  $C_{in}$  and  $C_{out}$  are the concentration coming in and going out of the compartment, respectively. Each compartment can be described using this equation. For example, the fat compartment will be written as (2.2) to reflect Figure 2.2. Other details, to include ADME principles, can be incorporated in the equation. The final mathematical description of a PBPK model is a series of nonlinear coupled differential equations ((41)).

$$\frac{dX_{fat}}{dt} = Q_{fat} * (C_{art} - C_{fat}^*) \quad (2.2)$$

The organisms' parameters in PBPK modeling and relevant organs are included in the model. There are two main methods of including the organs needed for the model. One method is to start with as many organs as possible and the cutback as necessary. The other method is to start with the simplest model, with organs lumped together, and proceed to adding organs as necessary. Tissues lumped together are assumed to possess similar anatomical properties (42). We used the latter method. We started with the model composed of four compartments, fat, liver, slowly perfused and rapidly perfused. We, then, added more compartments: lungs, muscles, brain, liver, lumen, gut and kidney (Figure 2.4). Each of the organs in our model is represented by its anatomical and physiological properties; for example, volume and perfusion blood flow rates. In Figure 2.4a model, the drug/chemical is administered orally and transported along the lumen to the gut. The compounds can also be administered via other routes, inhalation (Figure 2.4b), injection (Figure 2.12), skin contact, etc. The administration route determines how a drug is distributed in the body.

The second type of input parameters into PBPK models is the drug properties. These properties are compound-specific parameters. They include the physicochemical parameters of the compound (e.g. lipophilicity, solubility, pKa (unique physicochemical property that controls its ionization state when in solution (46)), etc.) that can usually be determined in vitro. Particularly relevant for oral administration, drug solubility determines the availability for absorption of a compound in the GI tract. Figure 2.3, adapted from (45) represents a breakdown of parameters that are input in the PBPK model.

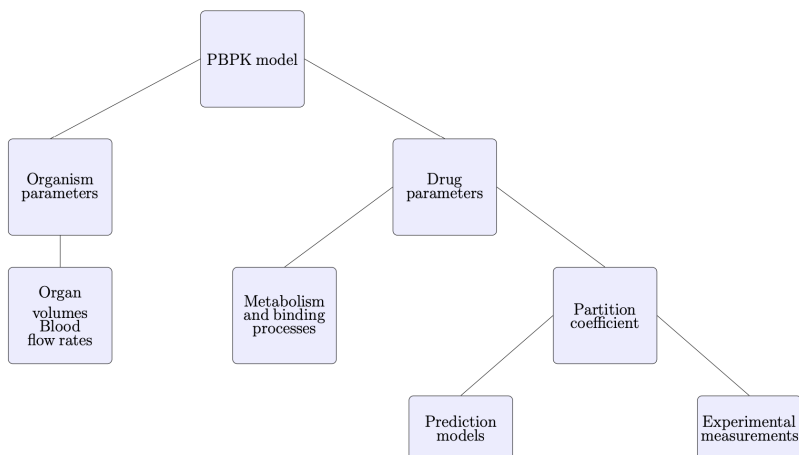


Figure 2.3: A Breakdown schematic of parameters input in the PBPK models

### 2.2.2 Ingestion and Inhalation Models

For humans, we used two types of PBPK: ingestion and inhalation. The schematic of the model are displayed in Figure 2.4. Most compartments in both the ingestion and inhalation models are represented by (2.3) where  $X_i$  is the amount in compartment  $i$ ,  $Q_i$  is blood flow through compartment  $i$ ,  $C_a$  is the incoming concentration from the arteries,  $V_i$  is the volume of compartment  $i$  and  $P_i$  is the partition coefficient between the compartment and the blood. As suggested by ingestion and inhalation, the main difference resides in the uptake of  $\text{CCl}_4$ . For the ingestion, the chemical is orally taken. It then goes to the lumen. From the lumen, the chemical distributes throughout the body as seen in Figure 2.4a and described in (2.4). For the inhalation method, on the other hand, the chemical is inhaled and goes to the lungs before distributing throughout the body as described in Figure 2.4b. The full equations of the PBPK models are found in Appendix (A.7), with corresponding acronyms in Table A.1.

$$\frac{dX_i}{dt} = Q_i * \left( C_a - \frac{X_i}{V_i * P_i} \right) \quad (2.3)$$

$$\frac{dX_{Lg_{ing}}}{dt} = Q * (C_v - C_a) \quad (2.4a)$$

$$\frac{dX_{Lg_{inh}}}{dt} = Q * (C_v - C_a) - \frac{Q_{resp} * C_a}{P_{air}} + Q_{resp} * C_{air} \quad (2.4b)$$

$$\frac{dX_{G_{ing}}}{dt} = Q_g * \left( C_a - \frac{X_{G_{ing}}}{V_g P_g} \right) + k_a * X_{Lm} \quad (2.4c)$$

$$\frac{dX_{G_{inh}}}{dt} = Q_g * \left( C_a - \frac{X_{G_{inh}}}{V_g P_g} \right) \quad (2.4d)$$

In (2.4),  $Q$  is the total blood flow from the veins (described in A.4). The first two equations, (2.4a) and (2.4b), are for the lungs, and the last two, (2.4c) and (2.4d), are for the gut compartment. (2.4a) and (2.4b) are the amount of the chemical in the lungs when the chemical is ingested and inhaled, respectively. The expression  $Q * (C_v - C_a)$  is present in both (2.4a) and (2.4b). In these equations, the amount is given by the product of the difference in concentration of the arteries,  $C_a$ , and veins,  $C_v$  (incoming and outgoing concentration, respectively) multiplied by the total blood flow,  $Q$ . (2.4b) has additional terms describing the inhalation and exhalation of the chemical.  $Q_{resp}$  is the inhalation/exhalation rate,  $P_{air}$  is the partition coefficient between blood and air and  $C_{air}$  is the concentration in the air.

The expression  $Q_g * \left( C_a - \frac{X_{G_{ing}}}{V_g P_g} \right)$  is in both (2.4c), ingestion, and (2.4d), inhalation, except for the subscript in expression  $X_{G_{inh}}$  of (2.4d) to highlight the inhalation. In the said expression, the amount is represented by product of the blood flow through the gut  $Q_g$  and the incoming concentration from the arteries minus the amount going to the veins (represented by the amount in the gut,  $X_{G_{ing}}$  divided by the product of the gut's volume,  $V_g$  and the partition coefficient between the gut and the blood,  $P_g$ ). In (2.4c), since the chemical is ingested, the amount coming from the lumen is added as well. The lumen's incoming amount is represented by the amount in the lumen,  $X_{Lm}$  multiplied by the absorption constant of the chemical,  $K_a$ .

$$\frac{dX_{Lv}}{dt} = (Q_{Lv} - Q_g) * C_a + \frac{Q_g X_g}{V_g P_g} - \frac{Q_{Lv} X_{Lv}}{V_{Lv} P_{Lv}} - \frac{\frac{\text{met}_{v_{max}} X_{Lv}}{V_{Lv} P_{Lv}}}{\text{met}_{km} + \frac{X_{Lv}}{V_{Lv} P_{Lv}}} \quad (2.5a)$$

$$\frac{dX_K}{dt} = Q_k * \left( C_a - \frac{X_k}{V_k * P_k} \right) - \frac{X_k}{V_k} * K_{Ext} \quad (2.5b)$$



Equation (2.5a) is the description of the chemical amount present in the liver compartment. This equation like others as well is represented by the difference in the amount coming in and going out. There is an additional expression  $\frac{\text{met}_{V_{max}^{X_{Lv}}}}{\text{met}_{K_m + \frac{X_{Lv}}{V_{Lv}P_{Lv}}}}$  that represent the loss of the chemical due to the liver metabolism. This equation is derived from the Michaelis-Menten kinetics and is described in (A.1.1).

Equation (2.5b) is the description of the chemical amount present in the kidney compartment. Like the liver compartment (2.5a) there is an addition expression  $\frac{X_k}{V_k} * K_{Ext}$  that represent the loss of the chemical due to kidney's metabolism.

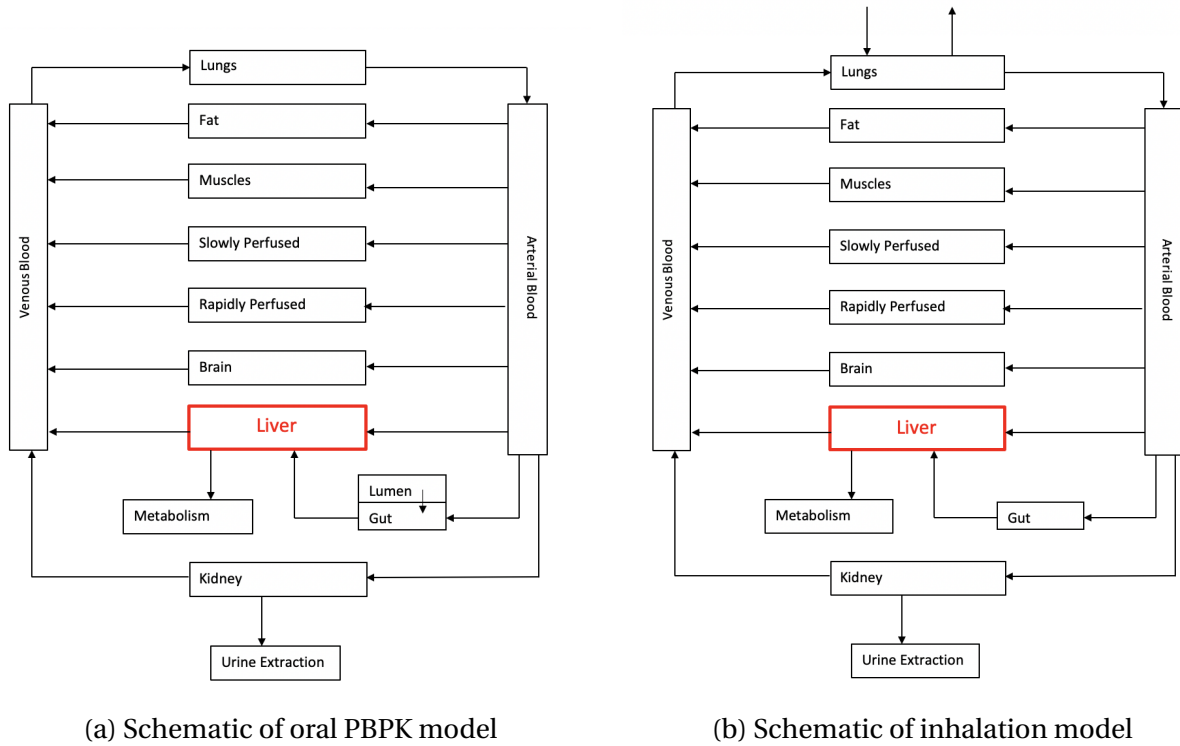


Figure 2.4: Schematic of PBPK model used in human.

### Estimation of some parameters in the PBPK model

For both of our PBPK models (Figure 2.4), almost all of the parameters (both in rats and humans) were found in literature (Table A.2, Table A.3, Table A.4 and Table A.5). However, the parameters for the absorption constant,  $k_a$ , and the Michaelis-Menten metabolite  $V_{max}$  were estimated using the rats data found in (47) along with the PBPK model for ingestion (Figure 2.4a). These two parameters were estimated, Table A.4, with the MATLAB (Version: 9.10.0.1602886 R2021a). The goal in the estimation is to find the unknown parameters that minimize the least square equation (2.6), where  $t$  represents the time-points,  $y_v$  is the venous concentration data points and  $y_e$  is the estimated venous output.

$$J(q) = \sum_{t=1}^T (y_v(t) - y_e(t))^2. \quad (2.6)$$

The simulation of  $\text{CCl}_4$  in the venous blood (using estimated parameters) was graphed on the same plot with the rats data, Figure 2.5. The estimated value of the metabolites  $V_{max}$  in the liver was assumed to be the same in the inhalation model (Figure 2.4b) as well. The estimated parameters, Table A.4, were also assumed to be similar to those in humans.

Based on the plot obtained by using the estimated parameters, Figure 2.5, the estimation was deemed adequate and satisfactory to continue with the project. Using the PBPK models (Figure 2.4) along with different initial  $\text{CCl}_4$  doses (as the PBPK models' initial conditions) in the lumen or air, the concentration of  $\text{CCl}_4$  in the liver were continuously estimated. The estimated  $\text{CCl}_4$  amount is then used to determine how certain biological processes (very low density lipoproteins and sterol-regulatory element binding protein 1C) are impacted.

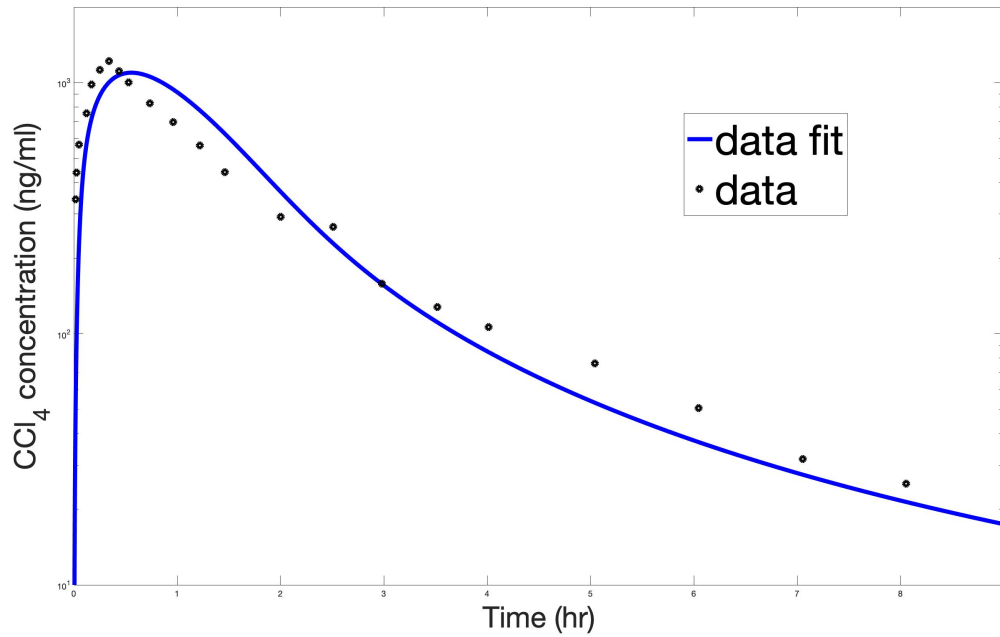


Figure 2.5: Data fit after estimating  $k_a$  and  $me t_{V_{max}}$ .

### 2.2.3 Parameter Impacted by Carbon Tetrachloride

The existence of carbon tetrachloride in the body impacts biological processes. Certain interrupted processes have an active role in the amount of triglycerides in the liver. Two parameters that are known to change when  $\text{CCl}_4$  is present are very low density lipoproteins and sterol regulatory element binding protein 1C.

#### Very Low Density Lipoproteins (VLDL)

Very low-density lipoprotein (VLDL) is a major carrier of lipids from liver to general blood circulation. The effect of carbon tetrachloride on the liver and the VLDL had been extensively researched. Changes of lipoprotein secretion and composition in response to  $\text{CCl}_4$  treatment have been established in monolayer cultures of rat primary hepatocytes (48).  $\text{CCl}_4$  impacts the secretion of very low density lipoproteins (VLDL) (49). If the VLDL values

is higher, then more triglycerides will be carried out of the liver. On the other hand, a small value of VLDL result to little TGs being carried out of the liver. The inhibition of lysosomal lipase activity and very low density lipoproteins (VLDL) secretion might be the reason for the accumulation of triglyceride in liver (50).

Using data from (51), an exponential equation describing the relationship between the concentration of  $CCl_4$  and the percentage of VLDL in the liver was determined, Figure 2.6. The exponential equation in (2.7) indicates that the value of the VLDL drops as the concentration of  $CCl_4$ ,  $x$ , in the liver increases. The equation was used to estimate the continuous levels of VLDL based on the  $CCl_4$  concentrations found in the liver from the PBPK models. To simplify things, the maximum concentration of  $CCl_4$  was used to determine the corresponding VLDL value, and this value was the only one used later on the Liver model (Figure 2.9). Before finding the  $VLDL$ , the  $CCl_4$  value in the liver was first converted to  $\mu M$  according to (2.8) in order to match the data units.

$$VLDL_{percent} = 79.79 e^{-0.09904 x} + 20.14 e^{-0.001841 x} \quad (2.7)$$

$$M = \frac{\text{Mass(mg)} * 10^{-3} \text{g}}{\text{MW (g/moles)} * 1 \text{mg} * L} = \frac{\text{moles}}{L} \quad (2.8a)$$

$$\mu M = M * 10^6 \quad (2.8b)$$

In equations (2.8), the Mass(mg) is the mass found in the rats liver after administration of  $CCl_4$  in one of the PBPK models. The mass is multiplied by  $10^{-3}$  to convert to grams in order to match the units of the molecular weight (MW) of  $CCl_4$ . The results are moles per volume, molarity (equation 2.8a)). We then multiply the results of (2.8a) by  $10^6$  to obtain (equation 2.8b)) which has units, micro-molarity ( $\mu M$ ), that match the data in Figure 2.6. Finally the VLDL amount depending on PBPK liver concentration ( $\mu M$ ) was estimated (2.7).

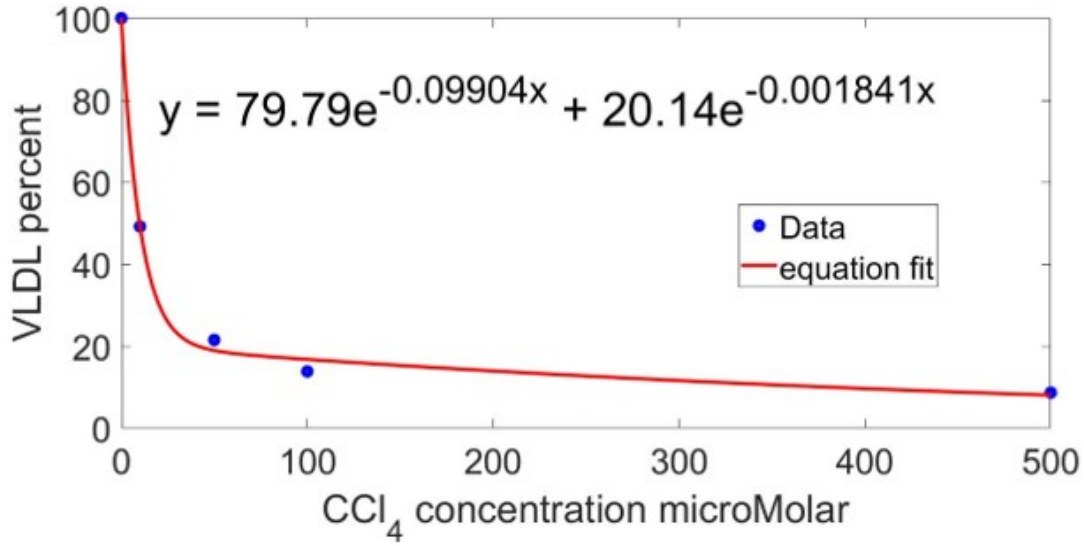


Figure 2.6: Instantaneous effect of CCl<sub>4</sub> on VLDL

### Sterol-Regulatory Element Binding Protein 1c (SREBP-1c)

Lipogenesis enclose the processes of fatty acid synthesis and subsequent triglyceride synthesis (52). The literature (29) shows that simulating insulin resistance (IR) leads to higher levels of hepatocytes TG. An additional simulation of lipogenesis, by increasing the sterol-regulatory element binding protein 1c (SREBP-1c) which has been shown to be increased by carbon tetrachloride in the body (53), leads to even higher levels of hepatocytes TG. Using the euclidean norm of the CCl<sub>4</sub> amount in the liver from the PBPK model and the data from (53), we can determine the value of sreb-1c. In (53), there is only three data points, which are not enough to establish a strong relationship. However, seeing that the last two data points, Figure 2.7, are close to each other, we devise that the CCl<sub>4</sub> and sreb-1c relationship follows some type of logistic model. In (2.9), we use a logistic model to establish the representation of CCl<sub>4</sub> vs sreb-1c. In the logistic equation (2.9),  $x$  represent the euclidean norm value of CCl<sub>4</sub>.

$$\text{sreb-1c} = \frac{7.416}{1 + e^{-0.0836(x-182.4)}} + 4 \quad (2.9)$$

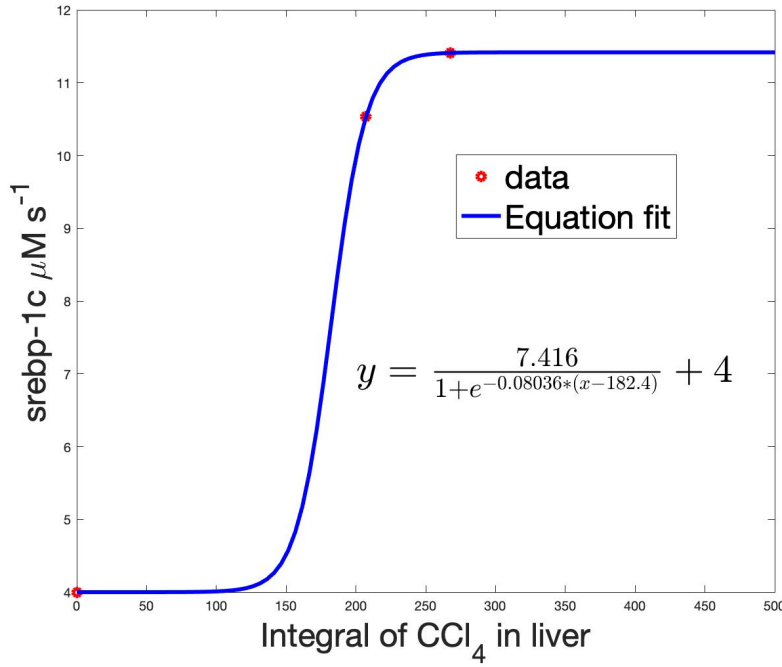


Figure 2.7: Logistic approximation of srebp-1c.

## 2.2.4 Liver model

The liver is the central organ that controls lipid homeostasis (32). The liver model used to simulate the levels of fat and triglycerides in the body is divided in two main compartments: the liver and the body. The liver compartment is further divided into eight compartments, Figure 2.8 adapted from (29). In the model, the blood flows from the periportal blood (first compartment) to the subsequent compartments of the liver until its pericentral compartment (eight and last compartment of the liver). From the distal pericentral compartment, the blood goes into the rest of the body where it will interact with other organs such as pancreas.

The liver for this computational model is divided to include zonation. The model is therefore able to represent changes in concentrations of metabolites and hormones as blood passes through the sinusoid, as well as the variation in sinusoidal hepatic enzyme expression. The position along the axis determines the blood and surrounding hepatocytes

in the sinusoid (proximal periportal > distal pericentral). This model does not include the blood flow, blood oxygenation and hormonal changes occurring during exercise, it simulates an individual at rest.

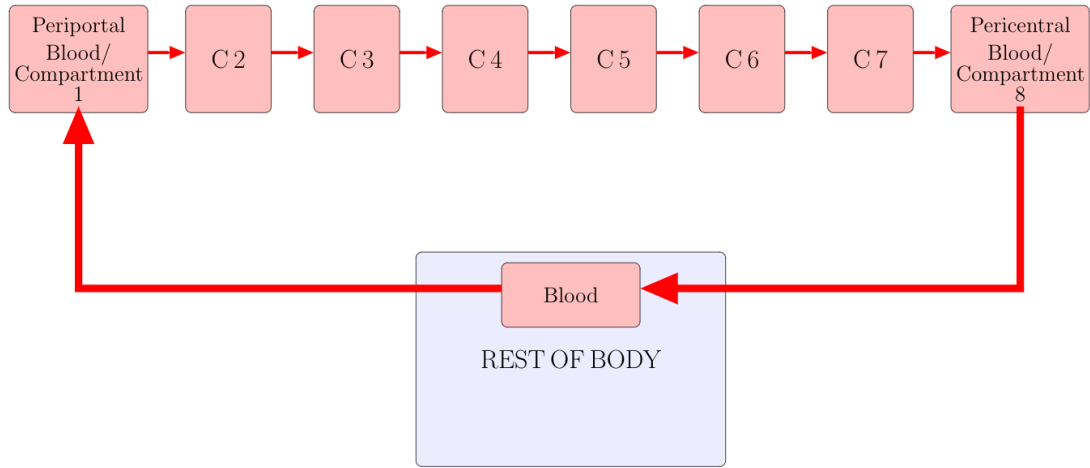


Figure 2.8: A schematic of the liver model. The porto-central axis of the sinusoid is considered to be the repeating unit of the liver.

