12-2016

Comparison of Two Methods for Developing Aggregate Population-Based Models

Oyebola E. Oyero
East Tennessee State University

Comparison of Two Methods for Developing Aggregate Population-Based Models

---

A thesis

presented to

the faculty of the Department of Mathematics

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Science in Mathematical Sciences

---

by

Oyebola Oyero

December 2016

Michele Joyner, Ph.D.

Jeff Knisley, Ph.D.

Edith Seier, Ph.D.

Keywords: Aggregate models, Nonlinear Mixed effect model, population models

ABSTRACT

Comparison of Two Methods for Developing Aggregate Population-Based Models

by

Oyebola Oyero

Aggregate models incorporate the variation between individual parameters of individual-based models to construct a population-based model. This thesis focuses on the comparison of two different methods for creating these population-based models. The first method, the individual parameter distribution technique (IPD) focuses on the similarities and variation of parameters in an individual-based model as calculated using individual data sets [4]. The second method we consider is the nonlinear mixed effect method (NLME), which is primarily used in modeling repeated measurement data. In the NLME approach, both the fixed effects and random effects of the parameter values are estimated in the model by assuming a normal distribution for the parameter values across individuals[1]. Using the variation in parameters estimated using the two different approaches, a population model was generated and then compared to the dynamics seen in the individual data sets. We compare three features of the concentration data to the simulated population models. The values for all three features were captured by both methods; however, the biggest difference observed is that there is a longer tail in the distribution for the population model developed using NLME than observed in the dynamics in the original data.

DEDICATION

This thesis is dedicated to God and the entire Oyero family.

ACKNOWLEDGEMENTS

I am grateful to God, who counted me among the living to be able to achieve this degree. I would like to express the deepest appreciation to my thesis supervisor, Dr. Michele Joyner, whose experience, understanding, kind guidance and support made it possible for me to work on a topic that is of great importance in mathematical modeling.

I am grateful to the other members of my thesis committee, Dr. Jeff Knisley and Dr. Edith Seier, for taking their time to be part of this milestone achievement in my academic career. My sincere gratitude goes to Dr. Robert Gardner, the Math department graduate coordinator for the role he played in my entire stay in the department of Mathematics and Statistics. I would also like to thank the Chair of the department, Professor Robert Price for the kind accommodation and support given to the international students. I am thankful to all my math colleagues, family and all my students that I taught during my graduate assistant obligation in the department, God bless you all.

TABLE OF CONTENTS

LIST OF FIGURES

1 INTRODUCTION

Mathematical modelers are faced with the challenges of modeling systems such as pharmacokinetics, ecology, dairy science, etc. in which the underlying procedure involves the collection of individualized data which must then be used to design a model which can capture the dynamics of the population[1]. For example, subjects in pharmacokinetic experiments usually give a series of blood samples after being administered a test agent; this is what we refer to as individualized data. For each individual, one may formulate a mathematical model describing the concentration of the test agent within the *individual* across time. Aggregate models introduce an approach which is built on either using these individual-based models or simply the individual data to develop a model for the population as a whole. In this thesis, we compare the results of two different methods for formulating an aggregate model: individual parameter distribution technique (IPD)[4] and the nonlinear mixed effect method (NLME)[1]. The two methods both use contributions from individuals to build a population model, but they use the individual attributes differently. In the IPD approach, it is assumed that there is an underlying unknown probability distribution across the model parameter values for each individual which must be estimated. Furthermore, this method assumes correlation may exist between parameter values which must be taken into account. The NLME method, also known as a hierarchical model, is the second method we considered. Unlike the IPD method, the NLME method assumes the parameter values have a normal distribution across the population and measures both the fixed and random effects. In both approaches, the assumption of the two methods we consider is based on the fact that individuals

behave differently and hence are modeled using different parameter values. We implement both methods on a pharmacokinetic model to determine the differences in the resulting population model.

Pharmacokinetic models require the body to be represented as a system of compartments. In pharmacokinetics, a compartment is defined as a group of tissues that have a similar blood flow and drug affinity [6]. Hence in compartment pharmacokinetic modeling, we assume that the rate of transfer between compartments is in the form of a first-order differential equation[7]. The solution of these differential equations gives a formal mathematical description of the concentration in the compartments at any time as a function of the parameter values.

In this thesis, we consider the simplest compartment model, a single compartment model, for the concentration of a drug in the body. In Section 2, we introduce the specific mathematical model and data we will use throughout this thesis. In Section 3, we will examine the implementation of both methods on the concentration model. In Section 4, we compare the results of each population model by considering three aspects of the individual data: the area under the curve (AUC), peak concentration and final concentration. We conclude with some final remarks and comments about future work in Section 5.

2 MODEL AND DESCRIPTION

In this section we will describe the basic model and the type of data used in this thesis. Moreover, we shall a give brief illustration of an individual-based (IBM) and a population model.

2.1 Individual and a Population Based Model

We can define both an individual and population based model according to their dynamics.

Definition 2.1 *Individual based models are simulations based on the global consequence of local interaction of members of a population*[5].

Definition 2.2 *A population model is a model that allows a better understanding of complex interactions among individual models.*[2]

In the former, the characteristics of each individual are typically measured as time varies. The latter relays a better understanding of variations possible among individuals. The first-order one compartment model that we shall describe shortly is a typical example of an IBM. In this example, the concentration of the blood is measured at various times throughout the time period for multiple subjects given some oral dose of a test agent. A population based model would not mimic the concentration levels in a specific individual as in IBM; instead, the population model determines the trend possible across an entire population.

2.2 Example Model

In this thesis, we applied the IPD technique and NLME modeling to a pharmacokinetic process where the concentration of drugs is measured according to a specific oral dose administered to different subjects. The one-compartment and two-compartment models are the most common compartment models in pharmacokinetic modeling [7]. In our case the one-compartment model was considered where the body is represented as single body compartment.

In the one-compartment model, an orally administered drug will usually flow through the compartment just as shown in Figure 1. Let t be the time following the drug administration. At the initial time $t = 0$, an oral dose D is immediately delivered into the blood stream from an arbitrary absorption location, e.g the stomach, resulting in a drug concentration $C_a(t)$ measured at time t [7]. It is assumed that drugs enter the compartment at absorption rate k_a and leave at elimination rate given by $k_e = Cl/V$. Cl is the clearance rate and V is the volume of the blood that is diluted by the dose at the absorption location.

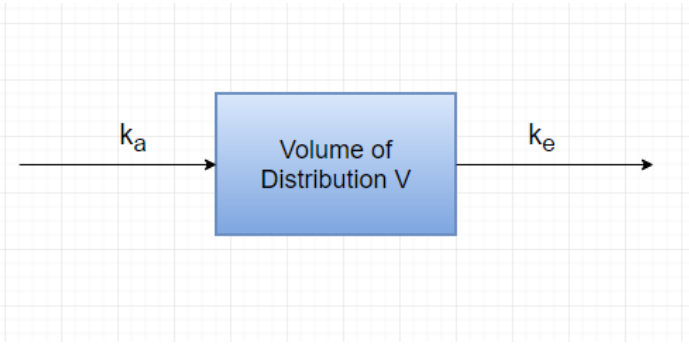

Figure 1: One-compartment model with first-order absorption and elimination.

Assuming first-order linear kinetics, the drug concentrations at the absorption lo-

cation and in the blood are denoted as $[C_a(t), C(t)]^T$ respectively and can be described by the following linear system of differential equations[7]:

$$\begin{aligned} \frac{dC_a(t)}{dt} &= -k_a C_a(t), & C_a(0) &= \frac{D}{V} \\ \frac{dC_a(t)}{dt} &= -k_a C_a(t) - k_e C(t), & C(0) &= 0. \end{aligned} \tag{1}$$

The solution of the above linear system of differential equations is given by Equation (2),

$$C(t) = \frac{Dk_a}{V(k_a - Cl/V)} \left\{ \exp\left(-k_a t\right) - \exp\left(\frac{-Cl}{V} t\right) \right\}. \tag{2}$$

2.3 Data Description

Individual-based models usually involves getting repeated-measure data on a number of subjects over a given time range. The data considered in this work was obtained from the repeated measurement of blood concentration carried out after an anti-asthmatic drug (theophylline) was administered orally to 12 subjects[3]. The data can be found with the name **Theoph** using the **R** package **nlme**[7]. The data sets contains 5 columns as follows: the subjects, the time since the administration of the drug, the concentration of the drug administered to the subjects, the doses administered to each subject and the weight of each subject. As can been seen in Figure 2, the concentration-time profiles follow the same trend for all the 12 subjects: the concentrations rise gradually before reaching a maximum (peak) concentration and then decaying gradually. However, the concentration profiles vary for the 12 subjects because of the random effects across individuals as well as the difference in dosage which is given in Table 1 below. The fit of the IPD model is shown with the data,

Table 1: Values of doses for the 12 patients

Patient	*Dose*, $D(mg/kg)$
1	4.02
2	4.40
3	4.53
4	4.40
5	5.86
6	4.00
7	4.95
8	4.53
9	3.10
10	5.50
11	4.92
12	5.30

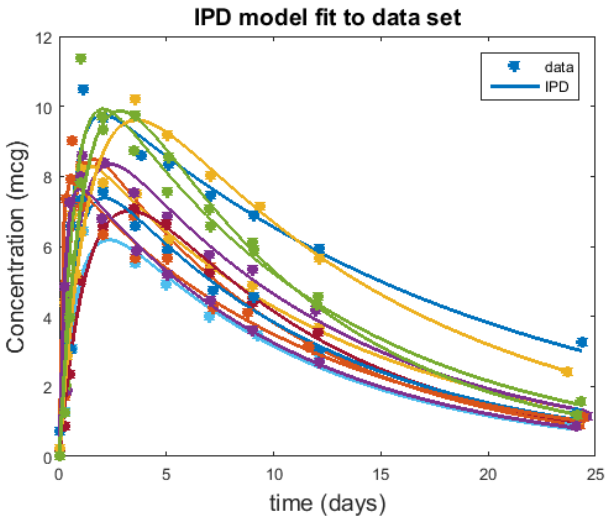

Figure 2: Concentration-time profiles for 12 subjects with the original dose.

but details on obtaining this fit will be given in Section 3.

3 AGGREGATE MODELS

In an aggregate model, a population model is formed from individual-specific models. In the following sections we give the detailed description of the techniques we used on the pharmacokinetic model described in Section 2 as fitted to the data shown in Figure 2.

3.1 Individual Parameter Distribution Technique (IPD)

The individual parameter distribution technique or IPD for short was first developed by Quijano et. al[4] to study the predation movement of a certain species of spiders. The approach in this method is based on the assumption that individuals within a population exhibit different behaviors and hence are modeled according to different parameter values that specifies these differences. For instance, the one-compartment model we described in Equation (2) would contain individual-specific parameter values for the rate of absorption k_a, the clearance Cl and the volume V. The model resulting from fitting Equation (2) to individual data shown in Figure 2 would be considered the individual based model (IBM).

In the IPD technique, the model parameters k_a, Cl and V were first estimated for each of the twelve individuals using the *fminsearch* function in MATLAB. This method uses the Nelder-Mead algorithm to find the parameter values q which minimize a given cost function. For our model, *fminsearch* seeks to return a vector of estimated parameters $q = [ka_{est}, Cl_{est}, V_{est}]^T$ which minimizes the cost function $J(q)$,

where

$$J(q) = \sum_{i=1}^{N} \mid \hat{C}_i - C(t_i; q) \mid^2 . \tag{3}$$

Here, $\hat{C}_i$ is the data at time t_i, $C(t_i; q)$ is the model at time t_i with parameter values $q = [k_a, \quad Cl, \quad V]$. We specified an initial estimate $[k_{a0} \quad Cl_0 \quad V_0]^T = [1.3034 \quad 0.0405 \quad 0.9236]^T$ of initial guesses for the parameter for each of the 12 subjects. Table 1 shows the estimated parameter values for the 12 subjects. It is obvious that there is variation in these parameters from individual to individual. Figure 3 shows the individual fit. There is a good fit for each individual data set. The task now is to develop an aggregate model in which the three parameters are drawn from an appropriate probability distribution which properly captures the variation displayed among the subjects.

Table 2: Estimates of the parameters of the model using IPD approach

$Subject$	k_{aest}	Cl_{est}	V_{est}
1	1.7774	0.0199	0.3693
2	1.9427	0.0448	0.4403
3	2.4536	0.0396	0.4858
4	1.1714	0.0374	0.4276
5	1.4714	0.0436	0.4931
6	1.1637	0.0511	0.5138
7	0.6797	0.0516	0.5046
8	1.3755	0.0465	0.5053
9	8.8656	0.0327	0.3773
10	0.6955	0.0324	0.4386
11	3.8490	0.0572	0.5834
12	0.8329	0.0420	0.3978

We now seek to specify the underlying probability distribution for each parameter value k_a, Cl and V across the population. Figure 4 shows three histograms for the