A simplified representation of blood flow was utilized because it is compartmentalized with average concentrations of variables in each compartment. In each time step, a proportion of the blood in each compartment moves to the succeeding compartment and is replaced by blood from the previous compartment. The plasma concentration of each metabolite or hormone in each compartment changes according to the following equations:

$$\frac{dM_{i=1:n}}{dt} = bf * M_{i-1} - bf * M_i = bf * (M_{i-1} - M_i) \quad (2.10a)$$

$$\frac{dM_0}{dt} = \frac{bf * (M_n - M_0)}{s}. \quad (2.10b)$$

In equation (2.10),  $n$  is equal to eight, the total number of liver compartments. Equation (2.10a) represent the liver compartments, while equation (2.10b) is for the rest of the body.

In various plasma and hepatic metabolites in the model, the concentrations are calculated in units of  $\mu\text{M}$  ( $\mu\text{Moles/L}$ ). The concentrations of hormones are calculated in  $\text{pM}$  ( $\text{pMoles/L}$ ). Triglycerides are esters derived from three fatty acids as well as a glycerol groups. Triglycerides do not have a unique molecular formula, hence there is not fixed molecular mass. This poses problems in calculating molarity. To resolve this issue, an average molecular mass of  $\frac{807.339\text{g}}{\text{moles}}$  was used for triglyceride (this amount is equal to that of tripalmitin, derived from the fatty acid palmitic acid.)

Each liver compartment is expanded to include other biological mechanisms. The storage of glucose as glycogen, the cycling between glucose and lactate, adenosine triphosphate (ATP) production, fatty acids (FA) production and the storage of free fatty acids (FFAs) as triglycerides are the main focuses of the model. The schematic of some of the variables and processes included in each hepatic compartment of the model is displayed in Figure 2.9, adapted from (29) .

More details including a reduced description of the representation of metabolism in each hepatic compartment in the model is provided in Tables 1 and 2 of (29). Table 1 contains the differential equations for each hepatic variable in terms of the metabolic processes. Table 2 defines the metabolic processes included in the model, it also references the sections of the supplementary material in which the full equations can be found. A detailed description of all equations in the model, along with the values of each constant and references used to set them is provided in S1 Text of (29).



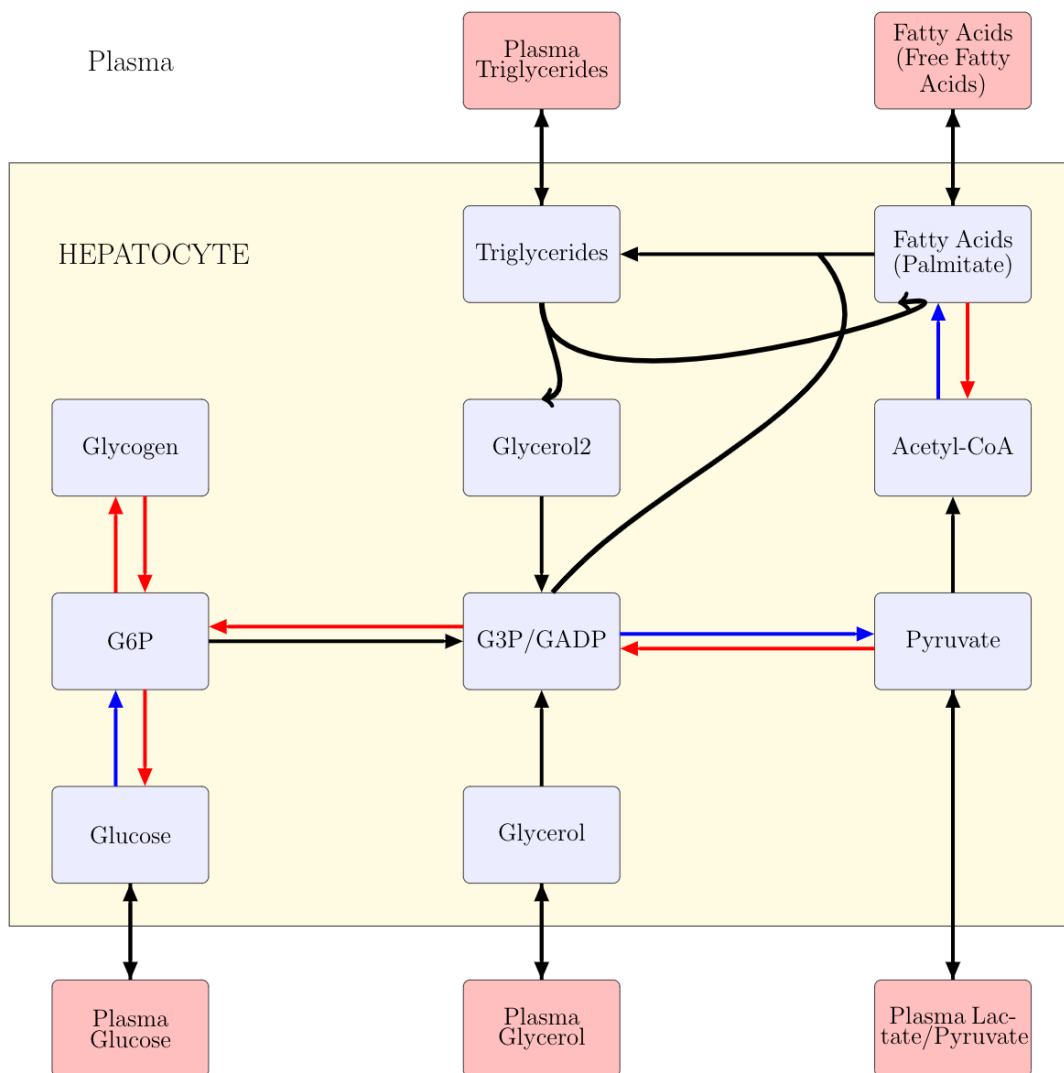


Figure 2.9: Some variables and conversions included in each hepatic compartment.

### 2.2.5 Complete model

The simulation of carbon tetrachloride and hepatocytes triglycerides dose response will follow the outline displayed in Figure 2.10. The model starts with a PBPK model (either oral or inhalation). The amount of  $\text{CCl}_4$  in the liver is taken to determine the value of VLDL, Eq (2.7), and srepb-1c, Eq (2.9). These two estimated values are plugged in the liver model. The

VLDL parameter is between Triglycerides and Plasma triglycerides (in Figure 2.9) because it carries triglycerides from the hepatocyte to the plasma. srebp-1c is between Acetyl-CoA and Fatty Acids.

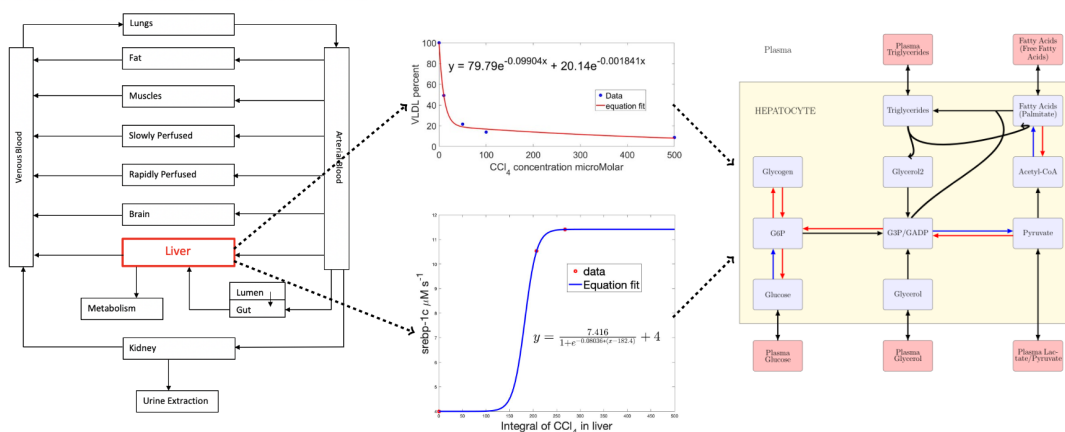


Figure 2.10:  $CCl_4$  v Triglycerides Outline of Simulation.

## 2.3 Model Validation in rats

The plan is to use the liver model and estimate the amount of  $CCl_4$  it takes for someone to develop hepatic steatosis. As outlined in 2.2.5, to accomplish the task, we will use the combination of a PBPK and liver model to estimate this amount. We will assume that  $CCl_4$  is either ingested or inhaled. However, in order to have confidence in the model and simulation, the results need to be validated. Since, we do not have experimental data in humans, we first needed to validate the outlined plan of dose response between  $CCl_4$  and hepatocyte triglycerides in rats by using existing rats data.

To validate, we changed the parameters of the models in the outline plan from humans to rats. We then used the model to simulate the results described in (53).

### 2.3.1 Liver model in rats

The liver model that we have is for humans. We wanted to make it compatible for rats first to proceed with our simulations. We kept the same model, but changed the parameters. The parameters were divided into two, parameters included in the script and parameter in the function (MATLAB was used). All the parameters in the script corresponded to the initial conditions of the molecules. These parameters were changed one at a time to ensure catching any parameters whose change in initial conditions led to big change in the output. Most parameters had negligible influence in the output. All parameters with low/lack of influence were changed immediately to either values found in the literature or by using (2.11), for those whose values we were unable to find in the literature. Glucose and triglycerides were among the initial conditions whose values were not found in the literature. However, changing the initial conditions of either one using (2.11) led to substantial change in the liver triglycerides. Because changing the initial conditions of glucose or/and triglycerides led to substantial change in hepatic triglycerides outputs, these initial conditions were estimated using least squares. (2.11) is referred to as allometric scaling, an empirical approach of changing biological parameters from one species to another (54) (55).

$$R = \text{Human parameter value} \left( \frac{\text{Rat weight}}{\text{Human weight}} \right)^{2/3} = \text{Human parameter value} \left( \frac{0.250}{70} \right)^{2/3} \quad (2.11)$$

where "R" represent the parameter values in rats while "H" represents the parameter values for humans.

Most of the parameters in the function remained unchanged with the exception of few of them: the diet, feeding cycles, rate of blood flow and rest of body to hepatic compartment ratio)]. The mentioned parameters along with the initial conditions of glucose and triglycerides were estimated by utilizing **fmincon** in MATLAB. The **fmincon** function minimizes function subject to the linear equalities  $Aeq * x = beq$  while defining a set of lower and upper bounds on the design variables in  $x$ , so that the solution is always in the range  $lb \leq x \leq ub$ .

In order to use the **fmincon** function, there need to be an existing known data of hepatocytes triglycerides to match the output of our model. We did not find known data whose units match that of the mechanistic model,  $\mu M(\frac{\text{micro moles}}{L})$ . However (56) has data with different units,  $\frac{mg}{g}$ . This posed a problem because there are different structures of

triglycerides, and they all have different molecular masses. We elected to use the same molecular weight proposed by (29). The proposed molecular weight was  $\frac{807.339g}{\text{moles}}$ . We used this molecular mass to convert  $\frac{mg}{g}$  to  $\mu M$  by using the following conversion formulas:

$$807.339g = 1\text{mole}, \quad (2.12a)$$

$$1g = \frac{1\text{mole} * 1g}{807.339g}, \quad (2.12b)$$

$$\frac{\text{mg of TG}}{\text{g of liver}} = \frac{1\text{mole} * 10^{-3}g * 10^3g}{807.339g * \text{g of liver} * 1L} = \frac{\text{moles}}{807.339L}, \quad (2.12c)$$

$$\mu M = \frac{10^6 \mu\text{moles}}{807.339L}. \quad (2.12d)$$

The parameters estimated as well as those changed by using the literature or (2.11) are displayed in Table A.6 and Table A.7. The simulation using estimated parameters is displayed in Figure 2.11. The results are not an ideal fit to the corresponding data. If we fit a trend line through the simulated dynamics, the slope seems more shallow than suggested by data. This is due to the estimated initial conditions of the hepatic triglycerides in rats as well as the parameter representing rest of body to hepatic compartment ratio. Regardless, the hepatocytes triglycerides output seemed to follow the pattern seen in humans. We decided this was satisfactory as long as the value from the data fit was within two to five fold of the data points. This is because the data was not consistent. It was the average, without the lower and upper bounds values. The average also was not of equal number. At one time point, a datum could be from only one animal, while at another time point it could be an average of more than 5 rats. Satisfied with the results from the parameters' estimation, we then proceeded to pairing the model with the PBPK model for injection in rats.

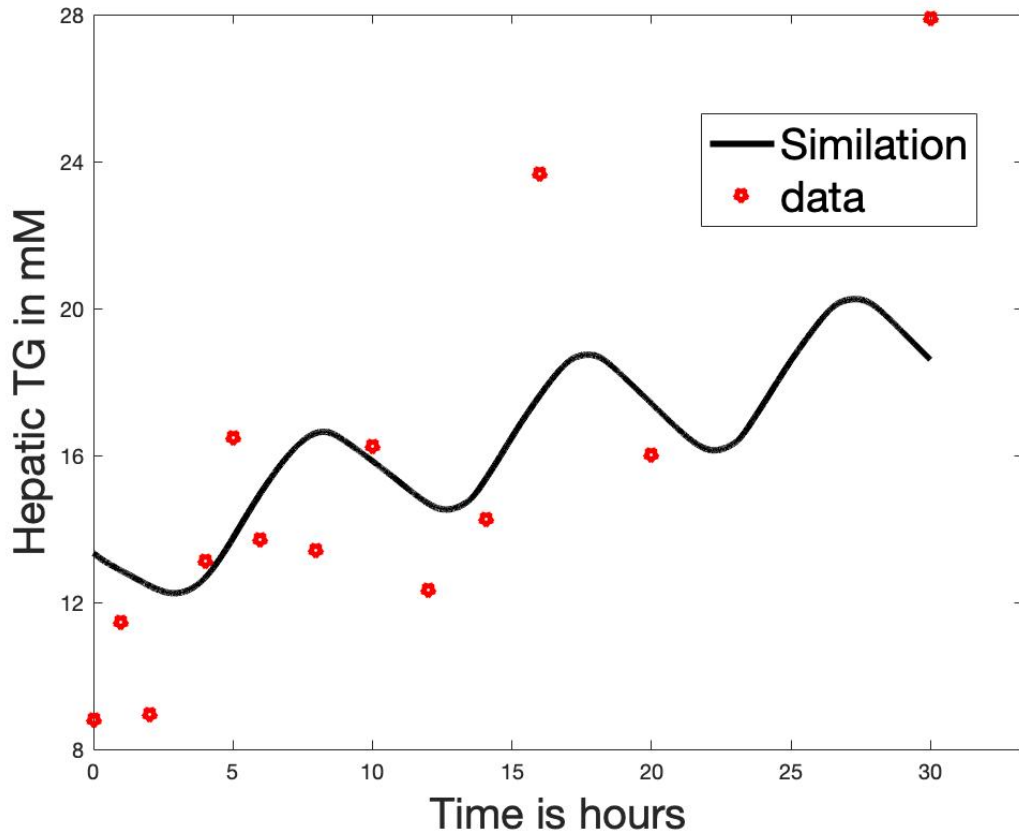


Figure 2.11: Parameter estimation for rats

### 2.3.2 Results in Rats

In (53), rats were given injections that contained  $CCl_4$ . We, therefore, simulate a PBPK model of injection. The model differ from the ingestion and inhalation in the route of administration. In this model, the  $CCl_4$  is injected in the venous blood as shown in Figure 2.12. The differential equations of the PBPK model for injection are similar to those of the inhalation and ingestion (A.7). Its lungs and gut equations are represented by (2.4a) and (2.4d), respectively. The difference lies on the concentration of the chemical in the venous blood. In the injection model, the venous compartment is defined as in (2.13), where  $V_{venous}$  is the total volume of the venous blood (57).

$$\frac{d X_{venous}}{d t} = \left( \frac{Q_B X_B}{V_B P_B} + \frac{Q_F X_F}{V_F P_F} + \frac{Q_K X_K}{V_K P_K} + \frac{Q_{Lv} X_{Lv}}{V_{Lv} P_{Lv}} + \frac{Q_{rp} X_{rp}}{V_{rp} P_{rp}} + \frac{Q_{sp} X_{sp}}{V_{sp} P_{sp}} + \frac{Q_M X_M}{V_M P_M} \right) - \frac{Q X_{venous}}{V_{venous}} \quad (2.13a)$$

$$V_{venous} = \frac{2}{3} (V_B + V_F + V_K + V_{Lv} + V_{rp} + V_{sp} + V_M + V_G + V_{lungs}) \quad (2.13b)$$

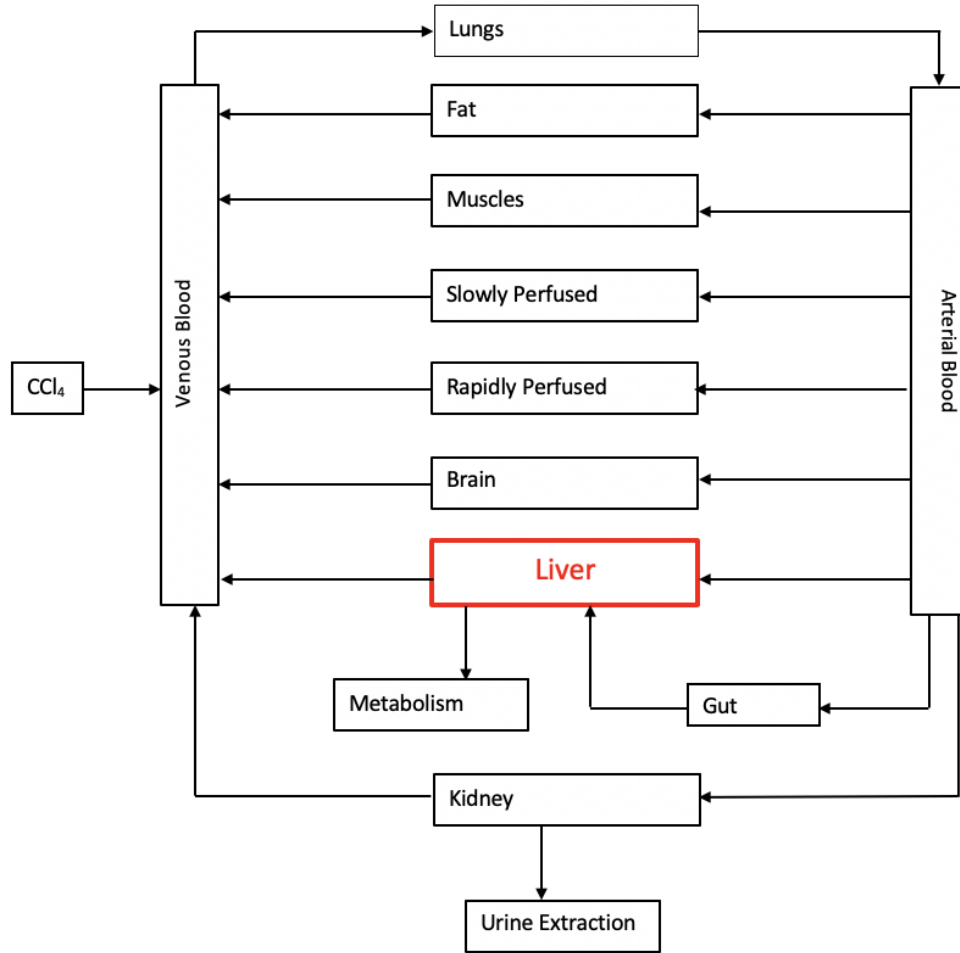


Figure 2.12: Schematic of injection PBPK model used to validate rats results.

To simulate the results in (53), we considered three scenarios as well (control, rat on high

fat (HF) diet with  $CCl_4$  for three weeks and rats on HF diet with  $CCl_4$  for five weeks). The PBPK with injection was run and the amounts of  $CCl_4$  in the liver were determined. The carbon-tetrachloride concentration was then converted to determine the percent VLDL amount, that would be later used in the liver model, using (2.8) obtained from the data in (51). In (53), we notice that high fat diet with  $CCl_4$  leads to high levels of SREBP-1c as well. The concept of high fat diet resulting in hepatocyte triglycerides increase had already been established (31).

Following the change in VLDL and SREBP-1c, we proceeded to change the diet to make it high fat. In (53), they did not mention by how they increase. We decided to increase it by 11% and looked at the results. During five and three weeks, the TGs levels respectively increased 2.78 and 2.31 folds. This is compared to the data increase of 2.52 and 2.5 in (53). We were satisfied with our results in rats to proceed to the humans. Figure 2.13 shows a plot of our results (blue part) and the reference (orange part). The vertical lines in the orange part of the reference represents the deviation from the mean.

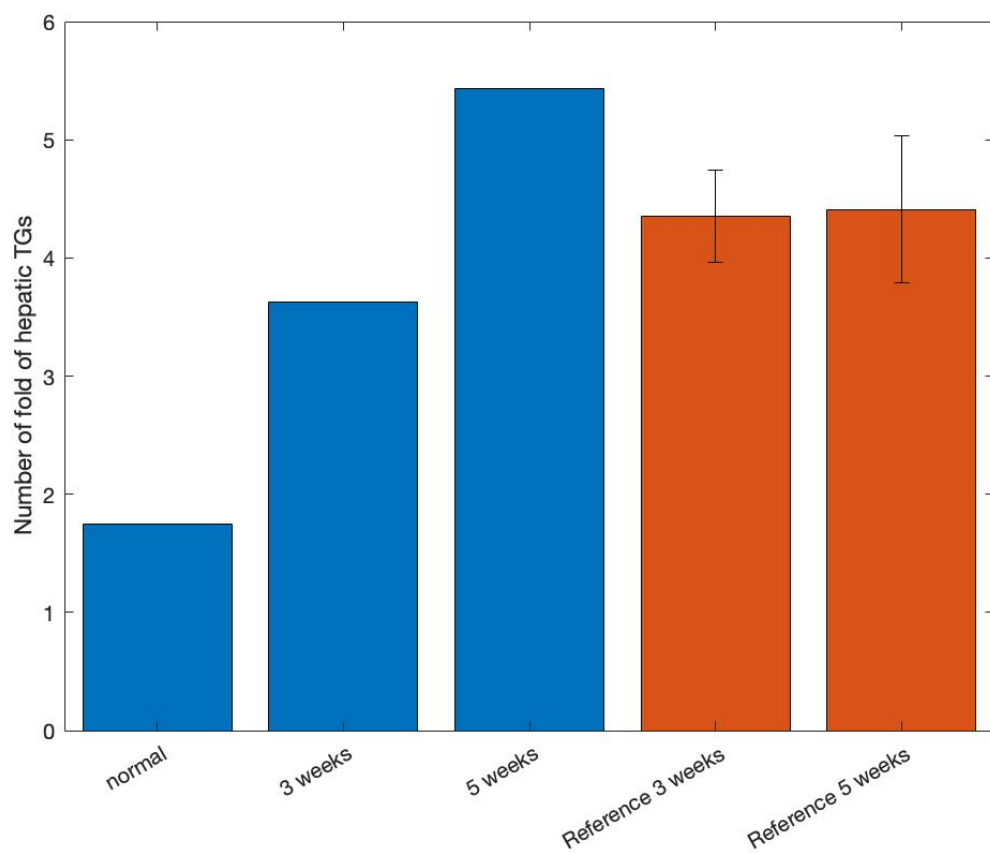


Figure 2.13: Plot of the number concentrations of hepatocytes triglycerides in rats in the presence of  $\text{CCl}_4$  and high fat diet.



## CHAPTER

### 3

## QST RESULTS

### 3.1 Simulation in the absence of carbon tetrachloride

In this section, we simulate the liver model in human under normal circumstances (no  $CCl_4$  chemical present in the body). We look at the impacts of diet, carbohydrates, fat and their combinations (Figure 3.1). In the absence of food, the simulation stops at approximately nine hours. This is probably due to the fact that the body needs sustenance to keep functioning. The absence of either carbohydrates or fatty acids leads to decreased levels of hepatocyte triglycerides. However, the decrease is sizeable in the absence of fatty acids (green line in Figure 3.1), implying that fatty acids contributes more to hepatocyte triglycerides. This could be due to the fact that 1g of carbohydrates correspond to 4 calories while 1 g of fat is 9 calories (58).