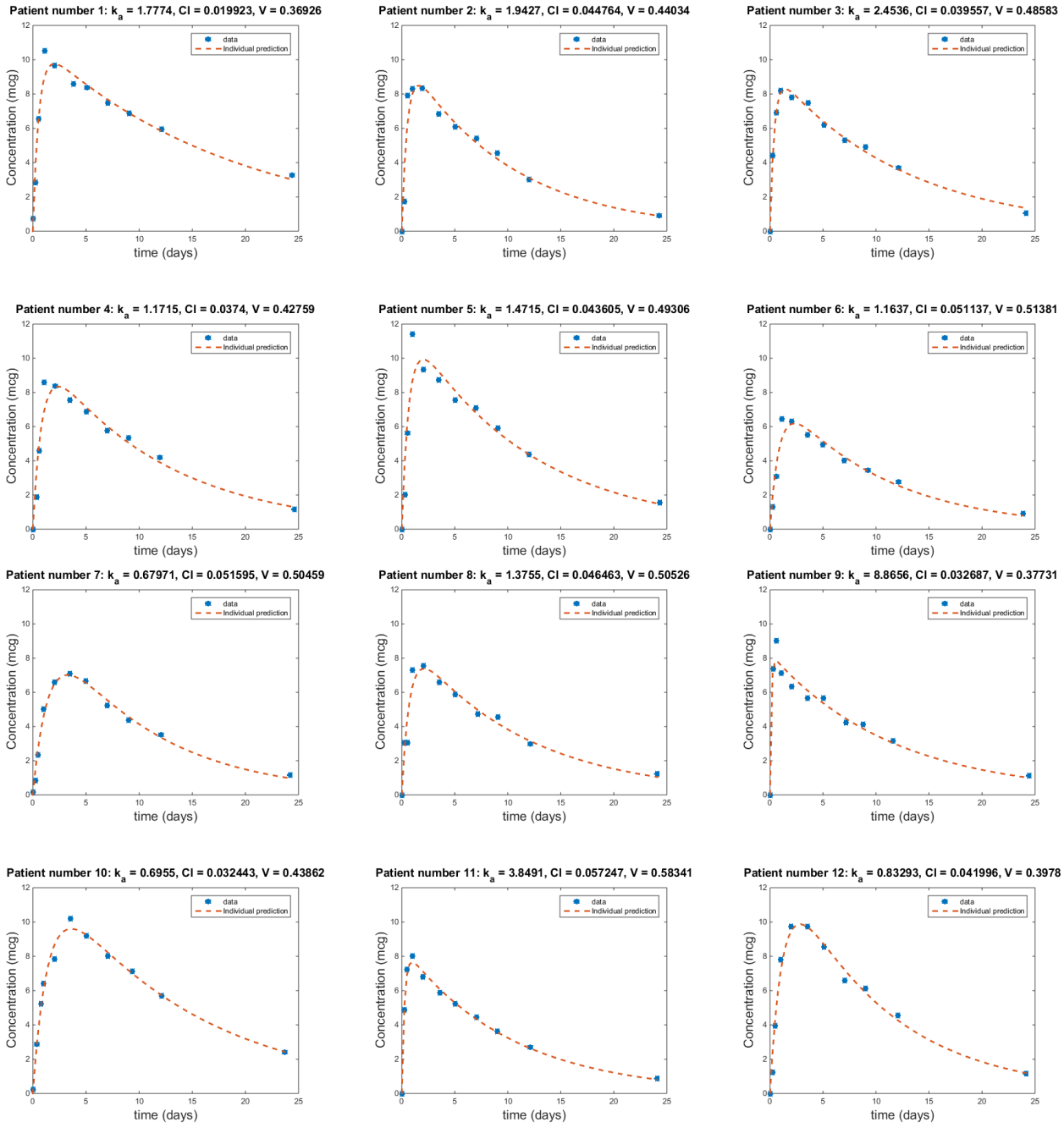

Figure 3: Individual fits to the model by minimizing Equation (3)

individual optimal parameter values for k_a, Cl and V. The three histograms indicate that none of the parameter values k_a, Cl and V follow a normal distribution. The

histogram of k_a appears to be skewed to the right while those of Cl and V do not show any skewness. It can also be noticed that the histograms of Cl and V look similar. We also created the matrix scatter plot for the three parameter values in

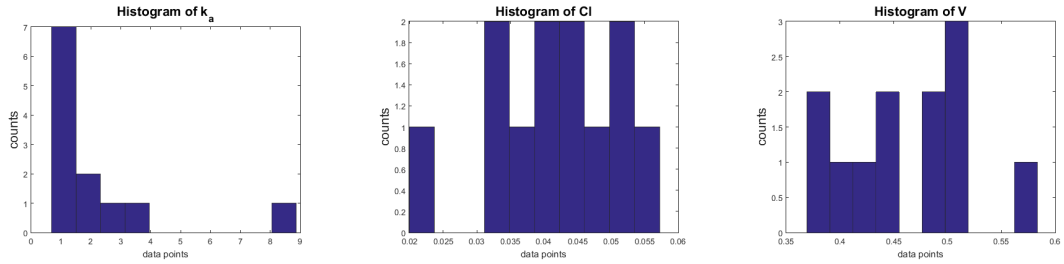

Figure 4: Histograms of the parameters of the model in Section 2

Figure 5. The scatter plot shows there is a linear relationship between Cl and V; therefore, we need to only specify the distribution for one of these (we chose V). Also the parameter k_a seemed not to have any association with either the parameter Cl or V.

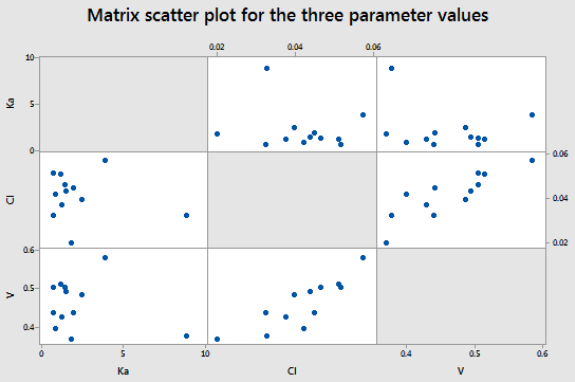

Figure 5: Scatter plot for the parameter values

The next task is to determine the appropriate probability distributions which describe this variation in k_a, Cl and V. Here, we tested the type of distribution that gives the best fit for parameters k_a and V. Each set of optimal parameter values for

k_a and V were fitted with different probability plots in MINITAB.

The outcome of these fits shows that k_a was best fitted with the 3-parameter Weibull distribution giving by

$$f(t) = \frac{\alpha}{\eta}\left(\frac{t-\gamma}{\eta}\right)^{\alpha-1} \exp^{-\left(\frac{t-\gamma}{\eta}\right)^{\alpha}} \tag{4}$$

α, $\gamma > 0$, $\eta > 0$ and $-\infty < \gamma < \infty$, where α is the shape parameter, γ is the location parameter and η is the scale parameter[?]. The 3-parameter Weibull distribution is the only distribution that gave the best capture for the optimal parameter values of k_a; that is, we had more of the optimal values for k_a falling on line in the distribution, see Figure 6, and this distribution also gave the largest p value. The large p value is an indication of how well the distribution fits the optimal values, compared to other distributions that we considered.

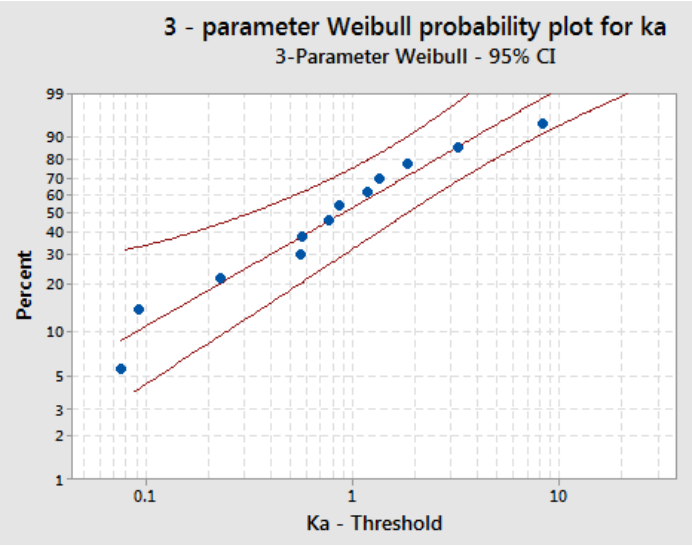

Figure 6: Probability plot for k_a

We also specified the distribution of V using the approach for k_a, but we did not

get a good fit to the optimal value of V with any probability distribution. We then generated 1000 random samples for parameter V using MINITAB and these 1000 random samples fitted a beta distribution with both shape parameters equal to 1.5. However, we had to rescale the original distribution Beta(1.5,1.5), because the original distribution of V values did not go from 0 to 1 but instead a shortened interval. The range of the optimal value for parameter V was (0.37 0.58), so we rescaled using using the linear function,

$$y = 0.21x + 0.37. \tag{5}$$

Since we will use the distribution of V to find a distribution for Cl, we had to further rescale the interval to capture the range of Cl values. The final linear rescale is given by

$$y = 0.32x + 0.28. \tag{6}$$

In Figure 7, we see that the form of association between the parameter values Cl and V is arguably positive linear due to the trend of the data in the plot. A linear model of the form

$$Cl = \beta_0 + \beta_1 V \tag{7}$$

was fitted, where we chose V as the explanatory variable and Cl the response variable. The values of the constants where estimated to be $\hat{\beta}_0 = -0.02$ and $\hat{\beta}_1 = 0.14$. The estimated value of $\hat{\beta}_0 = 0.00$ and it is close zero due to the assumption we have in Equation (2) ($K_e = Cl/V$). The plot in Figure 8 shows the fit together with the individual V and Cl values.

After the distribution was fitted to the data, we plotted a histogram of random variables for each of the parameter values k_a, Cl and V from the distributions we

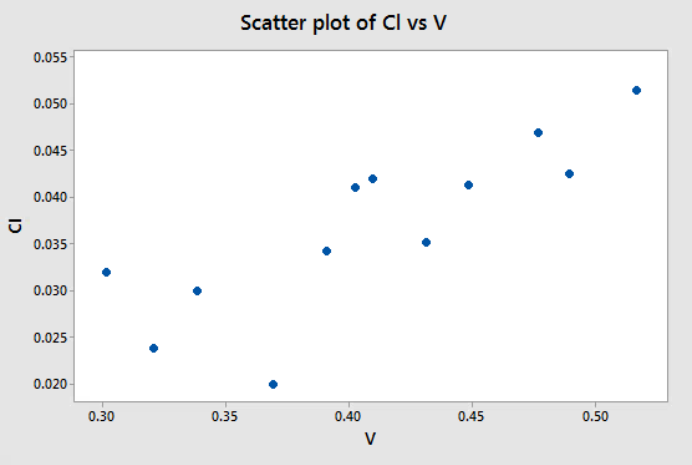

Figure 7: Scatter plot of Cl vs V

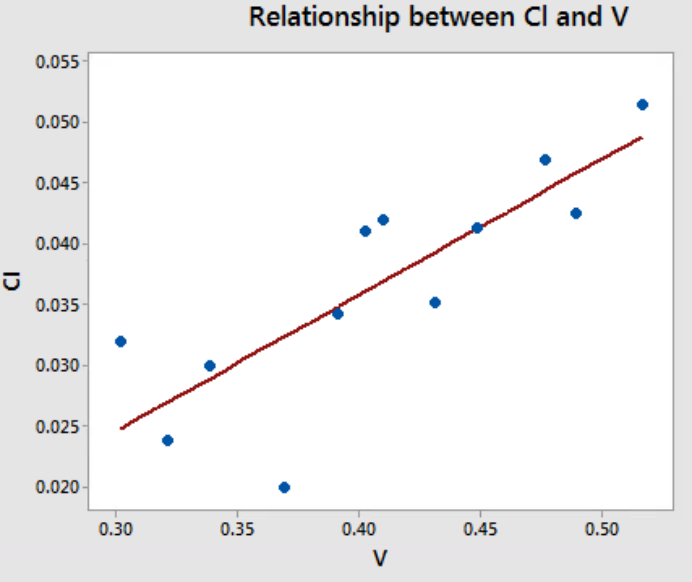

Figure 8: Plot of linear regression between Cl and V

determined above along with the parameter values and the distribution appears to describe the variation in the optimal parameter values.

The histogram for the distribution of parameter values k_a is given in Figure 9. Although all the parameter values are within the distribution, there is a long tail to

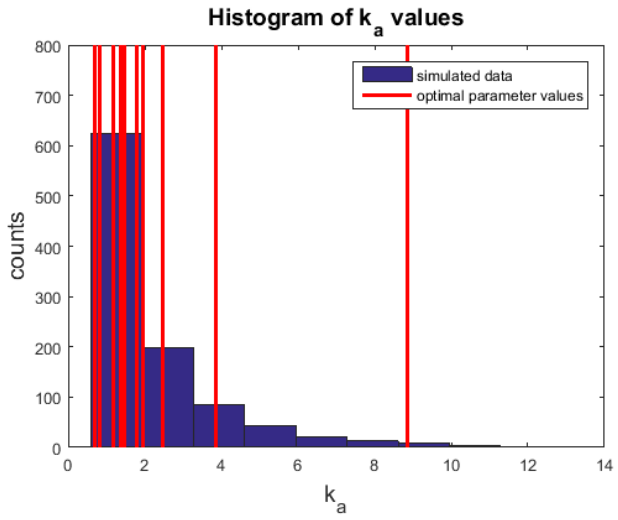

Figure 9: histogram of the distribution of k_a with shift random command

the distribution which is not seen in the original values. Notice that the histograms in Figures 9 and 10 where obtained after the rescaling of the Beta (1.5, 1.5) distribution. The red lines in each of the Figures 9-11 represent the actual values of the parameters for the individual-based models. It is obvious that each of the twelve individual parameter values fall within the distribution. This only suggests that the parameter values fall in the correct range. The most important feature of applying the IPD technique to create the population model is to be able to select a given parameter from a determined underlying probability distribution for that parameter and then using these values in the model, the model should capture the variation inherent in the population. The results of the population based model using these distributions for parameter values is given in Section 4.