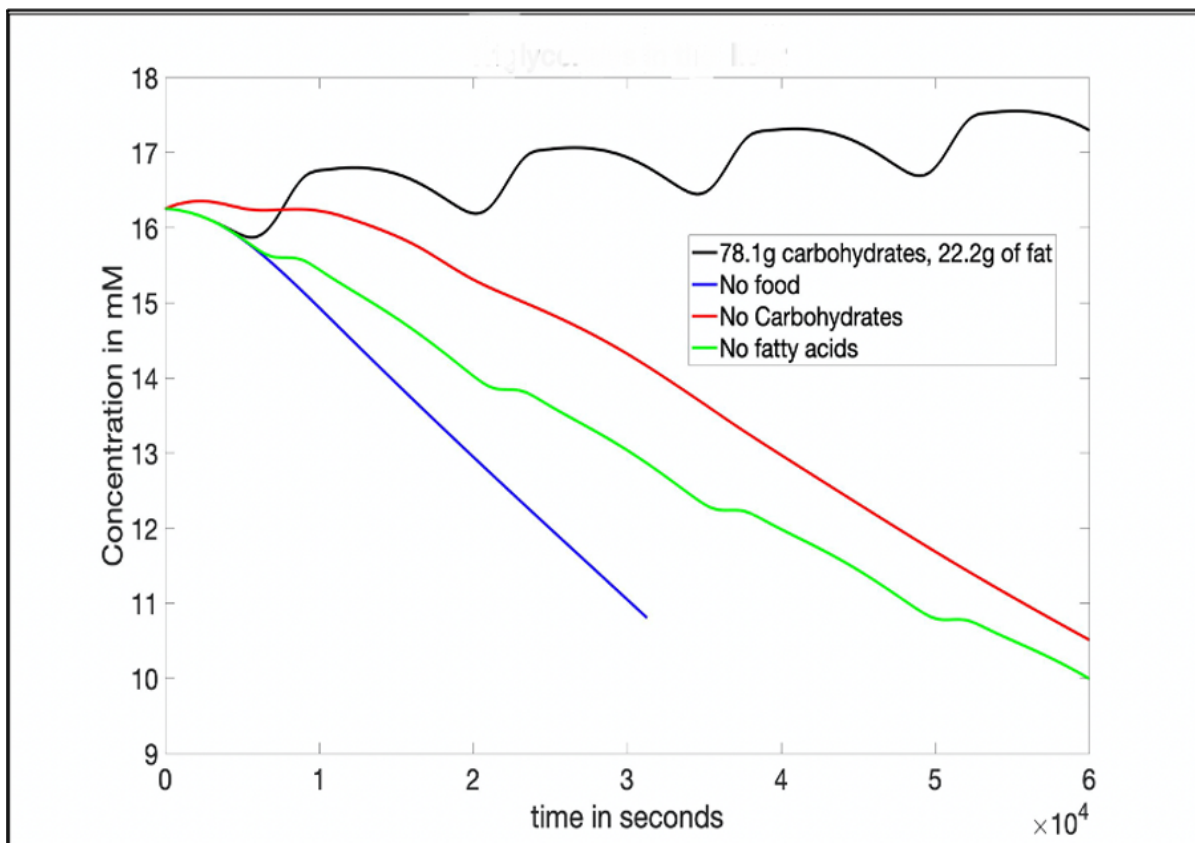


Figure 3.1: simulation of TGs under normal circumstances with or without food in humans.

### 3.2 Ingestion or Inhalation of carbon tetrachloride and the hepatocytes impact

Having validated the proposed outline for dose response (Figure 2.10), we proceeded to doing simulations in humans.

### 3.2.1 Carbon tetrachloride v. hepatic triglyceride in the absence of high fat diet

We start with the absence of high fat diet (Figure 3.2). In the model, it takes about 70 ppm of  $\text{CCl}_4$  in water for someone to develop hepatic steatosis over a period of 30 days. However, during the same period of 30 days, it takes about twenty times less for someone to develop hepatic steatosis when the chemical is inhaled. This is because with inhalation, people are constantly breathing in the chemical that even small amounts add up quickly. We note that the value inhaled dose 2.95394 ppm is less than the amount that can be smelled in the air, at least 10 ppm. This can be a problem because people could be unaware.

For the ingestion, we assumed that a person drink two liters of water per day. If someone were to drink water by following the recommendation of 2.7 liters per day for women and 3.7 liters per day for man (59), it will then take a smaller value of  $\text{CCl}_4$  ppm. We did not take into account that the chemical is digested via other means such as: drinks (coffee or tea, etc.) and food (prepared using the contaminated water).

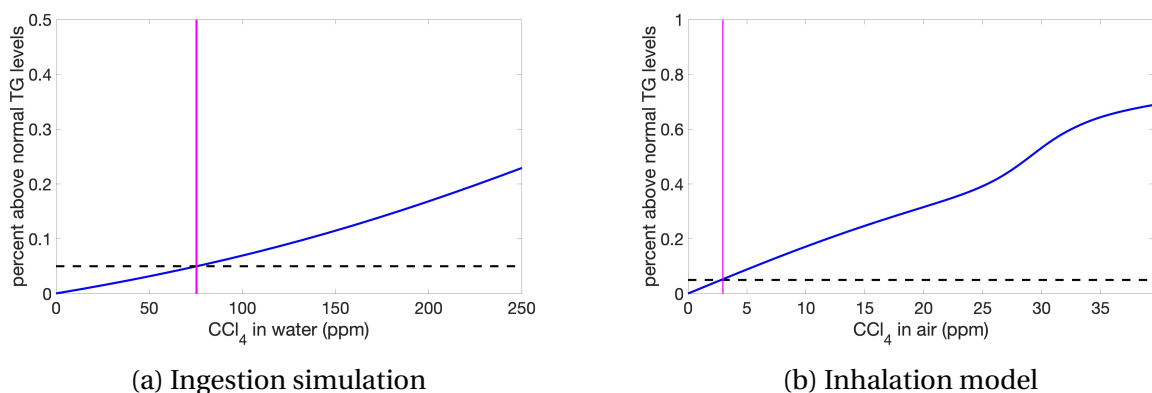


Figure 3.2:  $\text{CCl}_4$  vs. Hepatic triglycerides simulation

### 3.2.2 Carbon tetrachloride v. hepatic triglyceride in the presence of high fat diet

We also looked at the dose-response of carbon tetrachloride and hepatic triglyceride when someone is consuming a diet that is high in fat (11 % more). The results are in Figure

3.3. In the results, we see that over a period of 30 days with 0 ppm of  $\text{CCl}_4$ , the euclidean norm amount of hepatic triglycerides already exceed the normal one by 60%. This correctly support the literature that a high fat diet will lead to hepatic steatosis.

In the presence of carbon tetrachloride, the values of triglycerides get even higher. We note that these triglycerides values will eventually level off to a horizontal asymptotes.

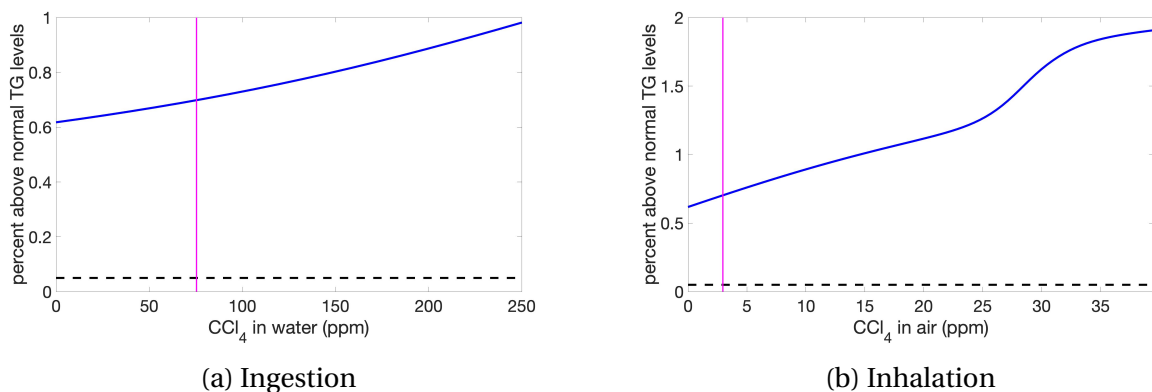


Figure 3.3:  $\text{CCl}_4$  and high fat diet vs. Hepatic triglycerides simulation

### 3.2.3 More carbon tetrachloride v. hepatic triglyceride results

From previous analysis and results, we know that a combination of VLDL decrease and srebp-1c increase lead to an increase in hepatic triglycerides. We mentioned along Figure 2.10 that the srebp-1c impacts takes place between acetyl-CoA and free fatty acids in the hepatocytes. This is due to the suggestion that FFA induce lipid accumulation by up-regulating SREBP-1c expression through the suppression of  $\text{PPAR}\alpha$  (60).

It is known that the elevation of exogenous free fatty acid level leads to insulin resistance (IR) in liver (61). The inhibition of SREBP-1c can be effective in improving IR (62). This appearance of indirect relation of srebp-1c and IR intimate the possibility of  $\text{CCl}_4$  causing insulin resistance. We simulated a decrease in VLDL, increase in srebp-1c, IR and their combination vs hepatic triglycerides, Figure 3.4.

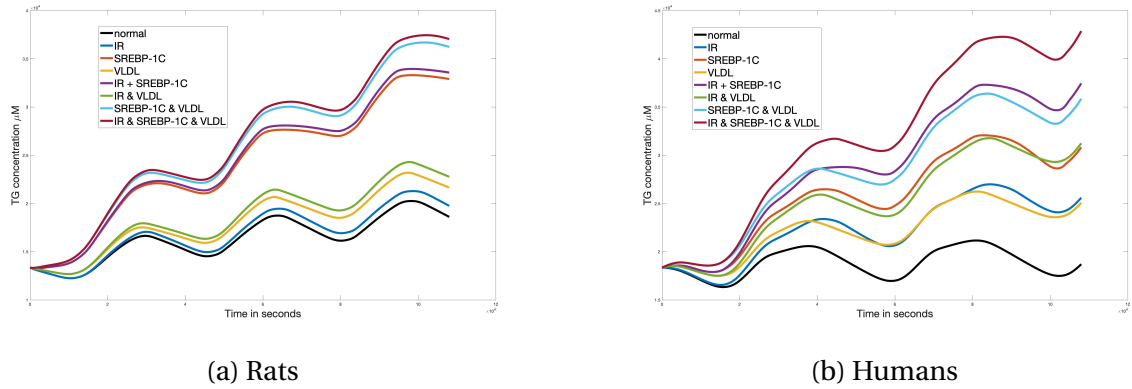


Figure 3.4: Plot of the concentrations of hepatocytes triglycerides in rats and humans.

Simulations shown in Figure 3.4 indicate that changing some parameters VLDL as well IR and SREBP-1c do indeed lead to a rise in the values of hepatocyte triglycerides. In both rats and humans, the largest deviation from normal hepatic triglycerides is during a combination of low VLDL value, IR and increase in srebp-1c.

The results are theoretical (because the parameters were changed without logical reason pertaining to the biochemistry of the body due to the lack of IR data when  $\text{CCl}_4$  is in the body). These simulations were done to confirm that adding IR to the model does indeed results in the rise of the hepatocyte triglycerides in the output.

The trend seen across the liver triglycerides in Figure 3.4 can also be seen across each of the eight liver compartments, Figure A.1.

### 3.3 Liver Model in the presence of Ethanol

As mentioned, steatosis can be caused by alcohol consumption as well. Alcohol is highly prevalent in most societies and more than 50% of Americans consume alcohol at least once a month, and medical treatment related to alcohol cost more than \$166 billions annually (28). With this knowledge, we wanted to see whether the alteration of the liver model, we used, exhibit what is known. Like  $\text{CCl}_4$ , ethanol (the chemical in alcohol) is known to decrease VLDL while increasing srebp-1c (28) (63) (64) (65).

Using the data in Figure 2 of (66), we used MATLAB optimization to increase srebp-1c while decreasing VLDL (the results are in Figure 3.5a). Using the results from the optimiza-

tion, we simulated how triglycerides in the liver will behave, Figure 3.5. We observe an increase of hepatic triglycerides over ten hours of alcohol consumption. The results are not surprising, but further validate our results of the carbon tetrachloride quantitative systems toxicology.

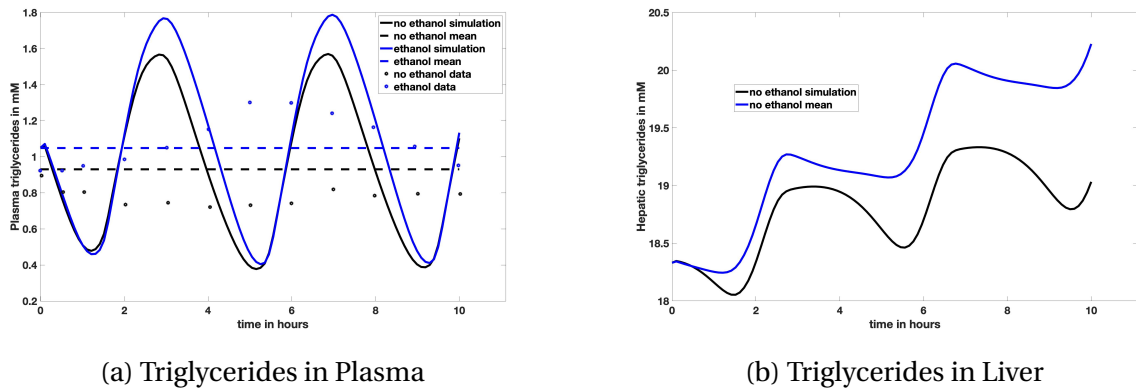


Figure 3.5: Triglycerides in. body as a response to ethanol consumption..

### 3.4 Sensitivity analysis of some parameters

Angrish et al. (31) provide a review of the biological process involved in hepatic lipid maintenance. The complex biological mechanisms that disrupt lipid homeostasis in hepatocytes is described by four apical key events central to hepatic lipid retention: hepatic fatty acid (FA) uptake, de novo FA and lipid synthesis, FA oxidation, and lipid efflux. We are interested in which process play a significant role. The global sensitivity analysis was performed for parameters that are linked to these four atypical events, Table 3.1 has the results.

1. **Lipogenesis:** Lipogenesis enclose the processes of fatty acid synthesis and subsequent triglyceride synthesis (52). It takes place in the adipose tissue, but also in the liver, muscle, heart and pancreas (67). In this work, we refer to lipogenesis as srebp-1c, because SREBP-1c has been implied to be a major factor in up-regulating lipogenic genes in response to carbohydrate feeding (68).

2. VLDL: As mentioned in 3.2.1, VLDL carries lipids from liver to general blood circulation. In the model there are two types of VLDL. One, we named it hepatic VLDL, that carries triglycerides from liver to plasma. The other VLDL, we named it plasma VLDL, is for carrying triglycerides from the plasma to the liver (see Figure 2.9).
3.  $\beta$ -oxidation: Beta-oxidation is a significant source of metabolic energy in the periods between meals or during high energy demand states, such as exercise (69). Fatty acid  $\beta$ -oxidation is the breakdown of fatty acids molecules to form acetyl-CoA.
4. Fatty Acids Transport: This is the uptake/output of hepatic fatty acids and free fatty acids. Fatty acids (FA) act as intermediates in lipid metabolism and as fuel for cells (70). Fatty acid transporters play a function in insulin resistance (71) (IR can lead to increased hepatic triglycerides Figure 3.4a)
5. Triglycerides Synthesis: In the model triglycerides are synthesized from 3 fatty acids. During the synthesis, glyceraldehyde-3-phosphate also contribute to the accumulation of TGs ((29) (Figure 2.9). The fatty acids present in TGs are mainly saturated (72).
6. Lipolysis: Lipolysis is the biochemical pathway responsible for the breakdown of complex triglycerides molecules stored in cellular lipid droplets (73). Lipolysis rates are regulated through hormonal and biochemical signals (74). Adipose tissue lipolysis is the process leading to the breakdown of triglycerides stored in fat cells and release of fatty acids and glycerol (75)
7. Insulin resistance: Insulin is a peptide hormone secreted by the  $\beta$  cells of the pancreatic islets (76). It maintains normal blood glucose levels by facilitating cellular glucose uptake, regulating carbohydrate, lipid and protein metabolism . Insulin resistance is an impaired biologic response to insulin stimulation of target tissues such as: the liver, muscle, and adipose tissue (77).

In the reference paper (29), sensitivity analysis was performed on the rate constants to identify the hepatic processes most likely to account for differences in susceptibility to steatosis. They performed the sensitivity by changing the baseline parameter constants,  $\mathbf{v}_b$ . These parameters were either increased or decreased by 10% of their reference values. The changes in the outputs (Cellular Fatty Acid ( $\mu M$ ), cellular Triglyceride ( $\mu M$ ), plasma FFA ( $\mu M$ ), plasma Triglyceride ( $\mu M$ )) were recorded. The difference between the analysis they

did and ours is in the method and the manner that we changed the parameters. In our case we use complex step method and performed global sensitivity. The complex step method slightly perturbs the parameters. In the global sensitivity, we changed each parameter more than once, 1000 different values to be exact. We assumed that the parameters values in the reference were means from normal distribution with standard deviations which were assumed to be a tenth of their respective means ( $\sigma = 0.1\mu$ ). The use of global sensitivity ensures that the sampling of the parameters reflect that of the population, as the parameters are likely to vary due to diversity in the population.

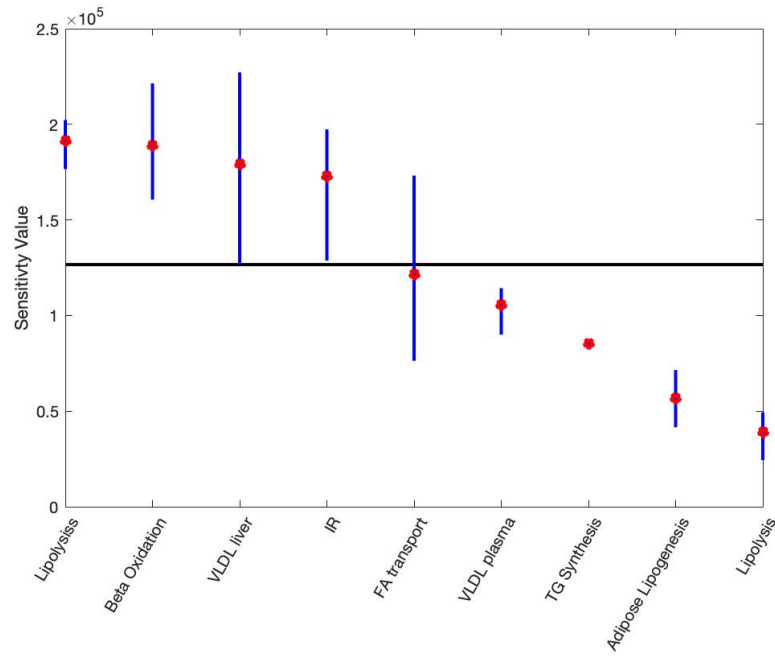


Figure 3.6: Plot of global sensitivity of some liver parameters

In Figure 3.6, the black line is the mean of all the values of the global sensitivity. The parameters above the black line are considered to be more sensitive, while the ones below the line are less sensitive. In this case, we consider these parameters: lipogenesis, beta oxidation, VLDL in the liver and insulin resistance contribute more to the production of triglycerides in liver.



Table 3.1: Global Sensitivity values

Parameters	Minimum	Average	Maximum
FA transport	$7.6313 E^4$	$1.2151 E^5$	$1.7321 E^5$
Insulin Resistance	$1.2875 E^5$	$1.2151 E^5$	$1.9735 E^5$
Lipogenesis	$4.1540 E^4$	$5.6449 E^4$	$7.1485 E^4$
TG Synthesis	$8.4274 E^4$	$8.5395 E^4$	$8.6921 E^4$
Beta Oxidation	$1.6060 E^5$	$1.8870 E^5$	$2.2139 E^5$
VLDL Plasma	$9.0023 E^4$	$1.0561 E^5$	$1.1430 E^5$
VLDL liver	$1.2676 E^5$	$1.7926 E^5$	$2.2716 E^5$
Lipolysis	$2.4374 E^4$	$3.9003 E^4$	$4.9367 E^4$
Adipose Lipolysis	$1.7660 E^5$	$1.9E9 E^5$	$2.0224 E^5$

### 3.5 Our contributions

We had one main novel contribution to our quantitative systems toxicology (QST) model for hepatic steatosis induction by carbon tetrachloride. Our novel contribution was in linking a PBPK model to the liver model via VLDL and srebp-1c, Figure 2.10. This approach can be used for other chemicals as well.

We had two other contributions. The first one was the observation that fatty acids contribute more than carbohydrates to the hepatic triglycerides, Figure 3.1, via simulation. The other contribution was in determining the sensitive parameters in the production of

triglycerides: Adipose Lipolysis, beta oxidation, VLDL in the liver and insulin resistance are the more sensitive processes.

Along with our collaborators, We are preparing two manuscripts, one to accompany these sensitivity analysis results, and the other on the QST model for hepatic steatosis induction by carbon tetrachloride.

## CHAPTER

### 4

# ANALYSIS OF IMMUNOGLOBULIN G PHARMACOKINETICS AND PHARMACODYNAMICS (PK/PD) IN INFANT RHESUS MACAQUES

Acquired immune deficiency syndrome (AIDS) was identified as a disease in 1981, and human immunodeficiency virus (HIV) was isolated as the cause of the disease (78). Since the discovery, the HIV pandemic has become part of the contemporary global landscape (79). It affects more than 1 million individuals in the United States (80) and almost 40 million HIV-infected individuals worldwide (81). In 2015, the World Health Organization estimated that 2.1 million new HIV infections were acquired (82).

The progression of HIV infection in untreated patients varies widely (83). The consequence of the disease contains three stages: an initial acute infection, a long asymptomatic period, and a final increase in viral load with simultaneous collapse in healthy CD4+T cell counts (84). HIV is chronic, but it can be treated with active combination antiretroviral

therapy (ART) with protease inhibitors, nucleoside-analogue and non-nucleoside reverse transcriptase (RT) inhibitors (85). Successes of antiretroviral therapy have reduced HIV to a chronic condition and made the progression to AIDS rare. Potent ART drug combinations allows a marked and sustained control of viral replication (86) (87). Studies have found that 3-drug regimen has led to a 60% to 80% decline in rates of AIDs, hospitalization, and death (88). In spite of the effort made by the scientific community, current treatments are unable to eradicate HIV from infected individuals, therapy must be lifelong, with the possibility of known and unknown side effects for short and long-term (89). Prolonged ART treatment has been associated with drug-induced toxicity, emergence of drug-resistant viral strains, etc. (90). To help manage the toxic effects of therapy and treatment failure, interruptions of ART treatment are being used (91). This is in spite of the fact treatment interruption may lead to a significant decline in CD4 T lymphocytes count and a rapid rebound in plasma viraemia (92). Treatment interruption has also been associated with the reemergence and predominance of a more sensitive viral population in patients with multi-drug resistant HIV.

These issues make it crucial for a need of a functional HIV cure, aiming at virological control, instead of the ART treatment for life. Passive immunization with broadly neutralizing antibodies of HIV-1 early in infection was found recently to lead to long- term control in a majority of Simian-Human Immunodeficiency Virus (SHIV) infected macaques, suggesting that HIV-1 remission may be more widely achievable (93).

Despite significant improvements, HIV/AIDS is a leading cause of infant mortality in sub-Saharan Africa where over one hundred thousands infants are infected with HIV annually (94). ART options are scarce and fewer for infants due to the formulation and pharmacokinetic limitations and the risk of resistance after exposure to drugs used to prevent mother-to-child transmission (95). One critical issue that the medical community was facing was the importance of the timely initiation of ART in HIV-infected infants. The World Health Organization (WHO) pediatric ART guidelines recommended that ART initiation in HIV-infected infants had to be deferred until specified clinical or immunologic criteria were met (96). This decision was because the majority of children who achieve durable suppression of HIV-1 replication early (before three months) lack persistent HIV-1 immunoglobulin G (IgG) antibodies (97). Early ART rapidly removes the antigenic stimulation needed to develop and sustain an HIV-1-specific antibody response (98).

This decision was however unsatisfactory to the Children with HIV Early Antiretroviral (CHER) since more than a third of infected and untreated infants die during infancy and about 50% by 2 years of age (95). Furthermore, infants who started ART at 6–12 weeks of

age, before meeting the WHO criteria, had a 76% lower mortality rate compared to those whom ART was deferred (96). In 2006, WHO guidelines recommended immediate initiation of ART in all infants who are identified to be HIV+ (96). The issue of balancing the quality of life and maintenance of immunology in children is of importance given that globally there are over 2 million children who are infected with HIV (99).