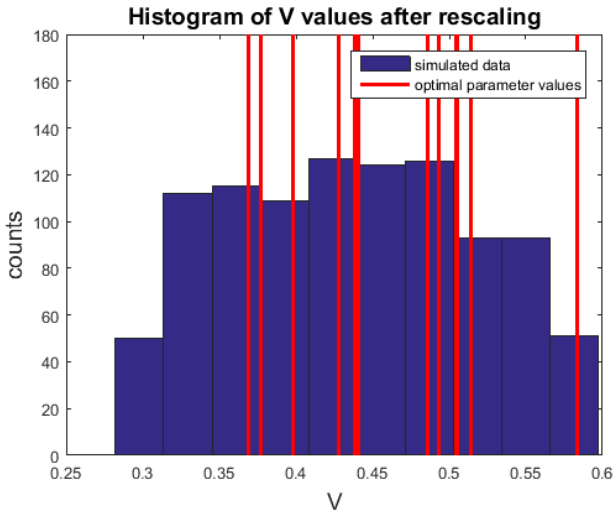

Figure 10: histogram of the distribution of V after rescaling

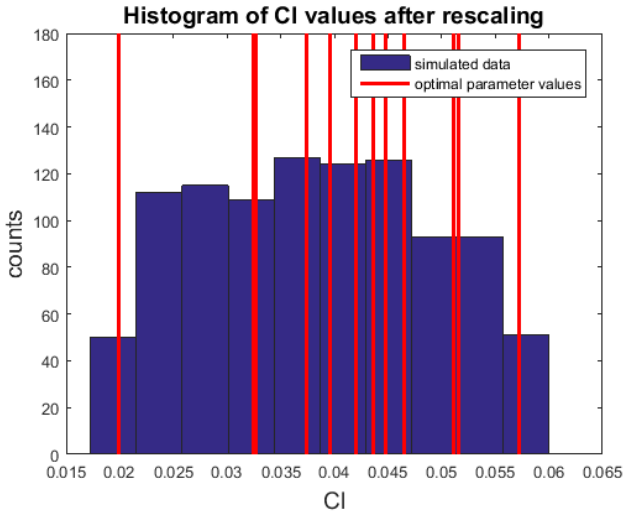

Figure 11: histogram of the distribution for Cl after rescaling

3.2 Nonlinear Mixed-Effect Method (NLME)

In NLME method, the population and individual-specific characteristics are taken into consideration, and these are referred to as fixed-effect parameters and random-

effect parameters respectively. An effect is anything that influences the value of a response variable at a specific setting of an explanatory or predictor variables[7]. Fixed effects represent population parameters, assumed to be the same each time data is collected. Random effects, on the other hand, are sample-dependent random variables[7]. In NLME modeling, random effects are seen as additional error terms that are assumed to be independently distributed with zero mean and constant variance across all measurements [7].

Let z_{ij} denote the jth observed response for individual i measured at time $t_{ij} = 1, 2, ..N$, $j = 1, 2, ..., N$. In the case of pharmacokinetics t_{ij} represents the j^{th} time the concentration is measured in the i^{th} individual.

In NLME described by Davidan et al[1], the governing equation is given as

$$z_{ij} = f(t_{ij}, u_i, \beta_i). \tag{8}$$

For our purpose, the function f in this equation stands for the concentration model described in Section 2, with $u_i = D_i$, the doses given in Table 1. The term β_i represents the parameters of f which is specific to individual i; in our case,

$$\beta_i = (k_{ai}, Cl_i, V_i)^T.$$

For each of these parameters, we assume there is a fixed effect as well as random effects. However, unlike [1], we do not use weight and creatinine clearance as given in the data. We simply assume that

$$k_{ai} = \exp(\beta_1 + b_{1i}),$$

$$V_i = \exp(\beta_2 + b_{2i}),$$

and

$$Cl_i = \exp(\beta_3 + b_{3i}) \tag{9}$$

where

$$\beta_i = [\beta_{1i}, \beta_{2i}, \beta_{3i}]$$

are the fixed effects and $b_i = [b_{1i}, b_{2i}, b_{3i}]$ are the random effects specific to individual i. We note that we are assuming that $\log(k_a)$, $\log(V)$ and $\log(Cl)$ are normal and b_i is normal with mean 0 and covariance Ψ. In what follows next, we will describe how we used the *nlmefit* package in MATLAB to estimate the fixed and random effects described for the concentration model and data described in Section 2.

The *nlmefit* attempts to estimate the fixed effects β and the covariance matrix for the normally distributed random effects Ψ by maximizing the marginal likelihood. We implemented the *nlmefit* on our model by estimating the logarithm of the parameter values. Our initial is given by

$$\beta_0 = [\log(k_{a0}) \quad \log(Cl_0) \quad \log(V_0)]^T = [0.2650 \quad -3.2070 \quad -0.795]^T$$

for the parameter estimates and dose D_i for individual subjects. The results obtained for the fixed effect and random effects for the logarithm of the parameter values are given by $\beta_1 = [-2.4547 \quad -3.2272 \quad -3.6929]^T$ and the random effect covariance vector

$$\Psi = \begin{bmatrix} 0.000 & 0 & 0 \\ 0 & 0.0279 & 0.0281 \\ 0 & 0.0281 & 0.4426 \end{bmatrix}.$$

The estimated covariance matrix Ψ of the random effects shows that the variance of the first parameter, $\log(k_a)$, is essentially zero, suggesting that we can set it as constant to simplify the model. It also indicated that the estimated random effects of

the parameters $\log(k_a)$ is not correlated with either estimated random effect of $\log(V)$ or $\log(Cl)$. Whereas, estimated random effects of the parameters $\log(V)$ and $\log(Cl)$ are correlated.

The model was refitted using the random effects for $\log(Cl)$ and $\log(V)$ only. In addition, the statistics we obtained from the *nlmefit*: the log-likelihood, ($\log(l) = -177.024$) without estimating random effects for $\log(k_a)$ random effect is identical to what we had ($\log(l) = -177.022$) when estimating all the random effects. The Akaike information criterion (AIC), which measures the quality of the model, is reduced from 370.05 to 368.05 and the Bayesian information criterion (BIC), which also determines the model to be preferred, is reduced from 373.92 to 371.42. The resulting small values obtained for both the (AIC) and (BIC) shows that the model with the smaller value from what we initially had should be preferred to refit the model. This led to a new covariance matrix for random effects while the fixed effect values remained the same. The new covariance matrix is given

$$\Psi_1 = \begin{bmatrix} 0.0280 & 0.0285 \\ 0.0285 & 0.4397 \end{bmatrix}.$$

The result for the combined estimates of the mixed effects from this method for each of the twelve subjects is given in Table 3, where this includes the estimated fixed effects plus the exact b_i values for each individual data set and parameter.

The model obtained using the individual estimated parameter values from the NLME approach is plotted in Figure 14 for all the twelve subjects together with the fits we saw in Figure 3 using the IPD technique. As with the previous aggregate model approach, the individual based models obtained using the NLME approach is a good approximation to the individual data sets. The solid colored lines represents

Table 3: Estimates of the parameters of the model using NLME approach

$Subject$	k_{aest}	Cl_{est}	V_{est}
1	0.0859	0.0397	0.0175
2	0.0859	0.0397	0.0252
3	0.0859	0.0397	0.0256
4	0.0859	0.0397	0.0234
5	0.0859	0.0397	0.0268
6	0.0859	0.0397	0.0289
7	0.0859	0.0397	0.0292
8	0.0859	0.0397	0.0277
9	0.0859	0.0397	0.0204
10	0.0859	0.0397	0.0224
11	0.0859	0.0397	0.0319
12	0.0859	0.0397	0.0236

IPD technique and the dashed lines are the NLME model fit. The points in the plots are the actual data sets for the model. As a side note, Figures 12 and 13 show the distribution of the optimal parameter values of V and Cl respectively using the normal distribution with mean and variance as described above. In these figures we see that we have longer tails which we believe causes the the main difference between the IPD technique and the NLME method.