Our project has two goals:

1. Compare the pharmacokinetics of the antibodies in data obtained by starting the administration of ART at different times,
2. Devise a model that predicts how antibodies responses develop, especially in infants rhesus macaques given a SHIV infection.

Our objectives will be accomplished by using three sets of longitudinal data, whose observations are viral load and antibodies. The data consists of SHIV-infected infants rhesus macaques, that were orally transmitted using breast milk . Each data set has 10 different subjects. The difference in the data groups is the time of antiretroviral therapy initiation. The first group, referred to as early ART, consists of ten animals that were initiated on ART four days post-infection. The second group, referred to as intermediate ART, consists of animals that were initiated on therapy two weeks post-infection. And the last group, or late ART, is of animals whose ART initiation was eight weeks post-infection. Each of these rhesus macaques were given a triple daily ART regiment of tenofovir  $\left(20 \frac{\text{mg}}{\text{kg}}\right)$ , emtricitabine  $\left(30 \frac{\text{mg}}{\text{kg}}\right)$ , and dolutegravir  $\left(2.5 \frac{\text{mg}}{\text{kg}}\right)$ . The subjects were treated with a ART until they were 60 weeks old before interruption. They were kept for 24 more weeks post ART interruption, to analyse viral load rebound, before necropsy.

## CHAPTER

# 5

## PHARMACOKINETIC (PK) ANALYSIS OF GP41 AND GP120 ANTIBODIES

We are interested in comparing the pharmacokinetics of the antibodies in data obtained by starting the administration of ART at different times. Our goal is to draw conclusion on whether certain PK metrics are similar or different.

### 5.1 Non-Compartment Analysis

Non-compartmental analysis (NCA) is a commonly used technique of pharmacokinetic data analysis (100). The NCA method has few underlying assumptions compared to the model-based approaches, and it can be readily be automated (101). The method practically relies on algebraic equations to estimate PK parameters (102). NCA normally applies to first order (linear) models. It, however, is often utilized in assessment of pharmacokinetics' drug, which are nonlinear, when multiple doses levels are administered. The NCA can determine pharmacokinetics parameters such as  $T_{\max}$ ,  $C_{\max}$  last, area under the curve (AUC), half-life  $\lambda_z$ , etc.

The PK metrics that are of interest to our project are: Cmax, Tmax, AUC, and the terminal half-life.

- Cmax is the maximum observed concentration. Cmax is usually utilized to glean insight into absorption rate of a drug (103). In case of HIV, the maximum observed concentration coupled with the time of ART initiation can inform us whether or not the initiation time is prudent.
- Tmax is the time of the observed maximum concentration. This time, given what we know about mortality in children not given treatment, is crucial in determining whether or not waiting for ART initiation could be harmful to children who are HIV+.
- Terminal half-life is the time required to divide the plasma concentration by two after reaching pseudo-equilibrium. Understanding the concept of half-life is useful for the steady-state concentrations for any specific drug (104), or antibodies in our case.
- AUC last refers to the area under the concentration curve from the beginning of the interval to the last concentration observed. By analyzing AUC over time rather than individual concentration measurements, a more accurate estimate of the overall antibodies is obtained (105).

### 5.1.1 Area under the curve

The area under the curve (AUC) is a measure of exposure in toxicological studies. NCA has no need of assumption of a specific compartment model for a drug or a metabolite. It uses the *trapezoidal rule* or the *log-linear trapezoidal rule* to measure the plasma concentration curve (101).

The trapezoidal rule uses the formula learned from calculus for estimating the area under the curve,

$$AUC_0^{t_{last}} = \frac{\Delta t}{2} \sum_{i=1}^n (C_i + C_{i+1}). \quad (5.1)$$

The magnitude of the errors depends on the size of the width of the trapezoid,  $\Delta t$ . It also depends on the curvatures of the profile as well. The trapezoidal method tends to over-estimate the area during the descending phase under the assumption of the elimination being first-order. When the size of  $\Delta t$  is large in relation to the half-life, the errors associated with the over/underestimation are even more pronounced. To avoid the over-estimation

of the linear-trapezoidal method, the log-linear trapezoidal method is sometimes used. The log-linear trapezoidal method rely on the exponential decline of the concentrations as detailed below,

$$C_{i+1} = C_i e^{-K(t_{i+1}-t_i)} = C_i \cdot e^{-K\Delta t}, \quad (5.2a)$$

$$K = \frac{\ln(C_i/C_{i+1})}{\Delta t}, \quad (5.2b)$$

$$AUC_i^{i+1} = \frac{C_i - C_{i+1}}{\ln(C_i/C_{i+1})} \Delta t, \quad (5.2c)$$

$$AUC_0^{t_n} = \sum_{i=1}^n \frac{C_i - C_{i+1}}{\ln(C_i/C_{i+1})} \Delta t. \quad (5.2d)$$

In equation (5.2),  $\Delta t$  is the time step,  $K$  is the elimination rate and  $AUC_i^{i+1}$  is the AUC within a time interval given by the difference between the concentrations divided by the slope  $K$ . In practice, a mixture of both linear and log-linear trapezoidal methods is used. The linear trapezoidal is used in the ascending or constant concentrations while the log-linear trapezoidal is used when concentrations are ascending.



Table 5.1: Area under the curve for intermediate and Late data

Late Data			Intermediate Data		
Animal ID	gp41	gp120	Animal ID	gp41	gp120
RIIm19	$1.62 E^5$	$3.03 E^5$	RFz19	$1.040 E^3$	$1.09 E^3$
RJm20	$1.20 E^5$	$9.52 E^5$	RKa20	$3.42 E^4$	$4.38 E^3$
RKg19	$1.17 E^6$	$1.87 E^6$	ROz19	$6.43 E^3$	$3.41 E^3$
RLg19	$2.22 E^5$	$2.26 E^6$	RQz19	$1.96 E^4$	$5,13 E^4$
RQc19	$1.28 E^5$	$2.02 E^6$	RZm20	$8.49 E^4$	$2.00 E^5$
RRm19	$4.02 E^5$	$2.16 E^6$	RGg20	$2.47 E^4$	$3.22 E^4$
RTp19	$9.14 E^5$	$2.76 E^6$	RYy19	$1.79 E^3$	
RVh19	$5.41 E^5$	$2.19 E^6$	RCa20	$5.88 E^4$	$4.46 E^4$
RVr19	$3.10 E^5$	$2.21 E^6$	RJa20	$2.47 E^4$	$4.84 E^4$
RWc19	$3.45 E^5$	$1.00 E^6$	RUy19	$8.02 E^3$	

NCA was performed to estimate the AUC of the gp41 and gp120 antibodies in both the late ART and intermediate ART, Table 5.1. The AUC values appear larger in the late ART compared to intermediate. We observe that in the intermediate ART, the AUC for subjects RYy19 and RUy19 are empty. This is because all the data corresponding to these subjects are censored, the antibodies never developed to the level that is detected by the instrument.

### 5.1.2 Maximum Concentration

We are also interested in the maximum concentration of the antibodies. This is the maximum concentration reached by each animal in the longitudinal data, Table 5.2. The results for subjects RYy19 and RUy19 in the intermediate data are again empty, because all data points are censored.

Table 5.2: Maximum concentration for Intermediate and Late data

Late Data			Intermediate Data		
Animal ID	gp41	gp120	Animal ID	gp41	gp120
RIm19	$8.50 E^3$	$3.48 E^4$	RFz19	$1.60 E^2$	$2.62 E^2$
RJm20	$1.40 E^4$	$3.08 E^5$	RKa20	$4.71 E^3$	$4.07 E^2$
RKg19	$1.31 E^5$	$1.61 E^5$	ROz19	$3.69 E^2$	$3.03 E^2$
RLg19	$1.42 E^4$	$2.07 E^5$	RQz19	$1.05 E^3$	$7.25 E^2$
RQc19	$6.71 E^4$	$1.84 E^5$	RZm20	$1.45 E^3$	$6.16 E^3$
RRm19	$6.87 E^4$	$3.08 E^5$	RGg20	$1.51 E^3$	$2.59 E^3$
RTp19	$4.09 E^4$	$5.74 E^5$	RYy19	$2.53 E^2$	
RVh19	$7.02 E^4$	$1.85 E^5$	RCa20	$4.910 E^3$	$1.61 E^3$
RVr19	$1.44 E^4$	$2.47 E^5$	RJa20	$2.31 E^3$	$2.25 E^3$
RWc19	$3.13 E^4$	$1.42 E^5$	RUy19	$6.36 E^2$	

### 5.1.3 Time to reach maximum concentration

The time to reach the maximum concentration is displayed in Table 5.3. This is the time corresponding to the values in Table 5.2, time of maximum concentration. Once again the subjects RYy19 and RUy19 are empty because their antibodies were always below the limit of detection.

Table 5.3: Time to reach maximum Concentration for intermediate and Late data

Late Data			Intermediate Data		
Animal ID	gp41	gp120	Animal ID	gp41	gp120
RIm19	8	10	RFz19	4	2
RJm20	6	8	RKa20	4	4
RKg19	8	9	ROz19	4	8
RLg19	9	9	RQz19	8	8
RQc19	10	10	RZm20	28	8
RRm19	8	8	RGg20	8	8
RTp19	8	8	RYy19	8	
RVh19	6	6	RCa20	4	16
RVr19	6	10	RJa20	8	4
RWc19	8	10	RUy19	4	

The values of AUC, Cmax, and Tmax were determined by using PkAnalix, an application of MonolixSuite.

#### 5.1.4 Terminal Half-life

Terminal half-life is the time required to divide the plasma concentration by two after reaching pseudo-equilibrium. To determine the half-life, the terminal slope of the curve known as  $\lambda_z$  is first calculated. To ideally determine  $\lambda_z$ , 3-4 half-lives should have elapsed. Normally, this is not usually possible, then a minimum of 3-4 observations is required (101).  $\lambda_z$  is related to steady state toward the end of data. The half-life is then given by,

$$t_{1/2} = \frac{\ln(2)}{\lambda_z}, \quad (5.3)$$

The first thing that we did in determining the terminal half-life was to mimic the results by another group that used the late data. The half-life of more than 60% of the data was similar to the results obtained by the other group, which used a different programming platform. Looking at the plots generated by Pkanalix, we decided to change the number of points, chosen by PkAnalix by default, by adding or reducing to the points used to generate the half-life. After changing the points needed for the terminal slope, the results matched those of the other groups. However, changing the points to get the results we wanted is a bad strategy as it hinders reproducibility.

Regardless, we tried the NCA method with the data obtained from the macaques that were started on ART after two weeks (intermediate group). However, the results generated did not make sense to the experts. The assays were redone, but still the results were still unsatisfactory.

Upon reflection, we realized that NCA was not the right technique to be used for this problem for several reasons. The problem is that we thought NCA could be used to estimate the terminal half-life since it works in the study of drugs. It is reasonable that it works in drugs because once the drugs are halted, they eventually clear out of the body after a certain period of time. This, however, is not the case with antibodies. They do not completely clear out, but the concentration goes towards a horizontal asymptote. There were other issues as well when it comes to using NCA to determine the terminal half-life of the antibodies. This is because this was not a typical Pk study: antibodies are developed differently in different individuals. The antibody response is dynamic, and the plasma concentration may depend on factors other than clearance.

## 5.2 Biphasic Model

Accompanying the realisation that the non-compartmental analysis was a wrong method to determine the terminal half-life, we looked into using the biphasic (106) (107), because the decay of antibodies in the data appears to have two different phase. A biphasic model is expressed with two exponential decay rates as follows,

$$A(t) = B e^{-\delta t} + C e^{-\gamma t} \quad (5.4a)$$

$$A(t) = A_0 r e^{-\delta t} + A_0 (1 - r) e^{-\gamma t} \quad (5.4b)$$

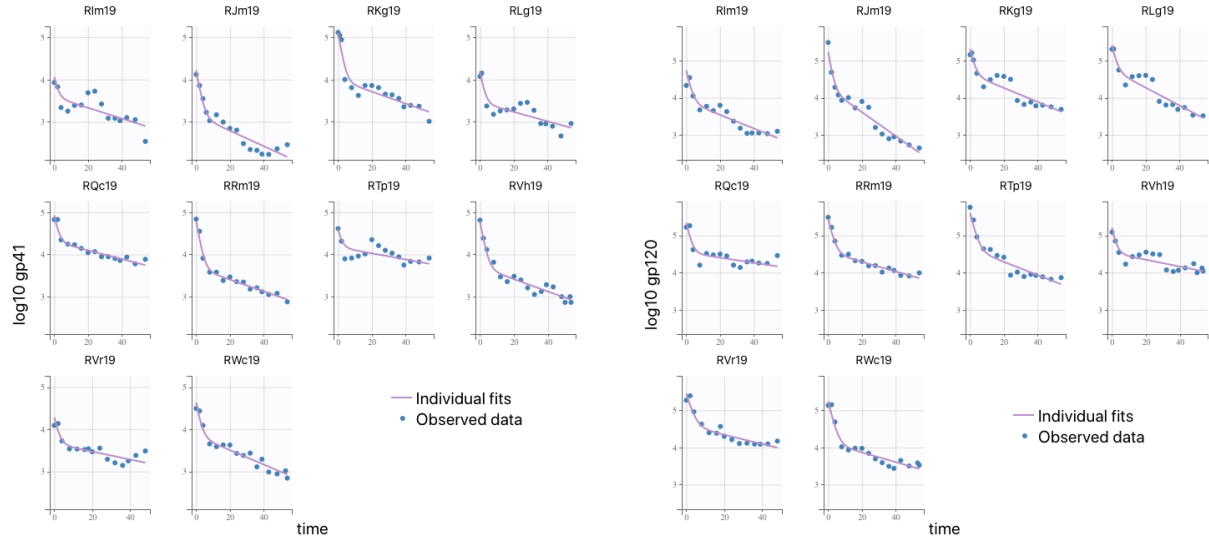
In the equation above, the parameters are defined as  $t$ : age of the macaques,  $\delta$ : the fast decay rate of antibodies,  $\gamma$ : the slow decay rate of antibodies,  $A_0$ : the initial antibody concentration, and  $A_0 = A(t = 0) = B + C$ , and  $r = \frac{B}{(B+C)}$ . In (5.4), it is assumed that the initial time corresponds to the time of the maximum concentration, and that the concentration keeps decline. But, this is not true with antibodies. Patients do not develop antibodies until post-infection. As time passes by, the concentration of antibodies increases, but it generally starts declining when subjects are put on ART. In the model, the post peak half-life and the terminal half-life can respectively be calculated using:  $(t_{1/2} = \frac{\ln 2}{\delta})$  and  $(t_{1/2} = \frac{\ln 2}{\gamma})$ . In accordance with (106), we made the following assumptions about the distribution of the parameters,  $A_0, \delta, \gamma$ : log-normal and  $r$ : logit-normal.

### 5.2.1 Late Antiretroviral Therapy Initiation

We start with the results from the late ART initiation. In this group we make the assumption that the peak concentration occurs at week 8, since we expect to see a decline after ART initiation. Using this logic, initial antibody concentration  $A_0$  is defined as the concentration at week 8. The plots generated from the estimated parameters are displayed in Figure 5.1.

The biphasic fits in Figure 5.1 are reasonable and satisfactory. In Figure 5.2 and Figure 5.3, we do a side by side comparison of the biphasic fit against the plots obtained from the NCA estimation of the terminal half-life. For gp41, subject RVr19 used all the points in NCA determination of terminal half-life. There is an obvious problem with the estimation because the antibodies start increasing after week 40, which NCA cannot handle as the technique expect concentration to keep decreasing. For the gp120, we see that subjects

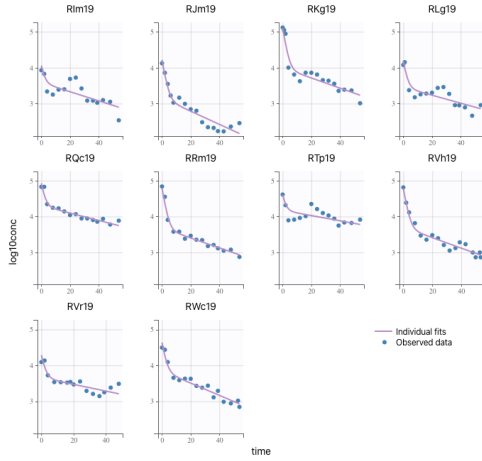
RJm19 and RKg19 used only the last three points to estimate the terminal slope (and the terminal half-life). In both subjects, there is a concentration decline after the peak followed by a slight increase (47 weeks for RJm19 and week 44 for RKg19) before the concentration starts decreasing again. The increase should not have affected the results because it is one data point for both subjects, and the increase was not dramatic.



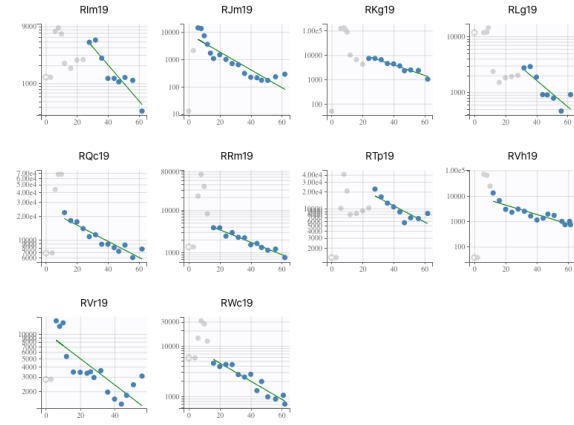
(a) gp41 biphasic result

(b) gp120 biphasic results

Figure 5.1: Fit of antibodies using Biphasic equation

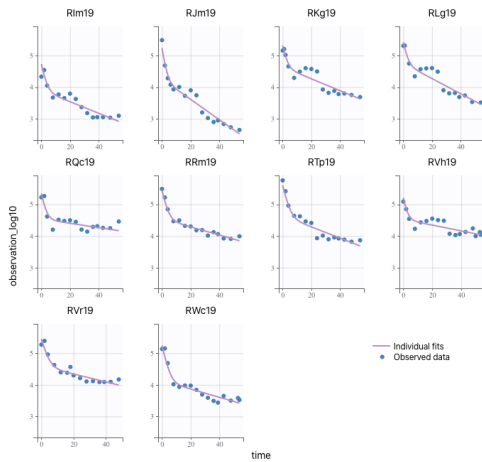


(a) gp41: biphasic

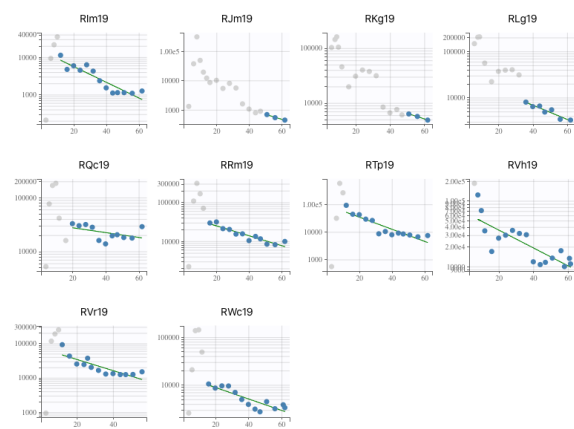


(b) gp41: NCS

Figure 5.2: (a) biphasic vs. (b) NCA gp41 for Late initiation ART data.



(a) gp120: biphasic



(b) gp120: NCA

Figure 5.3: (a) biphasic vs. (b) NCA gp120 for Late initiation ART data.

The terminal half-life resulting from the biphasic fit are in Table 5.4 and Table 5.5. In

both gp41 and gp120 antibodies, the results are not surprising. We expected the post-peak half-life to be shorter, because the concentration is decaying at a higher rate, than the terminal half-life.

Using the estimated parameters, we can calculate  $B$  and  $C$  and use them to find the phase transition time (the time that switches from short-lived to death rates to long-lived death rates of the antibodies). The phase transition time is defined by the equation below,

$$T_t = \frac{\log(B\delta) - \log(C\gamma)}{\delta - \gamma}. \quad (5.5)$$

The phase transition times should be more than the the post-peak half life and less than the terminal half-life. And our calculations were on par with the expectation, Table 5.6 and Table 5.7



Table 5.4: Half-life gp120 Late Data

Animal ID	delta	gamma	Post peak half-life	Terminal half-life
RIIm19	0.503	0.0412	1.378	16.843
RJm19	0.663	0.0727	1.046	9.532
RKg19	0.540	0.0426	1.285	16.262
RLg19	0.552	0.0541	1.257	12.817
RQc19	0.537	0.0157	1.290	44.017
RRm19	0.481	0.0286	1.442	24.259
RTp19	0.451	0.0387	1.536	17.899
RVh19	0.582	0.0207	1.192	33.502
RVr19	0.437	0.0276	1.585	25.115
RWc19	0.459	0.0285	1.509	24.357

Table 5.5: tab: Half-life gp41 Late Data

Animal ID	delta	gamma	Post peak half-life	Terminal half-life
RIIm19	0.559	0.0284	1.240	24.425
RJm19	0.551	0.0417	1.258	16.607
RKg19	0.573	0.0325	1.209	21.296
RLg19	0.568	0.0253	1.221	27.354
RQc19	0.549	0.0236	1.264	29.358
RRm19	0.557	0.0307	1.243	22.553
RTp19	0.570	0.0162	1.216	42.840
RVh19	0.539	0.0333	1.287	20.800
RVr19	0.554	0.0214	1.251	32.436
RWc19	0.547	0.0378	1.267	18.347

Table 5.6: The remaining parameters from the biphasic model for gp120: Late Data

Animal ID	r	B	C	$A_0$	$T_t$
RIml19	0.846	$5.45 E^4$	$4.61 E^4$	$8.36 E^3$	3.96
RJm19	0.896	$1.790 E^5$	$1.61 E^5$	$1.86 E^4$	3.21
RKg19	0.782	$2.09 E^5$	$1.64 E^5$	$4.55 E^4$	3.34
RLg19	0.778	$2.65 E^5$	$2.06 E^5$	$5.88 E^4$	3.12
RQc19	0.834	$2.26 E^5$	$1.88 E^5$	$3.76 E^4$	4.28
RRm19	0.865	$2.72 E^5$	$2.35 E^5$	$3.67 E^4$	4.50
RTp19	0.891	$4.03 E^5$	$3.59 E^5$	$4.37 E^4$	4.80
RVh19	0.785	$1.70 E^5$	$1.33 E^5$	$3.65 E^4$	3.58
RVr19	0.855	$2.84 E^5$	$2.43 E^5$	$4.12 E^4$	4.81
RWc19	0.924	$1.80 E^5$	$1.66 E^5$	$1.37 E^4$	5.32

Table 5.7: The remaining parameters from the biphasic model for gp41: Late Data

Animal ID	$r$	$B$	$C$	$A_0$	$T_t$
RIml19	0.655	$1.18 E^4$	$7.76 E^3$	$4.08 E^3$	2.97
RJm19	0.912	$1.73 E^4$	$1.58 E^4$	$1.52 E^3$	4.20
RKg19	0.920	$1.32 E^5$	$1.21 E^5$	$1.05 E^4$	4.26
RLg19	0.799	$1.53 E^4$	$1.22 E^4$	$3.07 E^3$	3.60
RQc19	0.741	$8.48 E^4$	$6.28 E^4$	$2.19 E^4$	3.47
RRm19	0.919	$6.25 E^4$	$5.75 E^4$	$5.04 E^3$	4.40
RTp19	0.613	$3.98 E^4$	$2.44 E^4$	$1.54 E^4$	3.16
RVh19	0.916	$6.36 E^4$	$5.82 E^4$	$5.34 E^3$	4.44
RVr19	0.752	$2.00 E^4$	$1.50 E^4$	$4.96 E^3$	3.56
RWc19	0.838	$4.48 E^4$	$3.76 E^4$	$7.25 E^3$	3.68

### 5.2.2 Intermediate Antiretroviral Therapy Initiation

In this section, we redo the same calculations done in the previous section. We, however, use the data from the macaques that were started on ART at two weeks post-infection. We fit the data, Figure 5.4 and Figure 5.5, and estimated the terminal half-life, Table 5.8 and Table 5.9.