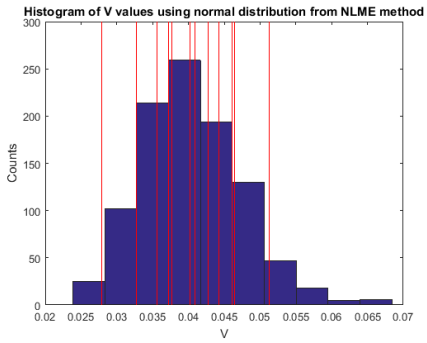

Figure 12: histogram of the distribution of V using NLME method

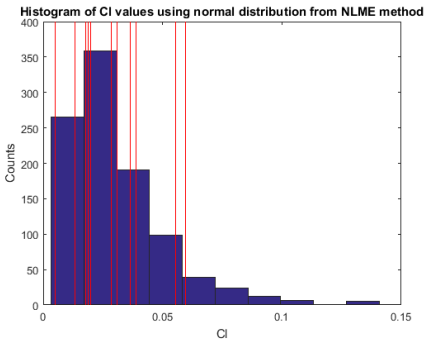

Figure 13: histogram of the distribution for Cl using NLME method

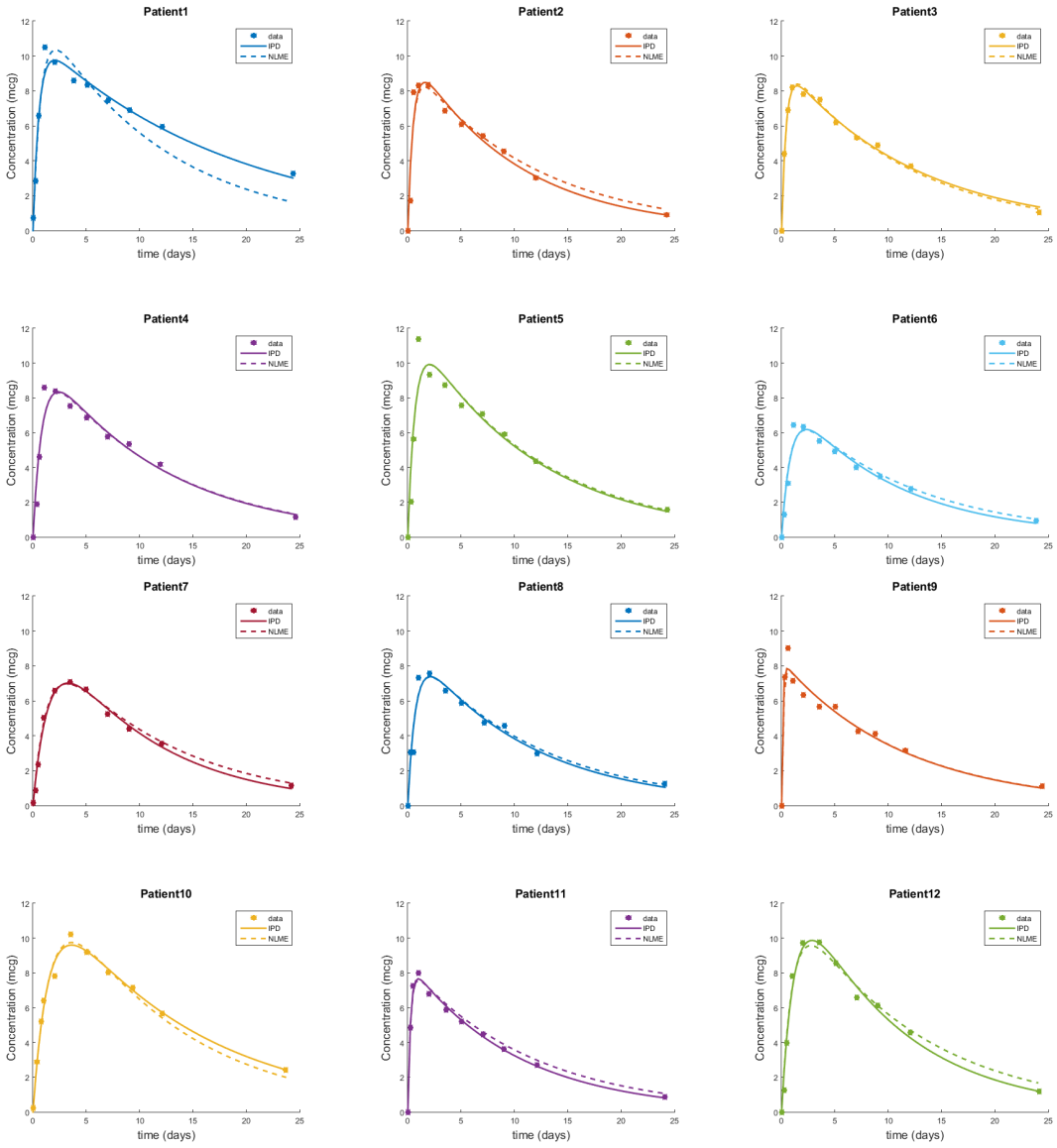

Figure 14: Individual fits to the model using the parameter for NLME method

4 COMPARISON OF THE POPULATION MODEL RESULTS USING THE TWO METHODS

In the two approaches we have considered in this thesis, we obtained different individual estimates for the parameter values of the one-compartment concentration model although both provided good fits to the individual data. In the first approach, IPD, we estimated the parameters of the model by first using ordinary least squares to estimate the parameters for the IBM. These estimates were then used to obtained the correlation between parameters Cl and V together with the underlying probability distributions for all the individual parameters of the model. In contrast, the NLME method assumes parameters are normally distributed and then estimated the fixed and random effects for the parameters in the model.

We further examined the comparison between the two methods by generating 1000 simulations of the model using the two approaches; this simulates the concentration profiles for 1000 random individuals in a population using the two different approaches for determining the parameters within the model. To generate the parameters for the model using the IPD technique, we randomly selected 1000 different draws from a Weibull distribution for parameter k_a and 1000 random samples from a beta distribution with both shape parameters 1.5 for the value of V as explained in Section 3. Then we calculated Cl using Equation (6). Plugging these values into Equation (2), we obtain simulated concentration profiles for a population of 1000 individuals as shown in the left hand plot in Figure 15. For this method, the twelve individual profiles seemed to fit well in the population model; however, there were several concentration profiles with higher peak concentrations than those represented

by the individual models. This is most likely due to the tail in the distribution of parameter values for k_a.

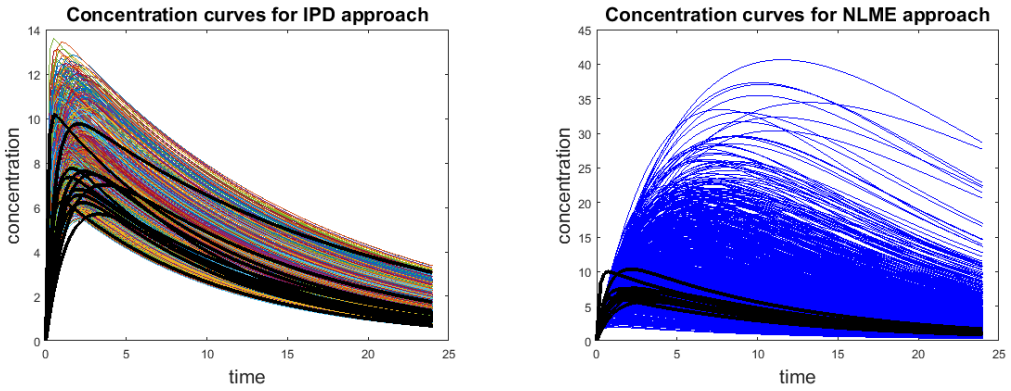

Figure 15: Concentration profiles for IPD (left) and NLME (right)

To generate the parameters using NLME method, the parameter k_a was assumed constant across the population. Meanwhile, the other two parameters were assumed to be normally distributed with mean and variance as defined in Section 3. As in the case of the IPD technique, once the parameters were obtained, they were substituted into Equation (2). Those concentration profiles are given in the right hand plot of Figure 15. There were also outliers in this case, but many of these outliers had both higher concentrations as well as a shift in the time of the peak concentration. This variation could also be due to the tail in the distribution. For comparison sake, we set the value of the doses for the 12 different patients as $D = 4.02mg/kg$ in the population model. We compared several derived quantities from these concentration curves, namely the area under the curve (AUC), the peak concentration and final concentrations for each of the 1000 simulated individuals.

In Figure 16, we examine the histograms for the AUC using the two different approaches specified. The red lines are computed from the individual based model

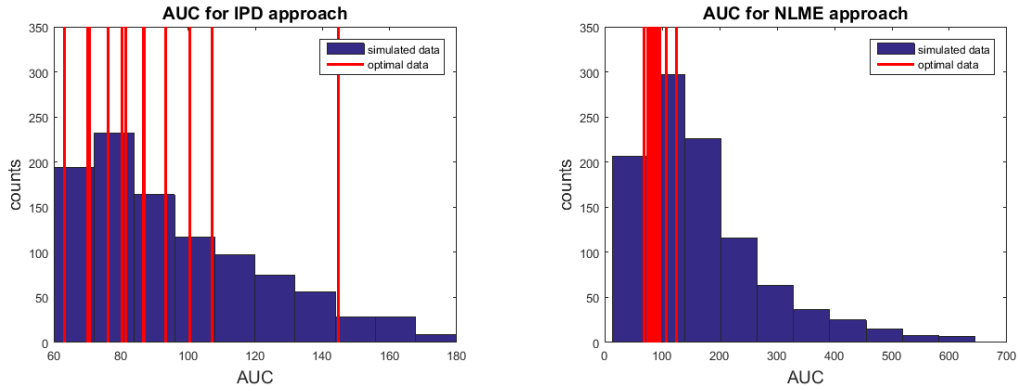

Figure 16: Histograms of AUC for IPD (left) and NLME (right)

for each method with a fixed dose of theophylline sample across all individuals. This allows us to more accurately compare the effects of the method as opposed to variations due to the effect of the dosing. The histogram gives the count of the 1000 simulations within a bin. We have that all the individual values of AUC are captured in the population model for both methods. However, the biggest difference is the tail of the population distribution for AUC when using the NLME method. This is not surprising given the concentration curves in Figure 15.

The distribution of the peak concentrations was obtained in a similar way using 1000 simulations of the population model generated from each of the two approaches. The peak concentration is the maximum concentration of the drug measured. In the first method, the IPD technique, we had the distribution of the calculated values from the individual models (indicated by the red lines in the left plot of Figure 17) lying inside a good portion of the distribution for the population based model given by the histogram. The NLME method still captures the calculated peak concentrations from the individual models, but there is a longer tail to the distribution for the peak concentration as compared to the more refined values for the calculations based on

the individual models. Again this is not surprising given the concentrations curves in Figure 15 when using the NLME method.