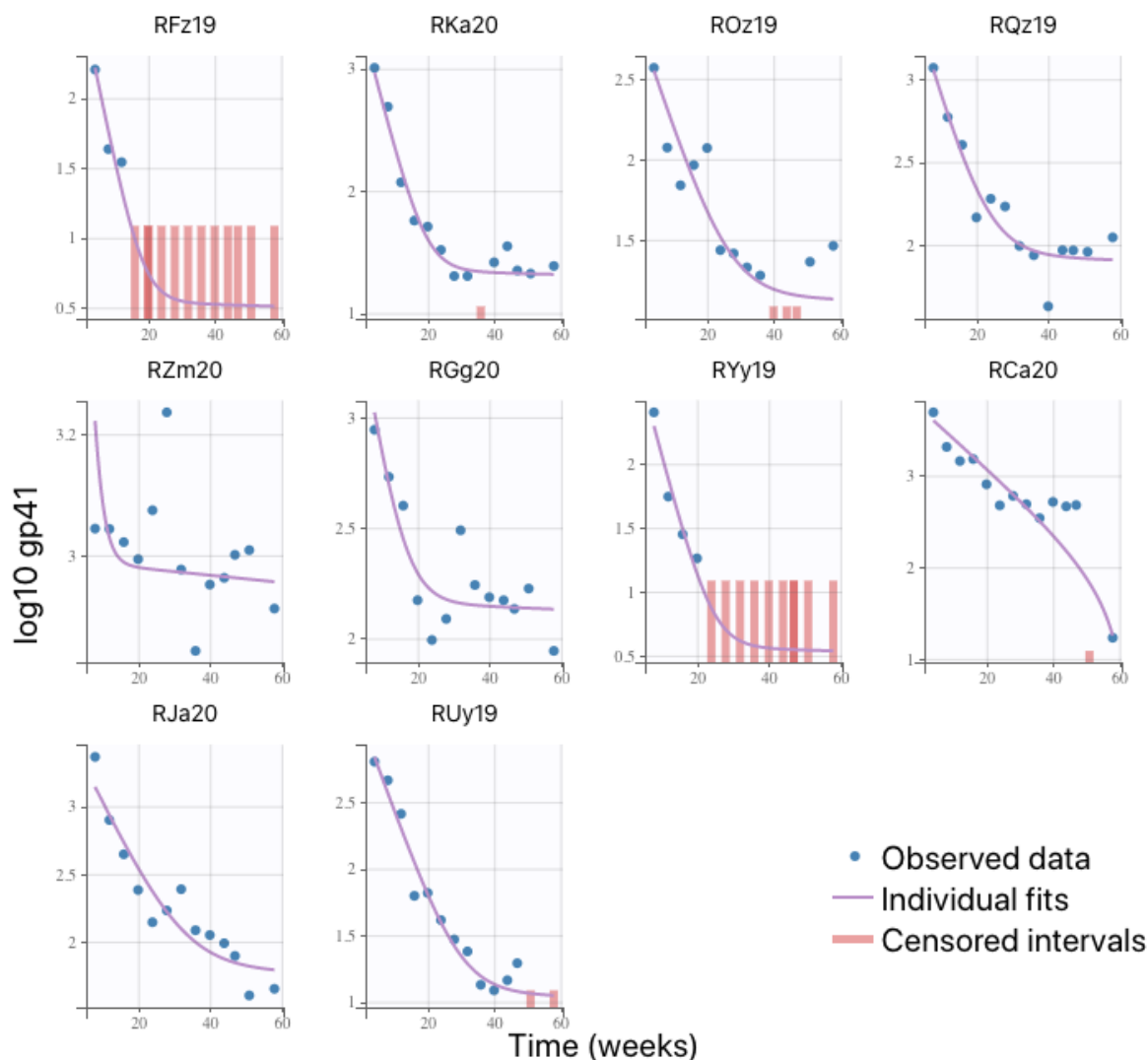


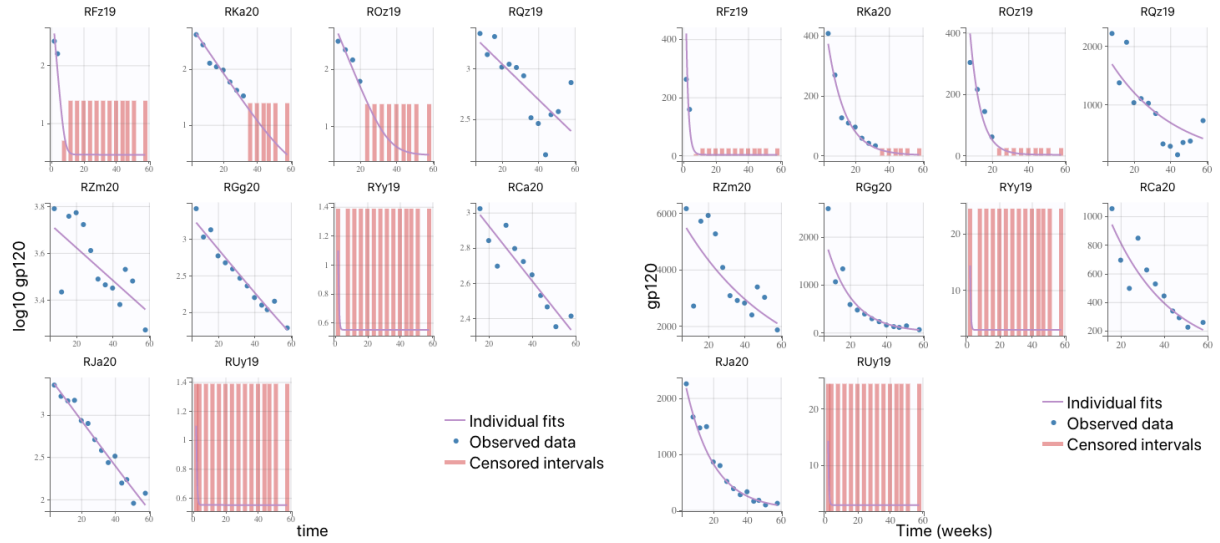
Figure 5.4: Fit of gp41 antibodies using Biphase equation for intermediate data

In Figure 5.4, the fit of subject RCa20 appears strange and not at all exponential in the way that one would expect. That appearance is because the estimation of parameters is influenced by the last data point. We looked at the Table 5.8 to analyze whether its half-life estimate are abnormal, compared to other subjects. The terminal half-life value, 159.82 weeks, is neither the highest nor the lowest. However, the post-peak half-life is more than twice the amount of 7 out of 9 other subjects. Looking at the remaining estimated

parameters  $r$ ,  $r_1 = 1.00944$ , we realized that there was an issue because the value is assumed to be exclusively less than 1,  $0 < r < 1$ . Treating the last data point of RCa19 as an outlier yielded better fit, but we kept the datum as there was nothing wrong with the assay that we were aware of. The parameters for the other parameters are in Table 5.11. In the case of subject RCa20, the values are left intentionally blank due to the last data point.

Table 5.8: tab: Half-life gp41 Intermediate Data

Animal ID	delta	gamma	Post peak half-life	Terminal half-life
RFz19	0.253	0.00410	2.739625548	168.927
RKa20	0.237	0.00267	2.929059057	259.781
ROz19	0.147	0.00260	4.716634552	266.176
RQz19	0.177	0.00268	3.921358553	258.533
RZm20	0.428	0.00298	1.621192131	232.797
RGg20	0.240	0.00275	2.892512281	251.614
RYy19	0.226	0.00801	3.073796155	86.565
RCa20	0.075	0.00434	9.209263383	159.820
RJa20	0.145	0.00775	4.794079432	89.390
RUy19	0.158	0.00640	4.390731257	108.302



(a) gp120 biphasic result

(b) gp120 biphasic results

Figure 5.5: Fit of gp120 antibodies using Biphasic equation for intermediate data

The biphasic model for the fitting of gp120 on intermediate ART data, Figure 5.5, was done twice. The first time it was done on with the output being on the log10 scale, Figure 5.5a. The biphasic model data fits appear to be linear with the exception of subjects RFz19 and ROz19, whose data points are dominated by the censored intervals. The fits are still biphasic, with two exponential decay, Table 5.9. In Figure 5.5b, the same data is fit (the y-axis is no longer on the log10 scale), we then see the appearance of two exponential decay. The estimated parameters, in both situations, were practically the same. Table 5.10 has the estimated parameter,  $r$ , as well as the remaining parameters that were algebraically calculated from the estimated parameters.

Table 5.9: Half-life gp120 Intermediate Data

Animal ID	delta	gamma	Post peak half-life	Terminal half-life
RFz19	0.774	0.000120	0.895	5777.333
RKa20	0.100	0.000123	6.898	5614.755
ROz19	0.186	0.000183	3.728	3782.151
RQz19	0.0395	0.000210	17.529	3300.842
RZm20	0.0137	0.000142	50.476	4888.720
RGg20	0.0694	0.000101	9.983	6873.528
RYy19				
RCa20	0.0343	0.000119	20.218	5848.850
RJa20	0.0622	0.000143	11.149	4854.244
RUy19				

Overall the values of the terminal half-life are higher in the intermediate data compared to the late data. This is because the concentration of antibodies in intermediate ART is smaller that after rapid antibodies decline the values appear to be leveled up. When values that give rise to the second exponential decay appear almost horizontal (as if not changing), the slow decay rate of antibodies,  $\gamma$ , is close to zero which yields to higher values of the half-life.



Table 5.10: tab: The remaining parameters from the biphasic model for gp120: Intermediate Data.

Animal ID	r	B	C	A <sub>0</sub>	T <sub>t</sub>
RFz19	0.9982	1750	3.1394	1757	8.4709
RKa20	0.9982	639	1.1772	640	56.2458
ROz19	0.9982	2000	3.6904	2006	30.9062
RQz19	0.9982	2450	4.5278	2457	127.3428
RZm20	0.9982	5501	10.0799	5511	347.5295
RGg20	0.9981	2960	5.5395	2964	80.276
RYy19					
RCa20	0.9981	1670	3.0573	1673	152.1698
RJa20	0.9981	3010	5.6187	3019	86.5445
RUy19					

Table 5.11: tab: The remaining parameters from the biphasic model for gp41: Intermediate Data.

Animal ID	r	B	C	A <sub>0</sub>	T <sub>t</sub>
RFz19	0.9856	414	6.0518.	420.118	14.5646
RKa20	0.9874	2279	29.0467	2308.41	16.4230
ROz19	0.9725	606	17.1164	622.731	22.8627
RQz19	0.9766	4418	105.678	4523.71	19.7628
RZm20	0.9666	25870	893.172	26760.9	8.5237
RGg20	0.9719	6463	187.161	6650.59	14.6813
RYy19	0.9923	1090	8.4253	1098.62	16.3755
RCa20	1.00944				
RJa20	0.9780	4938	111.317	5048.86	21.3223
RUy19	0.9862	1239	17.3824	1256.59	21.4248

## **5.3 Comparison of the the three stages of ART administration**

### **5.3.1 Statistical Test**

#### **One-way analysis of variance (ANOVA)**

One-way analysis of variance (ANOVA) is a statistical test designed to test whether the mean in more than two independent groups differ. The ANOVA seeks to test if the null hypothesis (the means among the groups being evaluated are equal). In the absence of the null hypothesis, the ANOVA test implies that at least one mean differs from the rest of the group.

ANOVA is based on assumptions that the errors have equal variance as well as independent and normally distributed. In violation of these assumptions, the test is not reliable. The information gleaned from ANOVA is the presence of evidence that at least one of the group is significantly from the others. In case of more than two groups, it is desirable to simultaneously compare all the pairs. When doing multiple comparisons in ANOVA, there are several follow-up tests that can inform exactly which condition/group differs from each other (108).

#### **Kruskal–Wallis Test**

The Kruskal-Wallis H test (sometimes also called the "one-way ANOVA on ranks") is considered the alternative to the one-way ANOVA. This is because the Kruskal–Wallis Test is a rank-based non-parametric test that, like ANOVA, can be used to determine if there are statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable (109). The non-parametric part of the Kruskal–Wallis Test means that there is no assumption that the data possesses a particular distribution. To avoid making wrong assumptions regarding our data, we use the Kruskal–Wallis Test.

## 5.3.2 Comparison Results

### Glycoprotein 41

In this section, we compare results from the NCA: AUC, Cmax and Tmax for gp41 as well as the terminal half-life (obtained from the biphasic model) for the intermediate and late ART initiation. We start by displaying the box plot. A boxplot displays a standardized distribution of data based on a five number five quartiles (“minimum”, first quartile (Q1), 25 percentile (Q2), median (Q3), 75 percentile (Q4), and maximum (Q5)). Box plots can also divulge information if the data is symmetrical or how tightly it is grouped. By scanning Figure 5.6, we see there is a possibility that median as well as the overall distribution for the AUC, Cmax and half-life for the late and intermediate stages appear different.

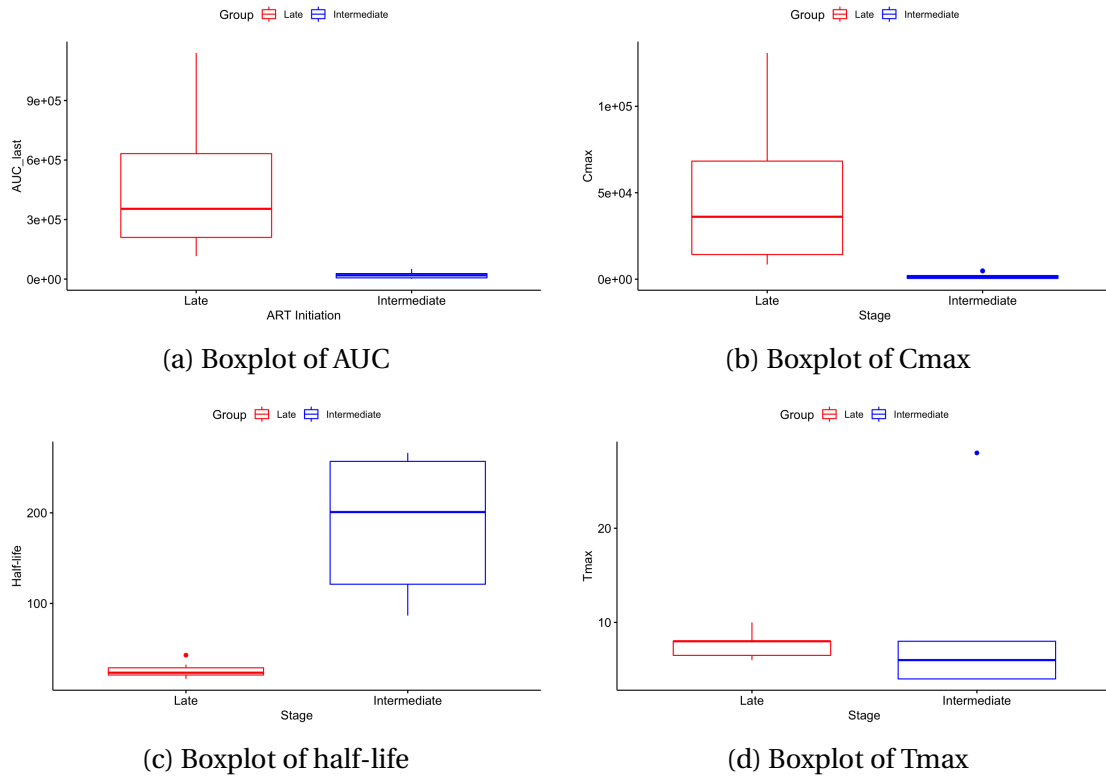


Figure 5.6: Box plots of gp41 for the AUC, Cmax, half-life and Tmax.

The examination of Figure 5.6 is not enough to conjecture about the similarities or

differences between the two datasets. We need a formal way to say with certainty whether or not our intuition is correct. We use the Kruskal–Wallis Test to draw statistical conclusion, Table 5.12. From the results table, the null hypothesis is rejected for Cmax, AUC and half-life. This means that the means of the PK parameters of gp41 for the animals who were introduced to ART after 8 weeks and 2 weeks are significantly different, except for the time it takes to reach the maximum concentration.

Table 5.12: Kruskal-Wallis Test for gp41

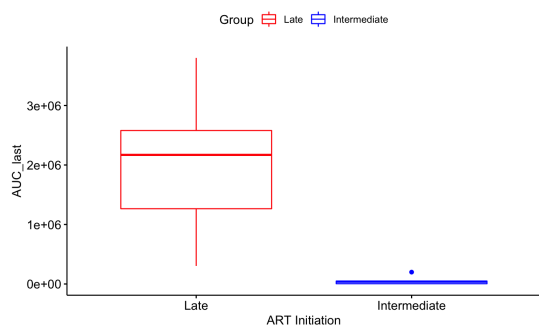
	Df	$\chi^2$	p-value
AUC	1	14.286	0.0001571
Cmax	1	14.286	0.0001571
Half-life	1	14.286	0.0001571
Tmax	1	2.0762	0.1496

### Glycoprotein 120

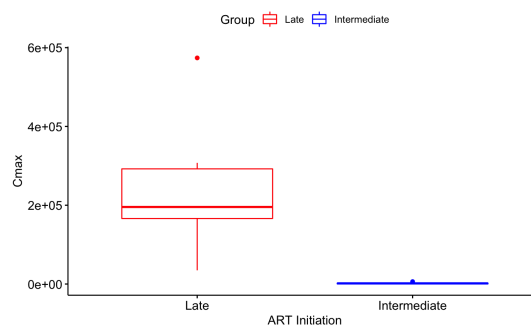
In this section we redo the analysis done with Glycoprotein 41 . By looking at Figure 5.7, we notice that the medians and distributions appear different for the Late and Intermediate in all categories: AUC, Cmax, half-life and Tmax. We perform the Kruskal-Wallis test to confirm the results of the box-plots, Table 5.12. The results rejects the null hypothesis and state that the means of the two groups are statistically different.

Table 5.13: Kruskal-Wallis Test for gp120

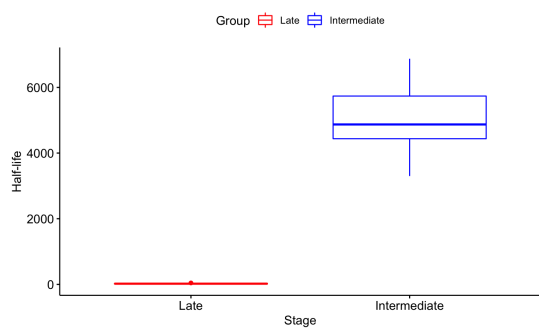
	Df	$\chi^2$	p-value
AUC	1	6.6269	0.01004
Cmax	1	12.645	.0003766
Half-life	1	14.286	0.0001571
Tmax	1	6.626	0.01004



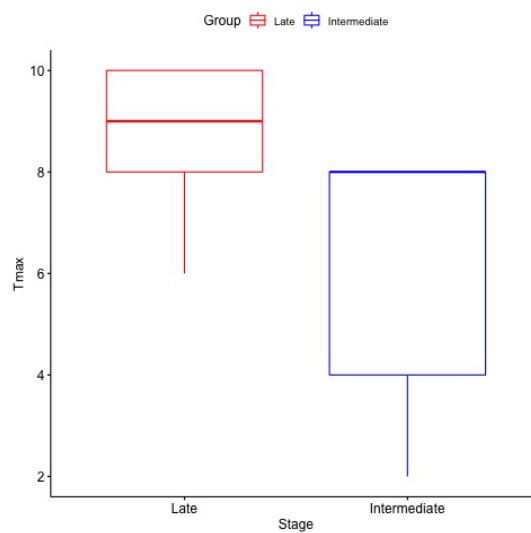
(a) Boxplot of AUC



(b) Boxplot of Cmax



(c) Boxplot of half-life



(d) Boxplot of Tmax

Figure 5.7: Box plots of gp120 for the AUC, Cmax, half-life and Tmax.

## CHAPTER

# 6

## MODELING VIRAL LOAD AND ANTIBODIES

The purpose of this chapter is to develop a mathematical model that can predict how antibodies responses develop in infants rhesus macaques subject to a SHIV infection, in the late ART treatment group.

### 6.1 Parameter Estimation

The parameters for the HIV model were estimated using a non-linear mixed effect model. We opt to use the NLMEM because we are interested in doing the analysis of repeated longitudinal measured data of ten rhesus macaques. We use MONOLIX to estimate the parameters. The output of each parameter estimation will have it respective standard deviation. Our output of interests,  $\log_{10}$  of the viral load and  $\log_{10}$  of antibodies, will also have associated residual errors. The parameters were estimated using the following five steps: model selection, sensitivity analysis, K-means clustering, identifiability and parameter fitting of the sensitive identifiable parameters.



### 6.1.1 Model Selection

There are several models in the literature that describe infected diseases. We, first, introduce a simple model (6.1) (110). This model describes three variables: uninfected cells  $T$ , infectious cells  $I$ , and viral load  $V$ . In the model,  $T$ , uninfected cells, proliferate at a rate  $\lambda$ , die at rate  $d$  and are converted to infectious cells at a rate  $\beta V T$ . Infected cells die at a rate  $\delta$ . The death of infected leads to the production of viral particles at a rate  $\pi I$ . The free virus particles decay at a rate  $c$  (111).

$$\frac{dT}{dt} = \lambda - dT - \beta V T \quad (6.1a)$$

$$\frac{dI}{dt} = \beta V T - \delta I \quad (6.1b)$$

$$\frac{dV}{dt} = \pi I - c V \quad (6.1c)$$

In the equations above, there is no antiretroviral therapy, ART, to slow down the virus. In the presence of ART, (6.1a & 6.1b) are modified to incorporate the presence of the disease therapy by multiplying the conversion rate  $\beta V T$ , from uninfected cells to infected cells, to  $(1 - \epsilon)$ . The parameter  $\epsilon$  describes the drug efficiency and has range  $0 \leq \epsilon \leq 1$ . We also consider the presence of the immune response  $Z$  (110). The immune response decreases the viral load and proliferates with respect to antigens at a rate  $r I Z$ . The decay rate of the immune response is given by the rate  $\mu$  (6.2).

$$\frac{dT}{dt} = \lambda - dT - \beta(1 - \epsilon) V T \quad (6.2a)$$

$$\frac{dI}{dt} = \beta(1 - \epsilon) V T - \delta I - s I Z \quad (6.2b)$$

$$\frac{dV}{dt} = \pi I - c V \quad (6.2c)$$

$$\frac{dZ}{dt} = r I Z - \mu Z \quad (6.2d)$$

We assume that antibodies are the immune response in (6.2). The antibodies in the model are of an adaptive (acquired) immune response, which are highly specific for a particular pathogen (112). We modified the model to properly include a simplified mechanism

of antibodies to remove the virus. The body's immune system is capable of producing an immune response via B-cells. The virus prompts the B-cell to divide and mature into antibody secreting cells called plasma cells. The plasma cells are fundamental in secreting antibodies at a much faster rate. However, large B-lymphocytes also secrete antibody, though at a lower rate. The antibodies flow in the blood to bind to the original antigen. Incorporating this mechanism in the model, the model is as described in Figure 6.1 and (6.3) for its mathematical interpretation. The meaning of the parameters and their respective units are in Table 6.1

$$\frac{dT}{dt} = \lambda T \left( 1 - \frac{T+I}{T_m} \right) - \beta(1-\eta)VT \quad (6.3a)$$

$$\frac{dI}{dt} = \beta(1-\eta)VT - \delta I - s * I * A \quad (6.3b)$$

$$\frac{dV_I}{dt} = \pi I - c V_I - k_p * A * V_I \quad (6.3c)$$

$$\frac{dV_{NI}}{dt} = k_p * A * V_I - c_{ni} V_{NI} \quad (6.3d)$$

$$\frac{dB}{dt} = p_i * B * V_I + p_{ni} * B * V_{NI} + \gamma_1 * u * B * \left( 1 - \frac{B}{Bm} \right) - \gamma_2 * (1-u) * B - \mu_1 B \quad (6.3e)$$

$$\frac{dP}{dt} = \gamma_2 * (1-u) * B - \mu_2 * P \quad (6.3f)$$

$$\frac{dA}{dt} = r_1 * B + r_2 * P - \mu_3 A - k_p * A * V_I \quad (6.3g)$$

$$\eta \text{ is drug efficiency: } \eta = \begin{cases} 1 & \text{With ART} \\ 0 & \text{No ART} \end{cases}$$

$u(0 \leq u \leq 1)$  is the fraction of B-cells

$(1-u)$  is the fraction that differentiate into the plasma cells

In equation (6.3), we assume that ART is 100 % effective. The parameters  $p_i$  and  $p_{ni}$  are the rate that induces the formations B-cells from the infectious and non-infectious viral load. The B-cells proliferate with constant birth rate  $\gamma_1$  and differentiate into plasma cells at a constant rate  $\gamma_2$ . The proliferation of B-cells are multiplied by the logistic equation because memory B-cells never wanes and also cannot be sustainably boosted. The death rate of the large B-cells and the plasma cells are respectively  $\mu_1$  and  $\mu_2$ . There is an assumption of  $\gamma_1 > \mu_1$  as an insurance of a net growth in the large B-cells population. Due to the clinical

knowledge that both the B-cells and the plasma cells produce antibodies with plasma cells secreting them at a much faster rate than the B-cells, it is assumed that  $r_1 < r_2$  (113).  $\mu_3$  is the death rate of antibodies.

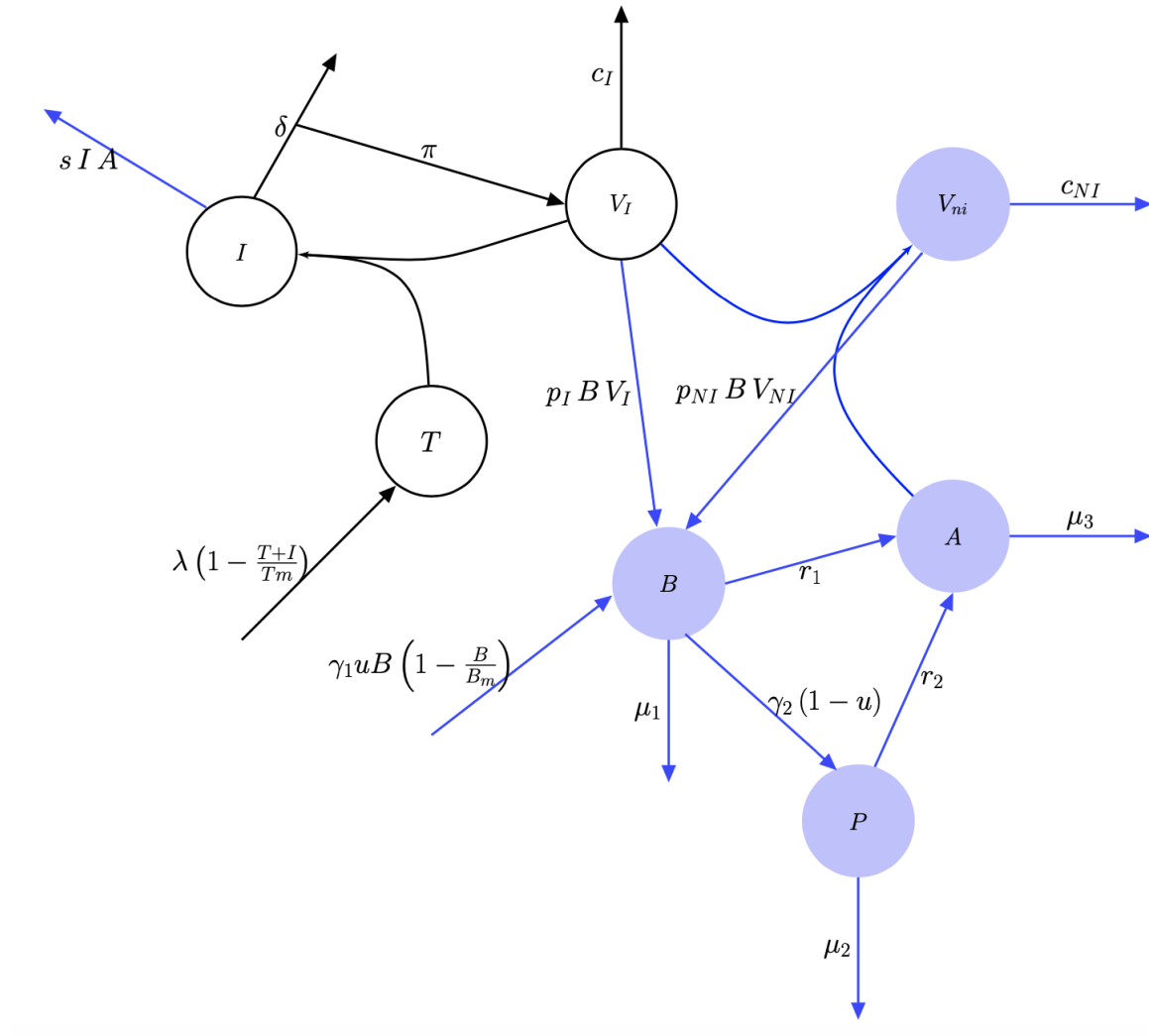


Figure 6.1: Schematic of antibodies and viral load model

Table 6.1: Parameters definition and units of the HIV model

Parameter	Definition	units
$\lambda$	T-cells proliferation rate	week <sup>-1</sup>
$T_m$	T-cells proliferation rate	cells mL <sup>-1</sup>
$\beta$	infectivity rate constant	mL week <sup>-1</sup>
$\delta$	infected cell killing rate	week <sup>-1</sup>
$s$	rate of Infected cells removed by antibodies	cells <sup>-1</sup> mL week <sup>-1</sup>
$\pi$	virus production rate	week <sup>-1</sup>
$c$	virus clearance rate	week <sup>-1</sup>
$k_p$	antibody binding rate	mL week <sup>-1</sup>
$c_{ni}$	Non-infected virus clearance rate	week <sup>-1</sup>
$p_I$	B-cells production from infectious virus	week <sup>-1</sup>
$p_{NI}$	B-cells production from non-infectious virus	week <sup>-1</sup>
$\gamma_1$	maximum proliferation of B-cells	week <sup>-1</sup>
$B_m$	B-cells carrying capacity	cells mL <sup>-1</sup>
$\gamma_2$	rate of change from B-cells to plasma cells	week <sup>-1</sup>
$\mu_1$	death rate of B-cells	weeks <sup>-1</sup>
$\mu_2$	death rate of B-cells	week <sup>-1</sup>
$r_1$	rate Antibodies production from B-cells	week <sup>-1</sup>
$r_2$	rate Antibodies production from plasma cells	week <sup>-1</sup>
$\mu_3$	Plasma cells decay rate	week <sup>-1</sup>

### 6.1.2 Sensitivity Analysis

Subsequent to selecting a model and estimating all the parameters, the sensitivity analysis was performed. We use the complex step method because we wanted to take a small step size,  $h = 10^{-16}$  and avoid cancellation error. We assumed the parameters estimated to be true, and proceeded to performing the global sensitivity analysis, Figure 6.2, assuming the mean and standard deviation are the population parameters and the standard deviation obtained from MONOLIX NLMEM. For the sake of simplicity, we ignored the viral load, whose parameters change depending on whether they are simultaneously being estimated

with gp41 or gp120. We only looked at the sensitivity of gp120 and gp41.

### 6.1.3 K-means clustering

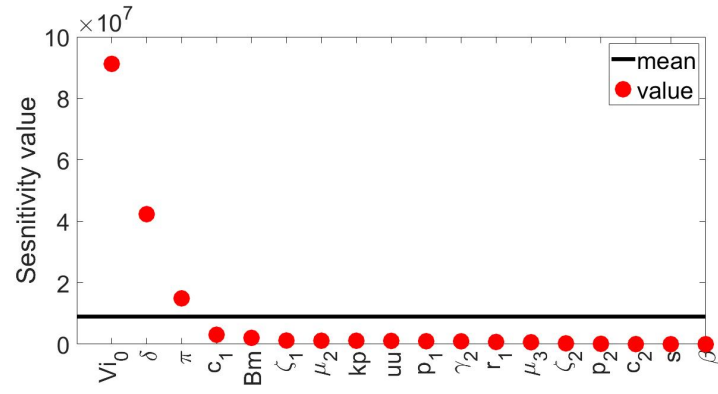
There are cases where the sensitivity values obtained from (1.16) are nicely spread out, Figure 1.4. If so, we can use k-means clustering, with  $k = 3$  for three clusters, to classify the sensitivity coefficients as ultra sensitive, sensitive, and non-sensitive, Figure 1.5. In such cases, we consider the sensitive parameters to be those that are ultra-sensitive and sensitive.

#### gp120

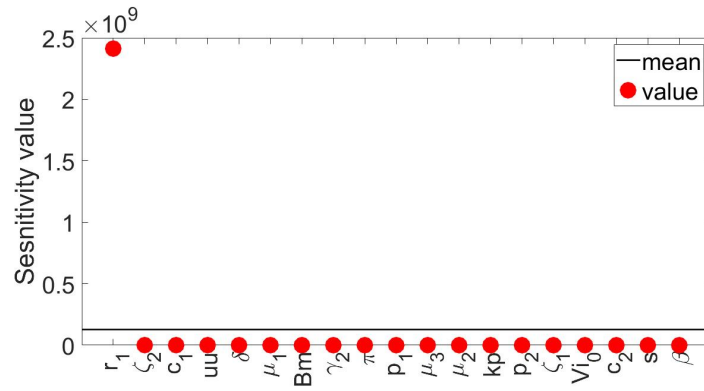
Our goal was to classify the sensitivity coefficients into three clusters: non-sensitive, sensitive, and ultra-sensitive. However, the values are not nicely dispersed. We, nonetheless, tried to use k-means with three clusters, only two parameters values were considered sensitive, Figure 6.3a. Two parameters values being sensitive is not necessarily an issue. However, from our global sensitivity plot, Figure 6.2a, we have three values that are above the mean line (which is another technique that has been used to differentiate between sensitive and non-sensitive parameters). We alternatively used the k-means clustering where  $k = 4$  to see if it will differentiate between the third and fourth point, which are both close to the mean line. In Figure 6.3b, we notice that the third and fourth point are then separated. We, therefore, decided to keep the k-means with four clusters, where three parameters are sensitive (because it can also be justified by taking the parameters above the mean line).

#### gp41

For gp41, we use k-means clustering to divide the sensitivity values into three clusters, Figure 6.4a. Cluster 1 and cluster 3 appear peculiar because by visually inspecting the figure, they seem to be of equal values. This is because the first sensitivity value in Figure 6.4a, corresponding to the parameter  $r_1$ , is considerably larger compared to the rest of them (that when graphed all the other sensitivity values appear to be equal). We plot the rest of the parameters values without  $r_1$ , Figure 6.4b, we notice that the three points (that are in cluster 3 in Figure 6.4a) are a bit removed from the rest of them (those in cluster 1 in Figure 6.4a). Hence, we conclude that there are four sensitive parameters, those in cluster 2 and cluster 3.

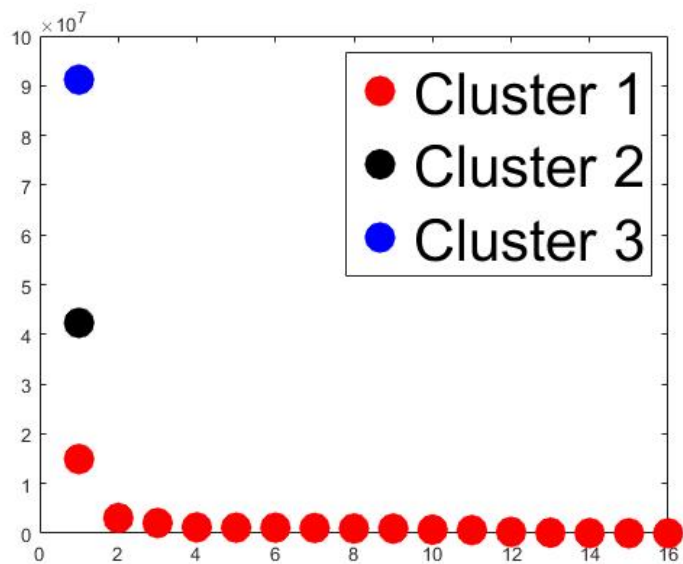


(a) gp120.

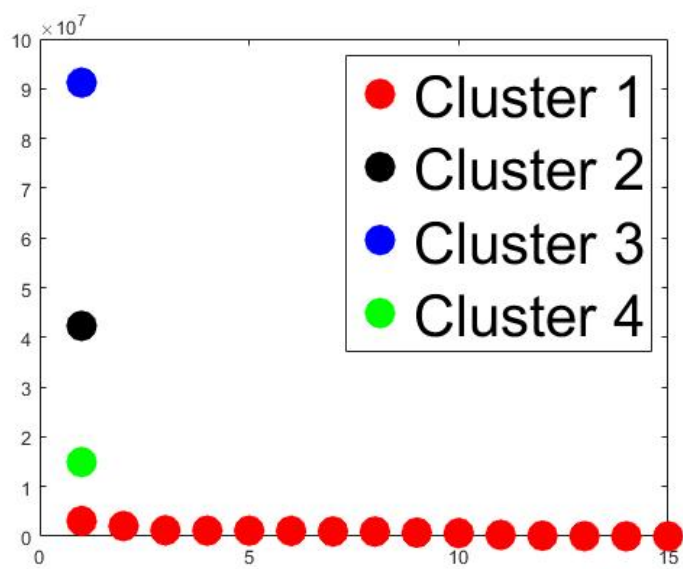


(b) gp41.

Figure 6.2: Plot of sensitivity values.

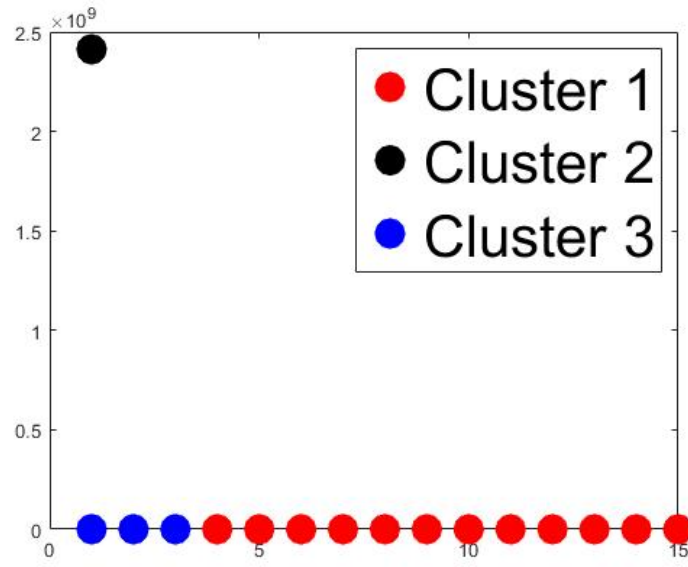


(a) gp120.

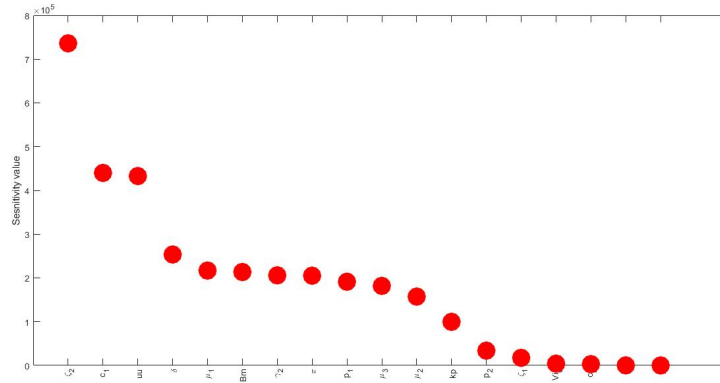


(b) gp120.

Figure 6.3: Plot of sensitivity clusters: gp120.



(a) Plot of sensitivity clusters: gp41



(b) sensitivity plot by removing the parameter  $r_1$

Figure 6.4: Plot of sensitivity clusters: gp41.

### 6.1.4 Identifiability

Knowing the sensitive parameters in gp41 and gp120, we carry out the identifiability analysis.



## gp120

In Figure 6.3b, the first three parameters in the sensitivity plot( $V_{i0}$ ,  $\delta$ , and  $\pi$ ), are either ultra-sensitive or sensitive. We create a sensitivity matrix of these parameters using their respective vectors obtained from the local sensitivity analysis. While performing the identifiability, we discover that all three parameters are identifiable.

On the other hand, looking at Figure 6.4a, we see that the first four parameters ( $r_1$ ,  $\zeta_2$ ,  $c_1$ , and  $u$ ) are considered sensitive. We once again create a sensitivity matrix of these parameters using their respective vectors obtained from local sensitivity analysis. While performing the identifiability, all the eigenvalues in the sensitivity matrix were greater than  $10^{-4}$ , which was taken to be our threshold.

In both case, we proceeded into running the model into MONOLIX once again. This time around, we fixed all parameters considered non-sensitive or non-identifiable, Table 6.2. Given that the parameter  $r_2 > r_1$  and  $\gamma_1 > \mu_1$  in(6.3), we set:  $r_2 = r_1 + \zeta_2$  and  $\gamma_1 = \mu_1 + \zeta_1$ .

Table 6.2: Parameters that are non-sensitive or non-identifiable. NA refer to sensitive and identifiable parameters

Parameter	gp120	gp41
$T_0$	$3.54646 E^7$	$1.646432 E^8$
$\lambda$	$5.78368 E^{-8}$	$1.54731 E^{-8}$
$Tm$	$6.57571 E^9$	$1.55652 E^8$
$\beta$	$2.44363 E^{-4}$	$1.09338 E^{-3}$
$\delta$	NA	NA
$s$	$2.86421 E^{-3}$	$3.94034 E^{-3}$
$\pi$	NA	$9.56078 E^{-3}$
$c_i$	1.30526	NA
$kp$	$5.78368 E^{-6}$	$1.19002 E^{-5}$
$c_{ni}$	$6.81887 E^{-3}$	$6.46100 E^{-3}$
$p_i$	$8.76884 E^{-1}$	$5.317143 E^1$
$p_{ni}$	$4.21905 E^{-1}$	$8.14621 E^1$
$\gamma_2$	$6.50842 E^{-1}$	$4.74917 E^{-1}$
$\mu_1$	$5.78367 E^{-8}$	1.87742
$\mu_2$	$3.41962 E^{-1}$	$9.690769 E^{-2}$
$\mu_3$	$1.526737 E^{-1}$	1.53916
$\zeta_1$	7.69272	$4.32718 E^{-5}$
$r_1$	$4.27837 E^{-1}$	NA
$\zeta_2$	$1.87496 E^{-1}$	NA
$Vi_0$	NA	2.796057

### 6.1.5 Parameter fitting of sensitive and identifiable parameters

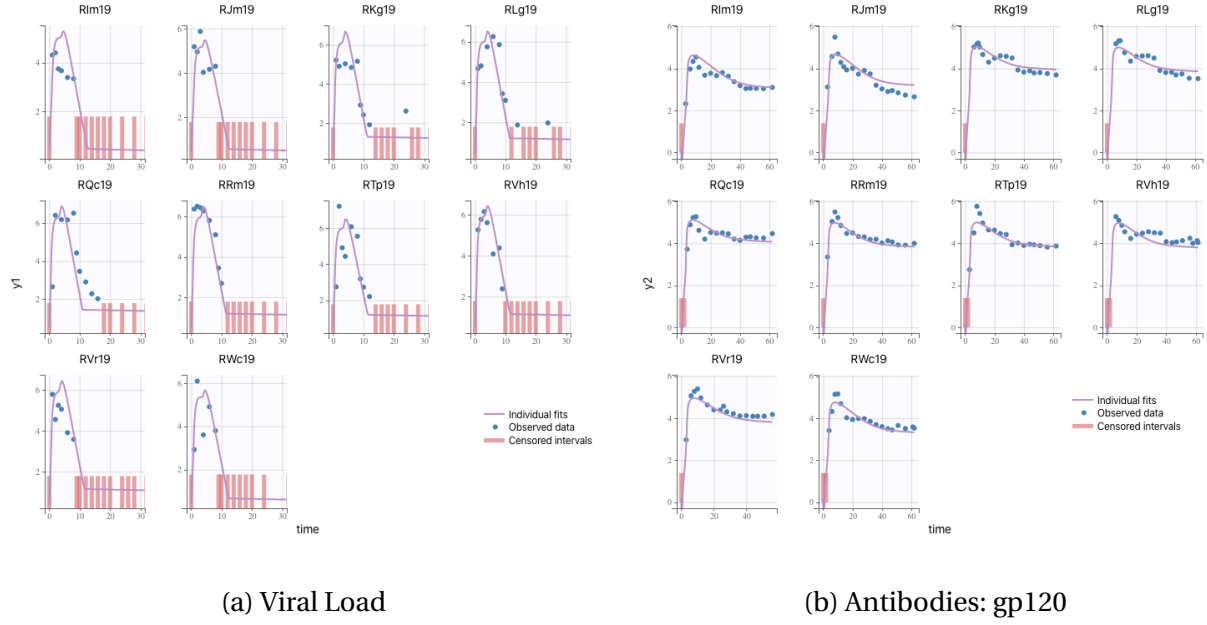
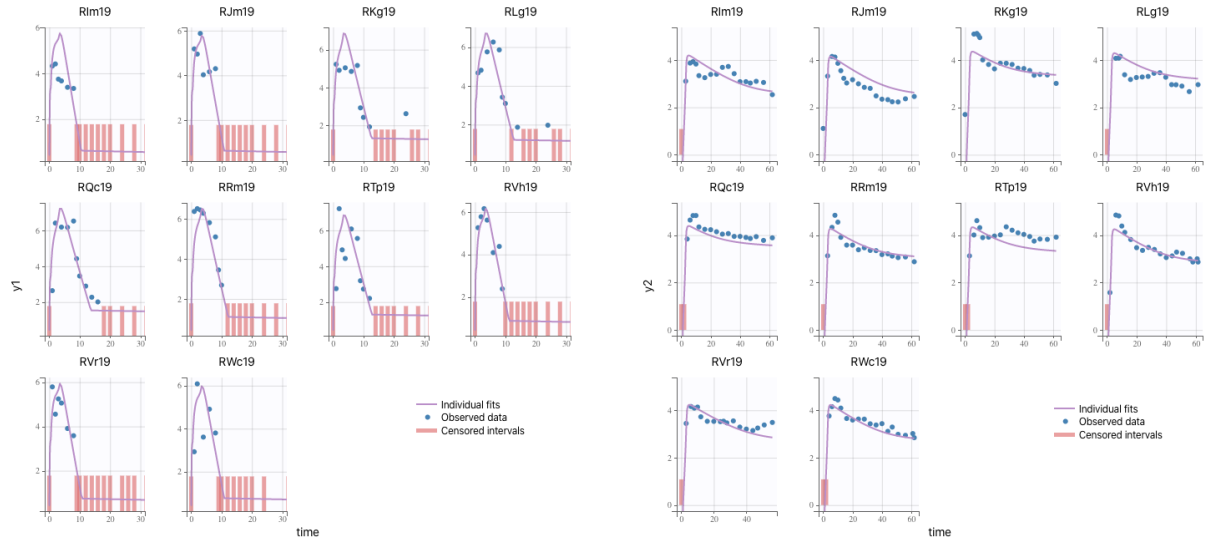


Figure 6.5: Model fit of viral Load and gp120 Antibodies.

Table 6.3: Sensitive and Identifiable parameters, gp120

Parameter	Value	Standard deviation
$V i_0$	$3.18394 E^{-2}$	5.0335441
$\pi$	$1.8254394 E^{-2}$	$3.169787 E^{-1}$
$\delta$	2.2657	$1.02179 E^{-1}$



(a) Viral Load

(b) Antibodies: gp41

Figure 6.6: Model fit of viral Load and gp41 Antibodies.

Table 6.4: Sensitive and Identifiable parameters, gp41

Parameter	Value	Standard deviation
$r_1$	$1.614665 E^{-8}$	2.64705
$\zeta_2$	$2.64525 E^{-3}$	$9.280496 E^{-2}$
$u$	$4.81734 E^{-1}$	$6.17128 E^{-2}$
$c_i$	1.422235	$1.27450 E^{-1}$

We used the **combined1** residual errors in MONOLIX. The observation models are then given by:

$$\log_{10}(V) = \log_{10}(V) + (a_1 + b_1 \log_{10}(V)) * \epsilon_1 \quad (6.4a)$$

$$\log_{10}(A) = \log_{10}(A) + (a_2 + b_2 \log_{10}(A)) * \epsilon_1 \quad (6.4b)$$

where the parameters  $a_1, b_1, a_2$  and  $b_2$  are given in Table 6.5.

Table 6.5: Model's residual errors

Parameter	gp41	gp120
$a_1$	0.12555	0.11619
$b_1$	0.19882	0.251797
$a_2$	0.67285	0.58853
$b_2$	$2.22045 E^{-16}$	$2.22045 E^{-16}$

The results of the individual fit (Figure 6.5 & 6.6) look acceptable by visual inspection. However, fitting these results is not enough to invoke confidence. Next, we validated the model by looking at the confidence interval.