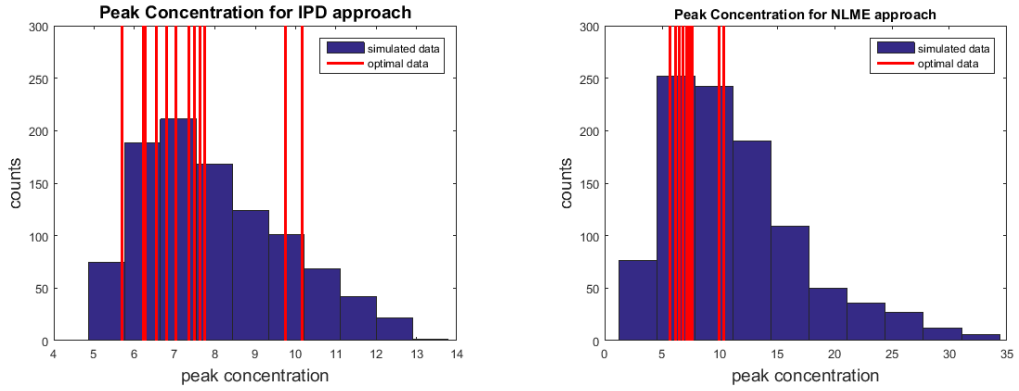

Figure 17: Histograms of Peak concentrations for IPD (left) and NLME (right)

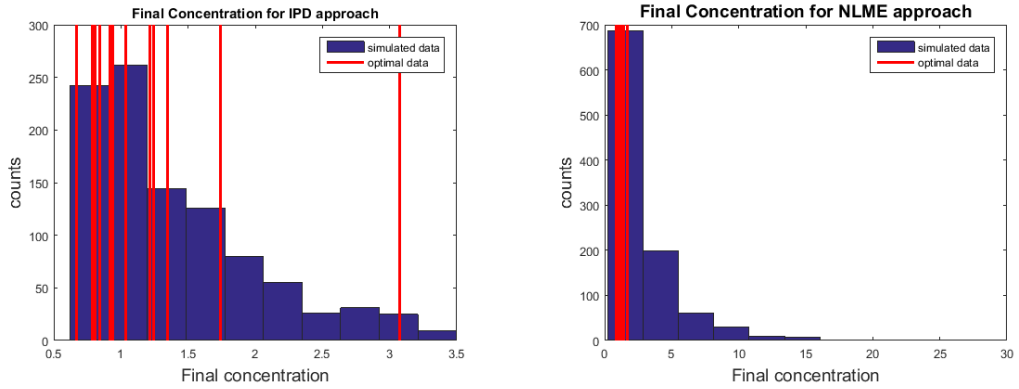

Figure 18: Histograms of Final concentrations for IPD (left) and NLME (right)

In Figure 18, we have the distributions for the final concentrations for the 1000 simulations we performed using the individual values. It is noticed that we also have our individual values being captured by a good portion of the population distribution. In the first graph to the left, the IPD technique has a tail in the distribution; we see a similar trend in the NLME method as well, although the tail is longer when using the NLME method as denoted by the values on the x axis.

5 CONCLUSIONS AND FUTURE WORK

In this thesis, we have explored two methods for creating a population-based model using individual data implemented on a one-compartment concentration model. The two methods have been used to obtain estimates for the distributions of parameters in the model. In the first method, we first estimated the parameters for the individual based models using the ordinary least squares approach and the twelve individual data sets. We then estimated whether correlations existed between the parameters; in this example, the parameters Cl and V were correlated. We then fitted distributions to k_a and V and used the linear relationship between Cl and V to estimate the distribution of Cl. In the process, we needed to rescale the probability distribution of the parameter, V so the resulting distribution for Cl captured all the individual parameter estimates of Cl. In future work, a better means of finding the probability distribution for such correlated parameters with little or no rescaling needs to be explored. We performed 1000 simulations of the model using the distributions found; our results from the population dynamics showed that the trend of the actual data was captured well by the population model. This allows us to conclude that the first approach provides a reasonable method for formulating a population model based on individual data sets.

The second method we implemented, the NLME method, assumes the parameters are normally distributed and estimates both fixed and random effects for the parameters. The NLME method in this thesis was implemented using the *nlmefit* package in MATLAB. The results of implementing the NLME method indicated the variance of one of the parameters k_a was zero; therefore, this parameter was assumed to be

constant across the population. The mean and variances for the other two parameters, Cl and V, were also estimated. Just like the first method, we also constructed a population model from 1000 simulations of the concentration model using the normal probability distribution for the two parameters and the fixed value for the third parameter. Although this second approach also seemed to capture the dynamics of the three variables in which we were interested, there was also a tail in each population distribution which is not warranted based on the original data set.

In conclusion, the two aggregate methods we considered in this thesis were a good fit for the individual based-models while also capturing the dynamics with the population based models. However, more work needs to be done to consider whether there is a better way for determining the distribution for the parameter values in the first approach as well as whether there is a set of models for which the assumption of normal distribution for parameter values is accurate.

BIBLIOGRAPHY

[1] M. Davidian and D. M. Giltinan, *Nonlinear models for repeated measurement data*, vol. 62, CRC Press, 1995.

[2] D. E. Irwin, I. Milsom, S. Hunskaar, K. Reilly, Z. Kopp, S. Herschorn, K. Coyne, C. Kelleher, C. Hampel, W. Artibani, et al., *Population-based survey of urinary incontinence, overactive bladder, and other lower urinary tract symptoms in five countries: results of the epic study*, European urology, 50 (2006), pp. 1306–1315.

[3] J. C. Pinheiro and D. M. Bates, *Approximations to the log-likelihood function in the nonlinear mixed-effects model*, Journal of computational and Graphical Statistics, 4 (1995), pp. 12–35.

[4] A. J. Quijano, M. L. Joyner, E. Seier, N. Hancock, M. Largent, and T. C. Jones, *An aggregate stochastic model incorporating individual dynamics for predation movements of anelosimus studiosus.*, Mathematical biosciences and engineering: MBE, 12 (2015), pp. 585–607.

[5] C. Reynolds, *Individual-based models*, Retrieved November, 25 (1999), p. 2003.

[6] R. Salway and J. Wakefield, *Gamma generalized linear models for pharmacokinetic data*, Biometrics, 64 (2008), pp. 620–626.

[7] L. Wang, *Estimating Nonlinear Mixed-effects Models by the Generalized Profiling Method and Its Application to Pharmacokinetics*, PhD thesis, McGill University, 2007.

VITA

OYEBOLA OYERO

Education:	M.S. Mathematics, East Tennessee State University, Johnson City, Tennessee 2016 B.S. Mathematics, Obafemi Awolowo University, Ile-Ife, Nigeria 2008
Professional Experience:	Graduate Assistant, East Tennessee State University, Johnson City, Tennessee 2014-2016 Teaching Assistant, Obafemi Awolowo University, Osun State, Nigeria 2006-2008
Research Interest:	Applied Mathematics, Stochastic differential Equations, Big Data Analytics, Operation Research, Convex Optimization, Fluid Mechanics