## 6.2 Model Validation

To validate the model, we look at the interval to check whether the observed data are compatible with the population prediction, taking into account the inter-individual variability (114). For each individual a prediction interval is computed based on multiple simulations with the population parameters and the design structure of the individual. In Figure 6.7

and Figure 6.8, the confidence intervals, 90%, of the populations are displayed alongside the data from the ten rhesus macaques. Both figures show the confidence intervals of the viral load as well as antibodies. On the two figures, both gp41 and gp120 antibodies fall within the 90% confidence intervals. The same cannot be said for the viral load. However, for the data that are not in the interval, they are on the border. There is an exception for the subjects RKg19 and RQc19.

For RKg19, at 24 weeks time, there is a data point that is completely outside the interval. We can, confidently, say that the point is an anomaly because we expect the viral load to be suppressed at this point. For RQc19, the consecutive data points at weeks 8, 12 and 14 are outside the 90% margin. These points are compatible with the poor individual fit for the subject in Figure 6.5a. While this poor fit is not very concerning when taking into account the rest of the animals and the antibodies fit, we, nevertheless, expanded the interval to 95% (Figure 6.9). All RQc19 data points are either included (Figure 6.9a) or on the border (Figure 6.9b) of the interval of the 95% confidence interval. We can therefore conclude that the overall population can be described by our model.

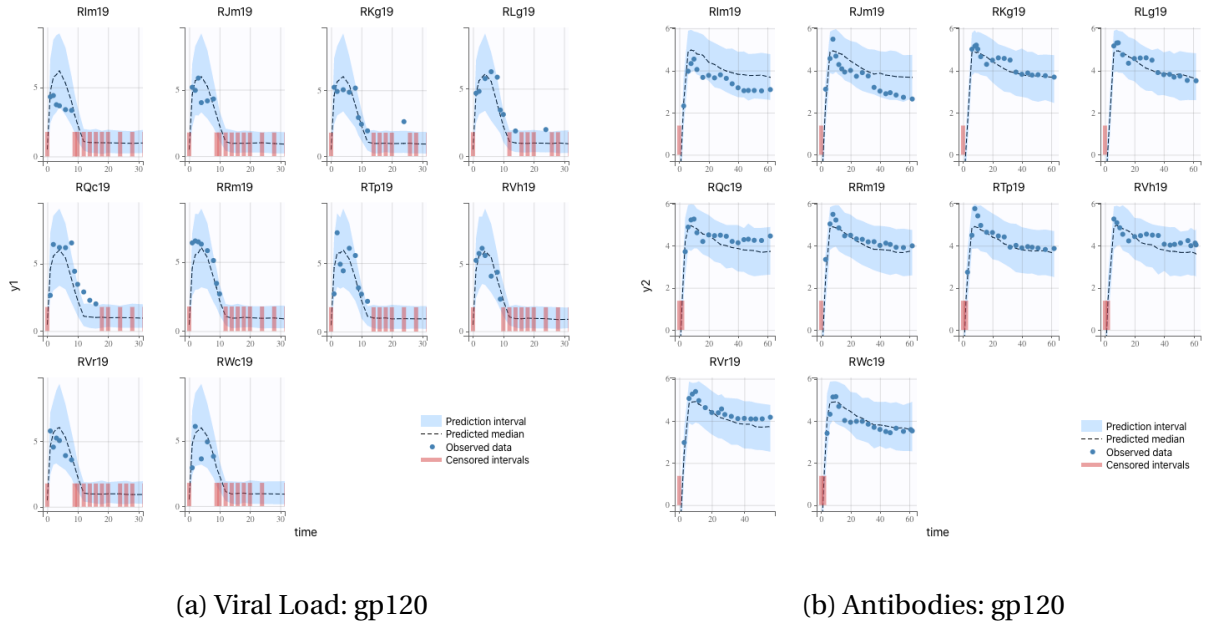
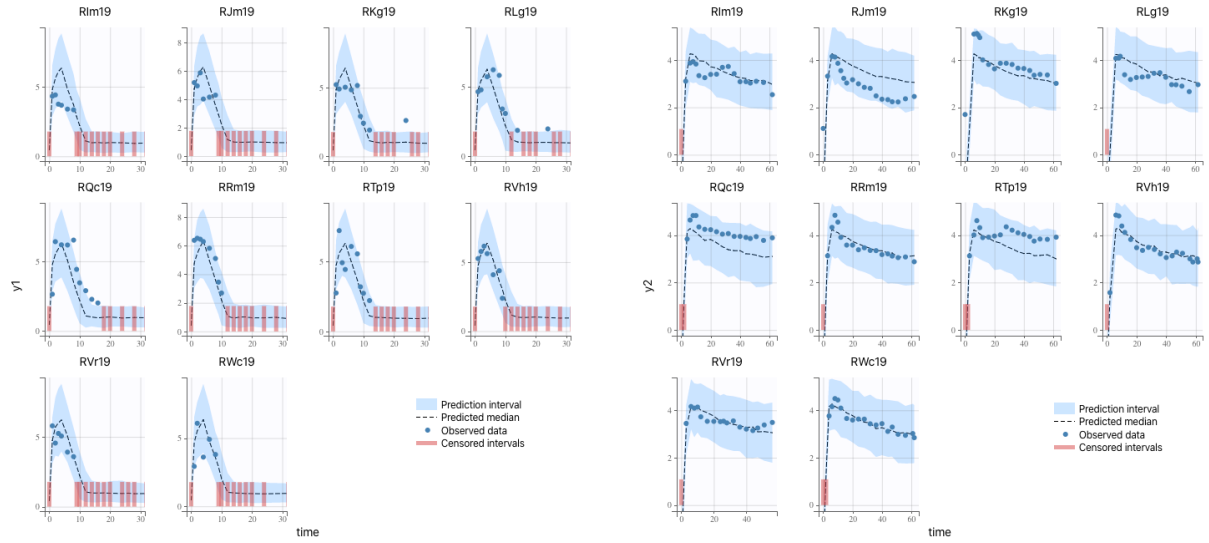


Figure 6.7: Model fit of viral Load and gp41 Antibodies on 90% distribution.

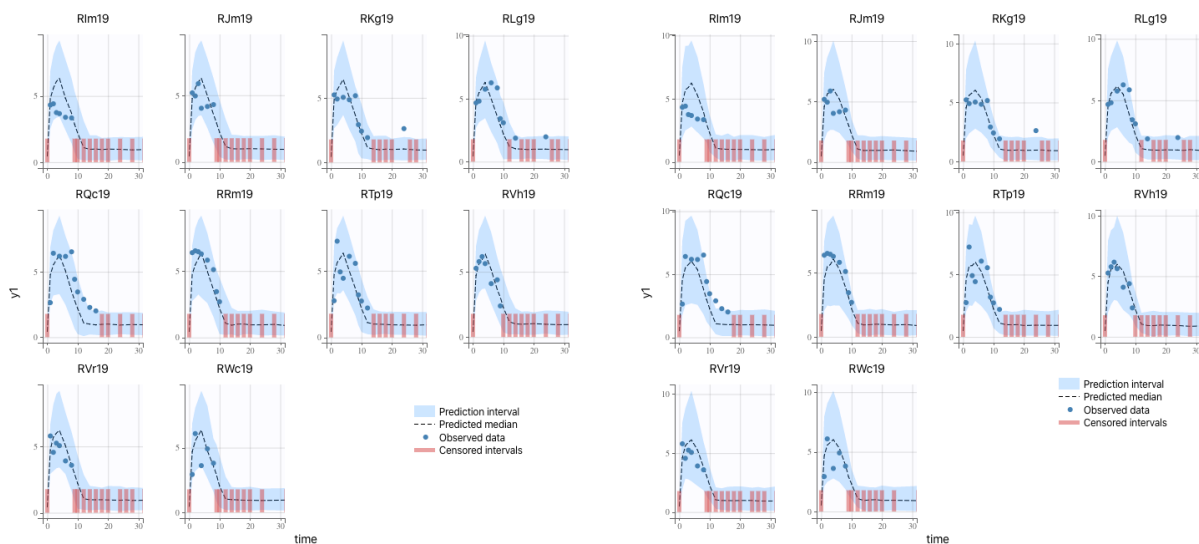


(a) Viral Load: gp41

(b) Antibodies: gp41

Figure 6.8: Model fit of viral Load and Antibodies on 90 % distribution.

We note that the main difference in Figure 6.9 is that Figure 6.9a is from the gp41 antibodies while Figure 6.9b is obtained using the parameters estimated on gp120 antibodies



(a) Viral Load: gp41

(b) Viral Load: gp120

Figure 6.9: Model fit of viral Load and Antibodies on 95 % distribution.

## 6.3 Our contributions

Using the rhesus macaques data, we showed that PK metrics ( $AUC$ ,  $C_{max}$ ,  $T_{max}$ , and terminal half-life) are statistically different. While the goal was also to compare both the intermediate ART and late ART to the early ART group, we deemed unnecessary to do it. It is because it is obvious from the data of early ART that the PK metrics are going to be significantly different from the other two groups. For gp120, five of ten rhesus monkeys did not develop any detectable antibodies. Of the five animals that developed gp120 antibodies, four of them had only one detectable data point. Essentially, only one subject developed gp120 antibodies when it was started on ART 4 days post-infection. In case of gp41, seven subjects had detectable antibodies, and five of them had four or more data points. There still was no need of performing a PK analysis as the values are visibly small, compared to the intermediate and late ART groups.

We developed a detailed HIV model that can simultaneously predict both the viral load and the natural antibodies (Figure 6.1, the blue part of the model is our contribution). The same model can be used for animals that started ART before 8 weeks, but the parameters will have to be re-estimated. For subjects that start ART after 8 weeks, our parameters



can be used as there is no much difference on antibodies concentration eight weeks-post infection. Our model could be modified (by introducing delayed differential equations) into a time-sensitive model where one could predict antibody responses regardless of the ART start time, that is the parameters remain with the exception of one parameter used by timing of ART.

The model was conceived with the desire if possible to be validated on human samples. The goal would be to predict how antibodies responses are developed in response to HIV infection in human babies. This is important because it will inform us when to give treatment to human infants for it to have the best beneficial effect in reducing the disease without crippling the immune response.

Along with our collaborators, we have finished writing the manuscript we intend to publish on the Pharmacokinetic (PK) Analysis of gp41 and gp120 antibodies. We are also working on writing another manuscript to accompany our results of Modeling Viral Load and Antibodies.

## REFERENCES

- [1] Douglas M. Hawkins. The problem of overfitting. *ChemInform*, 35(19), 2004.
- [2] Ashok Chaurasia and Ofer Harel. Using aic in multiple linear regression framework with multiply imputed data. *Health Services and Outcomes Research Methodology*, 12(2-3):219–233, 2012.
- [3] Manpreet Singh Minhas. Techniques for handling underfitting and overfitting in machine learning, Jun 2021.
- [4] David F. Findley and Ching-Zong Wei. Aic, overfitting principles, and the boundedness of moments of inverse matrices for vector autoregressions and related models. *Journal of Multivariate Analysis*, 83(2):415–450, 2002.
- [5] Shengping Yang and Gilbert Berdine. Model selection and model over-fitting. *The Southwest Respiratory and Critical Care Chronicles*, 3(12), 2015.
- [6] Sanjay.M. Why and how to cross validate a model?, Aug 2020.
- [7] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.
- [8] Adam R Attarian. *Patient Specific Subset Selection, Estimation and Validation of an HIV-1 Model with Censored Observations under an Optimal Treatment Schedule*. PhD thesis, North Carolina State University, 2012.
- [9] Gregory R. Carmichael, Adrian Sandu, and florian A. Potra. Sensitivity analysis for atmospheric chemistry models via automatic differentiation. *Atmospheric Environment*, 31(3):475–489, 1997.
- [10] George Bernard Lankford. *Optimization, modeling, and control: Applications to klystron designing and hepatitis C virus dynamics*. PhD thesis, North Carolina State University, 2016.
- [11] Christian Homescu. Adjoints and automatic (algorithmic) differentiation in computational finance, May 2011.

- [12] Sébastien Geeraert, Charles-Albert Lehalle, Barak A. Pearlmutter, Olivier Pironneau, and Adil Reghai. Mini-symposium on automatic differentiation and its applications in the financial industry. *ESAIM: Proceedings and Surveys*, 59:56–75, 2017.
- [13] J. Jr. E DENNIS and Robert B SCHNABEL. Front matter: Numerical methods for unconstrained optimization and nonlinear equations, 1983.
- [14] J. N. Lyness and C. B. Moler. Numerical differentiation of analytic functions. *SIAM Journal on Numerical Analysis*, 4(2):202–210, 1967.
- [15] J. N. Lyness. Numerical algorithms based on the theory of complex variable. *Proceedings of the 1967 22nd national conference on -*, 1967.
- [16] H. T. Banks, Kidist Bekele-Maxwell, Lorena Bociu, Marcella Noorman, and Kristen Tillman. The complex-step method for sensitivity analysis of non-smooth problems arising in biology, Oct 2015.
- [17] Joaquim R. Martins, Ilan Kroo, and Juan Alonso. An automated method for sensitivity analysis using complex variables. *38th Aerospace Sciences Meeting and Exhibit*, page 1–12, 2000.
- [18] William Squire and George Trapp. Using complex variables to estimate derivatives of real functions. *SIAM Review*, 40(1):110–112, 1998.
- [19] Harvey Thomas Banks and H. T. Tran. *Size-Structured Population Models/The Logistic Model*, page 247–249. CRC Press, Taylor & Francis Group, 2009.
- [20] Joseph G. Arthur, Hien T. Tran, and Philip Aston. Feasibility of parameter estimation in hepatitis c viral dynamics models. *Journal of Inverse and Ill-posed Problems*, 25(1), 2017.
- [21] Sergios Theodoridis, Aggelos Pikrakis, Konstantinos Koutroumbas, and Dionisis Cavouras. *Introduction to pattern recognition: a MATLAB approach*. Elsevier/Academic, 2010.
- [22] Nikki Shavon Smith. *A comparison of physiologically-based pharmacokinetic (PBPK) models of methyl-tertiary butyl ether (MTBE)*. PhD thesis, North Carolina State University, 2017.

- [23] Tom Quaiser and Martin Mönnigmann. Systematic identifiability testing for unambiguous mechanistic modeling – application to jak-stat, map kinase, and nf- $\kappa$  b signaling pathway models. *BMC Systems Biology*, 3(1), 2009.
- [24] Christoffer W. Tornøe, Henrik Agersø, E.niclas Jonsson, Henrik Madsen, and Henrik A. Nielsen. Non-linear mixed-effects pharmacokinetic/pharmacodynamic modelling in nlme using differential equations. *Computer Methods and Programs in Biomedicine*, 76(1):31–40, 2004.
- [25] Brett Matzuka, Jason Chittenden, Jonathan Monteleone, and Hien Tran. Stochastic nonlinear mixed effects: a metformin case study. *Journal of Pharmacokinetics and Pharmacodynamics*, 43(1):85–98, 2015.
- [26] Monolix 2018R1 user guide, url=<https://monolix.lixoft.com/single-page/>, journal=Monolix 2017, month=Aug, 2019.
- [27] Residual error model in monolix, url=<https://monolix.lixoft.com/data-and-models/errormodel/>, journal=Monolix 2017, month=Nov, 2020.
- [28] Min You and Gavin E. Arteel. Effect of ethanol on lipid metabolism. *Journal of Hepatology*, 70(2):237–248, Feb 2019.
- [29] William B. Ashworth, Nathan A. Davies, and I. David L. Bogle. A computational model of hepatic energy metabolism: Understanding zoned damage and steatosis in nafld. *PLOS Computational Biology*, 12(9), 2016.
- [30] Michelle M. Angrish, Charlene A. Mcqueen, Elaine Cohen-Hubal, Maribel Bruno, Yue Ge, and Brian N. Chorley. Editor’s highlight: Mechanistic toxicity tests based on an adverse outcome pathway network for hepatic steatosis. *Toxicological Sciences*, 159(1):159–169, 2017.
- [31] Michelle M. Angrish, Jonathan Phillip Kaiser, Charlene A. Mcqueen, and Brian N. Chorley. Tipping the balance: Hepatotoxicity and the 4 apical key events of hepatic steatosis. *Toxicological Sciences*, 150(2):261–268, 2016.
- [32] Michele Alves-Bezerra and David E Cohen. Triglyceride metabolism in the liver, Dec 2017.
- [33] Tilman Hahn, Konrad Botzenhart, and Fritz Schweinsberg. Toxic effects of solvent exposure, Mar 2019.

- [34] Rolf Teschke. Liver injury by carbon tetrachloride intoxication in 16 patients treated with forced ventilation to accelerate toxin removal via the lungs: A clinical report. *Toxics*, 6(2):25, 2018.
- [35] Robert D. Morrison, Brian L. Murphy, and Richard E. Doherty. Chlorinated solvents, Sep 2007.
- [36] Michael Dourson. Acsh explains: What’s the story on carbon tetrachloride?, Aug 2018.
- [37] Carbon tetrachloride, key=PubChem, url=https://pubchem.ncbi.nlm.nih.gov/compound/Carbon-tetrachloride, journal=National Center for Biotechnology Information. PubChem Compound Database, publisher=U.S. National Library of Medicine, 2022.
- [38] Obaid Faroon, Jessilynn Taylor, Nickolette Roney, Margaret E Fransen, Suzanne Bogaczyk, and Gary Diamond. Toxicological profile for carbon tetrachloride, Sep 2003.
- [39] Carbon tetrachloride, url=https://www.epa.gov/sites/default/files/2016-09/documents/carbon-tetrachloride.pdf, publisher=U.S. Environmental Protection Agency, 2000.
- [40] Role of modeling and simulation in scientific discovery, 2013.
- [41] Megan Elizabeth Sawyer. *Compartmentalizing the sunlight vitamin: Physiologically based pharmacokinetic modeling and Vitamin D*. PhD thesis, North Carolina State University, 2013.
- [42] Jennifer E Sager, Jingjing Yu, Isabelle Ragueneau-Majlessi, and Nina Isoherranen. Physiologically based pharmacokinetic (pbpk) modeling and simulation approaches: A systematic review of published models, applications, and model verification, Nov 2015.
- [43] Jane C. Caldwell, Marina V. Evans, and Kannan Krishnan. Cutting edge pbpk models and analyses: Providing the basis for future modeling efforts and bridges to emerging toxicology paradigms. *Journal of Toxicology*, 2012:1–10, 2012.
- [44] Xiaomei Zhuang and Chuang Lu. Pbpk modeling and simulation in drug research and development. *Acta Pharmaceutica Sinica B*, 6(5):430–440, 2016.
- [45] L Kuepfer, C Niederalt, T Wendl, J-F Schlender, S Willmann, J Lippert, M Block, T Eissing, and D Teutonico. Applied concepts in pbpk modeling: How to build a pbpk/pd model. *CPT: Pharmacometrics & Systems Pharmacology*, 5(10):516–531, 2016.

- [46] Thomas Hale and James Abbey. Drug transfer during breast-feeding. *Fetal and Neonatal Physiology*, 2017.
- [47] R. A. K. D. Thrall, M. E. Vucelick. Comparative metabolism of carbon tetrachloride in rats, mice, and hamsters using gas uptake and pbpk modeling. *Journal of Toxicology and Environmental Health, Part A*, 60(8):531–548, 2000.
- [48] Meinrad Boll, Lutz W.D. Weber, Eberhard Becker, and Andreas Stampfl. Hepatocyte damage induced by carbon tetrachloride: inhibited lipoprotein secretion and changed lipoprotein composition, 2001.
- [49] Takeshi Honma and Megumi Suda. Changes in plasma lipoproteins as toxicity markers for carbon tetrachloride, chloroform, and dichloromethane. *Industrial Health*, 35(4):519–531, 1997.
- [50] Yavar Mahmoodzadeh, Mohammad Mazani, and Lotfollah Rezagholizadeh. Hepatoprotective effect of methanolic tanacetum parthenium extract on ccl4-induced liver damage in rats, Aug 2017.
- [51] E. Becker, B. Messner And, and J. Berndt. Two mechanisms of cc14-induced fatty liver: Lipid peroxidation or covalent binding studied in cultured rat hepatocytes. *Free Radical Research Communications*, 3(1-5):299–308, 1987.
- [52] Sander Kersten. Mechanisms of nutritional and hormonal regulation of lipogenesis, Apr 2001.
- [53] Yong-Ping Zheng, Xian-Yang Zhong, Yu-Shu Huang, and Can-Bin Zheng. Hcbp6 is involved in the development of hepatic steatosis induced by high-fat diet and ccl4 in rats. *Annals of Hepatology*, 17(3):511–518, 2018.
- [54] AnroopB Nair and Shery Jacob. A simple practice guide for dose conversion between animals and human. *Journal of Basic and Clinical Pharmacy*, 7(2):27, 2016.
- [55] Van M. Savage, Eric J. Deeds, and Walter Fontana. Sizing up allometric scaling theory. *PLoS Computational Biology*, 4(9), 2008.
- [56] P.H. Stern, T. Furukawa, and T.M. Brody. Rat liver and plasma lipids after carbon tetrachloride administration. *Journal of Lipid Research*, 6(2):278–286, 1965.

- [57] Simon Gelman, David S. Warner, and Mark A. Warner. Venous function and central venous pressure. *Anesthesiology*, 108(4):735–748, 2008.
- [58] National Research Council. *Calories: Total Macronutrient Intake, Energy Expenditure, And Net Energy Stores*, pages 139–258, 273–290. National Academies Press, 1989.
- [59] How much water do you need to stay healthy?, url=<https://www.mayoclinic.org/healthy-lifestyle/nutrition-and-healthy-eating/in-depth/water/art-20044256>, journal=Mayo Clinic, publisher=Mayo Foundation for Medical Education and Research, month=Oct, 2020.
- [60] Shumin Qin, Yinjin Yin, and Keer Huang. Free fatty acids increase intracellular lipid accumulation and oxidative stress by modulating ppar $\alpha$  and srebp-1c in l-02 cells, 2016.
- [61] Xiu-Ping Bai, Feng Dong, Guo-Hua Yang, and Lei Zhang. Influences of sterol regulatory element binding protein-1c silencing on glucose production in hepg2 cells treated with free fatty acid - lipids in health and disease, Apr 2019.
- [62] Xia Chu, Liyan Liu, Lixin Na, Huimin Lu, Songtao Li, Ying Li, and Changhao Sun. Sterol regulatory element-binding protein-1c mediates increase of postprandial stearic acid, a potential target for improving insulin resistance, in hyperlipidemia, Feb 2013.
- [63] Margaret Sozio and David W. Crabb. Alcohol and lipid metabolism. *American Journal of Physiology-Endocrinology and Metabolism*, 295(1), Mar 2008.
- [64] Min You, Monika Fischer, Mark A. Deeg, and David W. Crabb. Ethanol induces fatty acid synthesis pathways by activation of sterol regulatory element-binding protein (srebp). *Journal of Biological Chemistry*, 277(32):29342–29347, May 2002.
- [65] Jinyao Liu. Ethanol and liver: Recent insights into the mechanisms of ethanol-induced fatty liver. *World Journal of Gastroenterology*, 20(40):14672, 2014.
- [66] Henry J. Pownall, Christie M. Ballantyne, Kay T. Kimball, Suzanne L. Simpson, Daniel Yeshurun, and Antonio M. Gotto. Effect of moderate alcohol consumption on hypertriglyceridemia. *Archives of Internal Medicine*, 159(9):981, May 1999.
- [67] Chiara Saponaro, Melania Gaggini, Fabrizia Carli, and Amalia Gastaldelli. The subtle balance between lipolysis and lipogenesis: A critical point in metabolic homeostasis, Nov 2015.

- [68] Angela K. Stoeckman and Howard C. Towle. The role of srebp-1c in nutritional regulation of lipogenic enzyme gene expression. *Journal of Biological Chemistry*, 277(30):27029–27035, 2002.
- [69] Jacob T Talley and Shamim S Mohiuddin. Biochemistry, fatty acid oxidation, Jan 2021.
- [70] James A. Hamilton. Fatty acid transport: Difficult or easy? *Journal of Lipid Research*, 39(3):467–481, 1998.
- [71] Robert W. Schwenk, Graham P. Holloway, Joost J.F.P. Luiken, Arend Bonen, and Jan E.C. Glatz. Fatty acid transport across the cell membrane: Regulation by fatty acid transporters. *Prostaglandins, Leukotrienes and Essential Fatty Acids (PLEFA)*, 82(4-6):149–154, 2010.
- [72] Michael W. King. *Lipids: Triglyceride and Phospholipid Synthesis*. McGraw-Hill Education, New York, NY, 2014.
- [73] Achim Lass, Robert Zimmermann, Monika Oberer, and Rudolf Zechner. Lipolysis - a highly regulated multi-enzyme complex mediates the catabolism of cellular fat stores, Jan 2011.
- [74] Robin E Duncan, Maryam Ahmadian, Kathy Jaworski, Eszter Sarkadi-Nagy, and Hei Sook Sul. Regulation of lipolysis in adipocytes, 2007.
- [75] Dominique Langin. Adipose tissue lipolysis as a metabolic pathway to define pharmacological strategies against obesity and the metabolic syndrome. *Pharmacological Research*, 53(6):482–491, 2006.
- [76] Gisela Wilcox. Insulin and insulin resistance, May 2005.
- [77] Andrew M. Freeman and Nicholas Pennings. Insulin resistance, Jul 2021.
- [78] Erik De Clercq. Anti-hiv drugs: 25 compounds approved within 25 years after the discovery of hiv. *International Journal of Antimicrobial Agents*, 33(4):307–320, 2009.
- [79] Michael H Merson, Jeffrey O'Malley, David Serwadda, and Chantawipa Apisuk. The history and challenge of hiv prevention. *The Lancet*, 372(9637):475–488, 2008.
- [80] Mary Masters and Beau Ances. Role of neuroimaging in hiv-associated neurocognitive disorders. *Seminars in Neurology*, 34(01):089–102, 2014.



- [81] Jacqueline Rosenthal and William Tyor. Aging, comorbidities, and the importance of finding biomarkers for hiv-associated neurocognitive disorders. *Journal of NeuroVirology*, 25(5):673–685, 2019.
- [82] Yunda Huang, Lily Zhang, Julie Ledgerwood, Nicole Grunenber, Robert Bailer, Abby Isaacs, Kelly Seaton, Kenneth H. Mayer, Edmund Capparelli, Larry Corey, and et al. Population pharmacokinetics analysis of vrc01, an hiv-1 broadly neutralizing monoclonal antibody, in healthy adults. *mAbs*, 9(5):792–800, 2017.
- [83] Caroline A. Sabin and Jens D. Lundgren. The natural history of hiv infection. *Current Opinion in HIV and AIDS*, page 1, 2013.
- [84] Esteban A. Hernandez-Vargas and Richard H. Middleton. Modeling the three stages in hiv infection. *Journal of Theoretical Biology*, 320:33–40, 2013.
- [85] Richard D. Moore and Richard E. Chaisson. Natural history of hiv infection in the\_era of combination antiretroviral therapy. *AIDS*, 13(14):1933–1942, 1999.
- [86] Lucia Palmisano and Stefano Vella. A brief history of antiretroviral therapy of hiv infection: Success and challenges, 2011.
- [87] Lorna Leal, Csaba Fehér, Valèria Richart, Berta Torres, and Felipe García. Antiretroviral therapy interruption (ati) in hiv-1 infected patients participating in therapeutic vaccine trials: Surrogate markers of virological response. *Vaccines*, 8(3):442, 2020.
- [88] Tyler R. Kemnic and Peter G. Gulick. *HIV Antiretroviral Therapy*. StatPearls Publishing, Treasure Island (FL), 2021.
- [89] Wei-Ti Chen, Cheng-Shi Shiu, Joyce P Yang, Jane M Simoni, Karen I Fredriksen-Goldsen, Tony Szu-Hsien Lee, and Hongxin Zhao. Antiretroviral therapy (art) side effect impacted on quality of life, and depressive symptomatology: A mixed-method study. *Journal of AIDS Clinical Research*, 04(06), 2013.
- [90] A. Oxenius, D. A. Price, H. F. Gunthard, S. J. Dawson, C. Fagard, L. Perrin, M. Fischer, R. Weber, M. Plana, F. Garcia, and et al. Stimulation of hiv-specific cellular immunity by structured treatment interruption fails to enhance viral control in chronic hiv infection. *Proceedings of the National Academy of Sciences*, 99(21):13747–13752, 2002.

- [91] Jody Lawrence, Douglas L. Mayers, Katherine Huppler Hullsiek, Gary Collins, Donald I. Abrams, Ronald B. Reisler, Lawrence R. Crane, Barry S. Schmetter, Thomas J. Dionne, Jennifer M. Saldanha, and et al. Structured treatment interruption in patients with multidrug-resistant human immunodeficiency virus. *New England Journal of Medicine*, 349(9):837–846, 2003.
- [92] Lidia Ruiz, Javier Martinez-Picado, Joan Romeu, Roger Paredes, Mohd Khalil Zayat, Silvia Marfil, Eugenia Negredo, Guillem Sirera, Cristina Tural, Bonaventura Clotet, and et al. Structured treatment interruption in chronically hiv-1 infected patients after long-term viral suppression. *AIDS*, 14(4):397–403, 2000.
- [93] Rajat Desikan, Rubesh Raja, and Narendra M. Dixit. Early exposure to broadly neutralizing antibodies may trigger a dynamical switch from progressive disease to lasting control of shiv infection. *PLOS Computational Biology*, 16(8), 2020.
- [94] Ilesh V. Jani, Bindiya Meggi, Osvaldo Loquiha, Ocean Tobaiwa, Chishamiso Mudenyanaga, Alcina Zitha, Dadirayi Mutsaka, Nedio Mabunda, Adolfo Vubil, Timothy Bollinger, and et al. Effect of point-of-care early infant diagnosis on antiretroviral therapy initiation and retention of patients. *AIDS*, 32(11):1453–1463, 2018.
- [95] Mark F Cotton, Avy Violari, Kennedy Otworld, Ravindre Panchia, Els Dobbels, Helena Rabie, Deirdre Josipovic, Afaaf Liberty, Erica Lazarus, Steve Innes, and et al. Early time-limited antiretroviral therapy versus deferred therapy in south african infants infected with hiv: Results from the children with hiv early antiretroviral (cher) randomised trial. *The Lancet*, 382(9904):1555–1563, 2013.
- [96] Leigh F Johnson, Mary-Ann Davies, Harry Moultrie, Gayle G. Sherman, Ruth M. Bland, Thomas M. Rehle, Rob E. Dorrington, and Marie-Louise Newell. The effect of early initiation of antiretroviral treatment in infants on pediatric aids mortality in south africa. *Pediatric Infectious Disease Journal*, 31(5):474–480, 2012.
- [97] Margaret Mcmanus, Jennifer Henderson, Anita Gautam, Robin Brody, Eric R Weiss, Deborah Persaud, Eric Mick, and Katherine Luzuriaga. Quantitative hiv-1 antibodies correlate with plasma hiv-1 rna and cell-associated dna levels in children on art. *Clinical Infectious Diseases*, 2018.
- [98] Paolo Palma, Margaret Mcmanus, Nicola Cotugno, Salvatore Rocca, Paolo Rossi, and

- Katherine Luzuriaga. The hiv-1 antibody response: a footprint of the viral reservoir in children vertically infected with hiv. *The Lancet HIV*, 7(5), 2020.
- [99] Jessica Haberer and Claude Mellins. Pediatric adherence to hiv antiretroviral therapy. *Current HIV/AIDS Reports*, 6(4):194–200, 2009.
- [100] Jürgen B. Bulitta and Nicholas H. G. Holford. Non-compartmental analysis. *Wiley StatsRef: Statistics Reference Online*, 2014.
- [101] Johan Gabrielsson and Daniel Weiner. Non-compartmental analysis. *Methods in Molecular Biology Computational Toxicology*, page 377–389, 2012.
- [102] Brandon Burch. What is noncompartmental pk analysis?, Nov 2021.
- [103] Yi Rang Han, Ping I. Lee, and K. Sandy Pang. Finding tmax and cmax in multicompartmental models. *Drug Metabolism and Disposition*, 46(11):1796–1804, 2018.
- [104] Jericho Hallare and Valerie Gerriets. Half life, Aug 2021.
- [105] Jeremy D. Scheff, Richard R. Almon, Debra C. DuBois, William J. Jusko, and Ioannis P. Androulakis. Assessment of pharmacologic area under the curve when baselines are variable. *Pharmaceutical Research*, 28(5):1081–1089, 2011.
- [106] Sinead E. Morris, Luise Dziobek-Garrett, Renate Strehlau, Juliane Schröter, Stephanie Shiau, Anet J. N. Anelone, Maria Paximadis, Rob J. De Boer, Elaine J. Abrams, Caroline T. Tiemessen, and et al. Quantifying the dynamics of hiv decline in perinatally infected neonates on antiretroviral therapy. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 85(2):209–218, 2020.
- [107] Daniel B. Reeves, Yunda Huang, Elizabeth R. Duke, Bryan T. Mayer, E. Fabian Cardozo-Ojeda, Florencia A. Boshier, David A. Swan, Morgane Rolland, Merlin L. Robb, John R. Mascola, and et al. Mathematical modeling to reveal breakthrough mechanisms in the hiv antibody mediated prevention (amp) trials. *PLOS Computational Biology*, 16(2), 2020.
- [108] Adam John Mackridge and Philip Rowe. *One-Way Analysis of Variance (ANOVA) – Including Dunnett’s and Tukey’s Follow Up Tests*, page 93–103. Wiley, 2018.
- [109] Stephanie Glen. Kruskal wallis h test: Definition, examples & assumptions, May 2021.

- [110] Akiko Murase, Toru Sasaki, and Tsuyoshi Kajiwaru. Stability analysis of pathogen-immune interaction dynamics. *Journal of Mathematical Biology*, 51(3):247–267, 2005.
- [111] Martin A. Nowak and Charles R. M. Bangham. Population dynamics of immune responses to persistent viruses. *Science*, 272(5258):74–79, 1996.
- [112] B Alberts, J Lewis, and A Johnson. The adaptive immune system, Jan 1970.
- [113] Shiferaw Feyissa and Sandip Banerjee. Role of antibodies: A novel paradigm in mathematical modeling for cancer treatment. *Bulletin of Mathematical Biology*, Jun 2010.
- [114] Individual fits in monolix, url=<https://monolix.lixoft.com/graphics/individual-fits/>, journal=Monolix 2017, month=May, 2018.
- [115] Boseung Choi, Grzegorz A. Rempala, and Jae Kyoung Kim. Beyond the michaelis-menten equation: Accurate and efficient estimation of enzyme kinetic parameters. *Scientific Reports*, 7(1), 2017.
- [116] Kenneth A. Johnson and Roger S. Goody. The original michaelis constant: Translation of the 1913 michaelis-menten paper. *Biochemistry*, 50(39):8264–8269, 2011.
- [117] Ute Deichmann, Stefan Schuster, Jean-Pierre Mazat, and Athel Cornish-Bowden. Commemorating the 1913 michaelis-menten paper die kinetik der invertinwirkung: Three perspectives. *FEBS Journal*, 281(2):435–463, 2013.
- [118] Ronald P. Brown, Michael D. Delp, Stan L. Lindstedt, Lorenz R. Rhomberg, and Robert P. Beliles. Physiological parameter values for physiologically based pharmacokinetic models. *Toxicology and Industrial Health*, 13(4):407–484, 1997.
- [119] Michael L. Gargas, Melvin E. Andersen, and Harvey J. Clewell. A physiologically based simulation approach for determining metabolic constants from gas uptake data. *Toxicology and Applied Pharmacology*, 86(3):341–352, 1986.
- [120] Cécil J.w.meulenberg and Henk P.m. Vijverberg. Empirical relations predicting human and rat tissue: Air partition coefficients of volatile organic compounds. *Toxicology and Applied Pharmacology*, 169(2):208, 2000.
- [121] Piotr Zabielski, Agnieszka Urszula Błachnio-Zabielska, Beata Wójcik, Adrian Chabowski, and Jan Górski. Effect of plasma free fatty acid supply on the rate of ceramide synthesis in different muscle types in the rat. *PLOS ONE*, 12(11), 2017.

- [122] Hidayaturrahmah, Heri Budi Santoso, Ridha Aulia Rahmi, and Dewi Kartikasari. Blood glucose level of white rats (*rattus norvegicus*) after giving catfish biscuit (*pangasius hypothalmus*). *BIO Web of Conferences*, 20:04005, 2020.
- [123] Genichi Watanabe, Toshiaki Ishiguro, Hisayuki Miura, Kazumasa Uemura, Yasuo Hiyoshi, Kuniaki Ozawa, Katsunori Nonogaki, Tatsuo Tamagawa, Nobuo Sakamoto, Akihisa Iguchi, and et al. Cns regulation of blood lactate concentration in anesthetized rats. *Life Sciences*, 54(20):1491–1499, 1994.

## **APPENDIX**

## APPENDIX

### A

## QUANTITATIVE SYSTEM TOXICOLOGY

### A.1 Physiological based pharmacokinetic (PBPK)

#### A.1.1 Michaelis-Menten Kinetics

For all the PBPK models, we use enzyme kinetics to describe the metabolism process in the liver. In the Michaelis-Menten Kinetics, free enzyme (E) reversibly binds with the substrate (S) to form the complex (ES), and the complex irreversibly dissociates into the product (P) and the free enzyme (E) (115). The reactions are represented by Equation



where  $k_1$  is the rate in which the enzyme and substrate react to form the complex  $ES$ ,  $k_{-1}$  is the rate in which the complex disassociate to convert back to the initial substrate (S) and enzyme (E), and  $k_2$  is the rate of the complex formation into the product  $P$ .

In 1913, scientists Leonor Michaelis and Maud Menten (116) (117) published their work on a mechanism for the catalysis of chemical reactions in biological systems. Michaelis and Menten described the mechanism of enzyme-catalyzed reactions and gave a relationship

between the reaction rates and the concentrations of enzyme and substrate (22). The model relates the enzymatic reactions rate,  $V$  to the concentration of a substrate  $S$ , as

$$v(S) = \frac{V_{max}S}{K_m + S}, \quad (A.2)$$

where  $V_{max}$  is the maximum production rate of  $P$  and  $K_m$  is Michaelis-Menten constant, the affinity of the enzyme for the substrate, defined as

$$K_m = \frac{k_{-1} + k_2}{k_1}. \quad (A.3)$$

(A.2) is used for the liver metabolite rate of the PBPK models. In the expression  $\frac{\text{met}_{V_{max}} X_{Lv}}{V_{Lv} P_{Lv}}$  in (2.5a),  $\text{met}_{V_{max}}$  is the maximum production,  $V_{max}$  in A.2,  $\text{met}_{K_m}$  is the affinity of the enzyme for the substrate  $K_m$ , and finally the substrate  $S$  is presented by the concentration  $\frac{X_{Lv}}{V_{Lv} P_{Lv}}$ .

### A.1.2 Non-compartment PBPK equations

The equation for the total blood flow (venous blood flow) is

$$Q = Q_B + Q_F + Q_K + Q_{Lv} + Q_{rp} + Q_{sp} + Q_M, \quad (A.4)$$

where the  $Q_i$ 's are the blood flow through the compartment of Figure 2.4 that go into the the veins.

The Arterial blood concentration  $C_a$  is given by

$$C_a = \frac{X_{Lg}}{V_{Lg} P_{Lg}} \quad (A.5)$$

The venous concentration  $C_v$  is given by

$$C_v = \frac{1}{Q} \left( \frac{Q_B X_B}{V_B P_B} + \frac{Q_F X_F}{V_F P_F} + \frac{Q_K X_K}{V_K P_K} + \frac{Q_{Lv} X_{Lv}}{V_{Lv} P_{Lv}} + \frac{Q_{rp} X_{rp}}{V_{rp} P_{rp}} + \frac{Q_{sp} X_{sp}}{V_{sp} P_{sp}} + \frac{Q_M X_M}{V_M P_M} \right), \quad (A.6)$$



where  $Q$  is given by A.4.

### A.1.3 Common Compartments PBPK equations

$$\dot{B} = Q_B * \left( C_a - \frac{X_B}{V_B * P_B} \right) \quad (\text{A.7a})$$

$$\dot{F} = Q_F * \left( C_a - \frac{X_F}{V_F * P_F} \right) \quad (\text{A.7b})$$

$$\dot{K} = Q_K * \left( C_a - \frac{X_K}{V_K * P_K} \right) - \frac{X_K}{V_K} * K_{\text{Ext}} \quad (\text{A.7c})$$

$$\dot{L}v = (Q_{Lv} - Q_G) * C_a + \frac{Q_G X_G}{V_G P_G} - \frac{Q_{Lv} X_{Lv}}{V_{Lv} P_{Lv}} - \frac{\frac{\text{met}v_{max} X_{Lv}}{V_{Lv} P_{Lv}}}{\frac{\text{met-km} + X_{Lv}}{V_{Lv} P_{Lv}}} \quad (\text{A.7d})$$

$$\dot{M} = Q_b * \left( C_a - \frac{X_m}{V_m * P_m} \right) \quad (\text{A.7e})$$

$$\dot{R}P = Q_{rp} * \left( C_a - \frac{X_{rp}}{V_{rp} * P_{rp}} \right) \quad (\text{A.7f})$$

$$\dot{S}P = Q_{sp} * \left( C_a - \frac{X_{sp}}{V_{sp} * P_{sp}} \right) \quad (\text{A.7g})$$

$$\dot{M}et = \frac{\frac{\text{met}v_{max} X_{Lv}}{V_{Lv} P_{Lv}}}{\frac{\text{met-km} + X_{Lv}}{V_{Lv} P_{Lv}}} \quad (\text{A.7h})$$

$$\dot{E}xt = \frac{X_k}{V_k} * K_{\text{Ext}} \quad (\text{A.7i})$$

$$\dot{L}m = -k_a * X_{Lm} \quad (\text{A.7j})$$

#### A.1.4 PBPK Equations acronyms

Table A.1: Equations acronyms

Abbreviation	Meaning
$Lg$	Lungs
$B$	Brain
$F$	Fat
$K$	Kidney
$Lv$	Liver
$M$	Muscles
$RP$	Rapidly Perfused
$SP$	Slowly Perfused
$G$	Gut
$L_m$	Lumen
$Met$	Liver Metabolite
$Ext$	Kidney Metabolite
$Q$	Total Blood flow
$C_a$	Arterial Concentration

### A.1.5 PBPK parameters

Table A.2: Percent of the organs' volume (in L)

Organ	Rats	Human
Lungs	0.0050 <sup>b</sup>	0.0076 <sup>b</sup>
Brain	0.0057 <sup>b</sup>	0.02000 <sup>b</sup>
Fat	0.0080 <sup>a</sup>	0.1000 <sup>a</sup>
Kidney	0.0073 <sup>b</sup>	0.0044 <sup>b</sup>
Liver	0.0366 <sup>b</sup>	0.0400 <sup>a</sup>
Muscles	0.4043 <sup>b</sup>	0.4000 <sup>b</sup>
Rapidly perfused	0.050 <sup>a</sup>	0.050 <sup>a</sup>
Slowly perfused	0.3584	0.3609
Gut	0.0270 <sup>b</sup>	0.0171 <sup>b</sup>

<sup>a</sup> Parameters from (47)

<sup>b</sup> Parameters from (118)

Slowly perfused is obtained by subtracting the sum of all organs from 1.

Table A.3: Percent of rate flow to the organs (in L/hr)

Organ	Rats	Human
Lungs	1.0000	1.000
Brain	0.0200 <sup>b</sup>	0.114 <sup>b</sup>
Fat	0.0400 <sup>a</sup>	0.060 <sup>a</sup>
Kidney	0.0141 <sup>b</sup>	0.175 <sup>b</sup>
Liver	0.183 <sup>b</sup>	0.277 <sup>b</sup>
Muscles	0.278 <sup>b</sup>	0.191 <sup>b</sup>
Rapidly perfused	0.0080	0.0280
Slowly perfused	0.1900 <sup>a</sup>	0.0800 <sup>a</sup>
Gut	0.1400 <sup>b</sup>	0.125 <sup>b</sup>

<sup>a</sup> Parameters from (47)

<sup>b</sup> Parameters from (118)

Rapidly perfused is obtained by subtracting the sum of all organs from 1.

Table A.4: Parameters used in the PBPK model for Carbon tetrachloride in both rats and humans

Parameter	Rats	Human
Cardiac Output (L/hr)	5.4 <sup>a</sup>	348 <sup>a</sup>
Body weight (L)	0.250 <sup>a</sup>	70 <sup>a</sup>
Urine extraction (L/hr)	0.8850	0.0.0208
Absorption constant	0.4871 <sup>d</sup>	0.4871 <sup>d</sup>
metabolites-vmax (mg/hr)	1.4632 <sup>d</sup>	1.4632 <sup>d</sup>
metabolites-km (mg/L)	0.275 <sup>b</sup>	0.275 <sup>b</sup>

<sup>a</sup> Parameters from (47)

<sup>b</sup> Parameters values chosen to be close (119)

<sup>d</sup> Parameters were estimated via **fmincon** in MATLAB

Table A.5: Partition coefficient tissue:air

Organ	Rats	Human
Blood	4.5200	3.16
Brain	20.09	8.73
Fat	359	177
Kidney	37.5	5.04
Liver	14.2	11.5
Muscles	4.57	6.43
Rapidly perfused	14.2	11.5
Slowly perfused	4.57	6.43
Gut	14.2	11.5

All the parameters are from (120).

The partition coefficient of the gut and rapidly perfused tissues were assumed to be the same as the liver, while that of the slowly perfused organs was assumed to be equal to the muscles.

## A.2 Liver Model

### A.2.1 Tables of Parameters that were changed

Table A.6: Parameters used in the liver model and their corresponding values

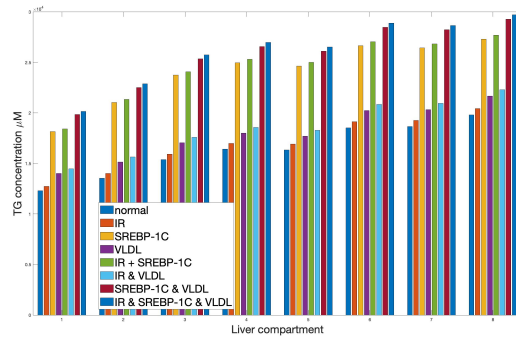
Parameters	Parameter definition	Human Values	Rats Values	Source
bf	Rate of blood flow	1.2 ( $\text{sec}^{-1}$ )	1.169 ( $\text{sec}^{-1}$ )	<b>fmincon</b>
s	Rest of body to hepatic compartment ratio	40	26.003	<b>fmincon</b>
timeref	Spiked diet inputs period	28800	$6.87 * 10^4 \text{ sec}$	<b>fmincon</b>
driveG	glucose diet	19.275	$19.266 \frac{\mu\text{M}}{\text{sec}}$	<b>fmincon</b>
driveF	fat diet	3.5	$3.719 \frac{\mu\text{M}}{\text{sec}}$	<b>fmincon</b>

Table A.7: Parameters used as initial conditions in the liver model and their corresponding values

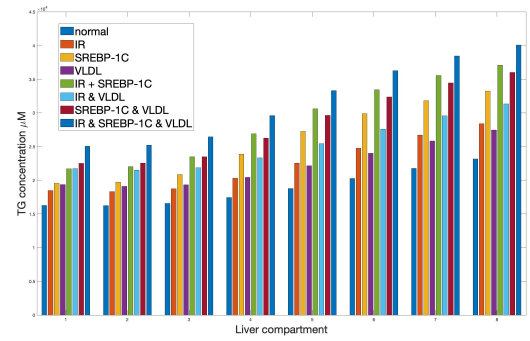
Parameter	Compartment number	Human values	Rats values ( $\mu M$ )	Source for rats values( $\mu M$ )
FFA	1-9	$5.00 * E^2$	$1.50 * E^2$	(121)
gCB	1-9	$5.000 * E^3$	$6.88 * E^2$	(122)
Lac	1-9	$1.000 * E^3$	1	(123)
Glucose	1-8	$2.0000 * E^4$	$1.80 * E^5$	parameter estimation
Triglycerides	1	$1.6243 * E^4$	$1.1622 * E^4$	parameter estimation
Triglycerides	2	$1.6261 * E^4$	$1.1831 * E^4$	parameter estimation
Triglycerides	3	$1.6537 * E^4$	$1.3215 * E^4$	parameter estimation
Triglycerides	4	$1.7251 * E^4$	$1.3450 * E^4$	parameter estimation
Triglycerides	5	$1.8303 * E^4$	$1.2118 * E^4$	parameter estimation
Triglycerides	6	$1.9505 * E^4$	$1.4792 * E^4$	parameter estimation
Triglycerides	7	$2.0716 * E^4$	$1.4263 * E^4$	parameter estimation
Triglycerides	8	$2.1835 * E^4$	$1.5478 * E^4$	parameter estimation
G6P	1-8		2.3832	body weight scaling



## A.2.2 Plots of Liver compartments



(a) Rats



(b) Humans

Figure A.1: Plot of the concentrations of hepatocytes triglycerides in rats and humans.