

Statistical Inference of Pharmacokinetic Models of Theophylline and Warfarin Data.

Annika Israelsson

April 27, 2017

Abstract

Pharmacokinetics is the study of how drugs are absorbed and distributed in the body and is used by pharmacists to ensure that they prescribe the appropriate dosage of medicine. The process used to determine the drug concentration in the body when medicine is prescribed is ADME; absorption, distribution, metabolism and excretion. The body is viewed as consisting of compartments between which a drug moves. This project focuses on a two compartment model where the body proceeds at a rate that is dependent on the concentration of the drug. Compartmental models can be described in terms of a set of linear, first-order, constant-coefficient, ordinary differential equations. A change of concentration in one of the compartments is a linear function of the concentrations in all other compartments.

There are two sets of data that were evaluated. The first is Theophylline which is a drug used for respiratory disease. The data given consisted of twelve patients who took one single oral dose of Theophylline and their drug concentration in blood was sampled multiple times over a period of 24 hours. The next data was from 31 patients who were given one single oral dose of Warfarin. Warfarin is an anticoagulant used as a blood thinner. The concentration of Warfarin was examined over a longer period of time, roughly five days.

The aim of the project was to fit and statistically evaluate the dynamics of the concentration of the drugs administered to the subjects. When the parameters were estimated and the estimates were plotted an interesting occurrence was discovered. The lack of rising concentration values for most of the Warfarin Patients and for Patient 9 from the Theophylline data cause a poor estimate for the rate of absorption. It proved that an over parameterization had occurred and that a one compartment model might have given a better outcome. Patients who had rising concentration values also had more accurate estimates, this can also be seen in the plots. 95% confidence intervals and standard errors were also calculated.

Lastly, correlations between the patients and the parameters were plotted and visually evaluated. The Theophylline patients had no noticeable correlations which could be a result of the small number of patients. The Warfarin correlation plots appeared to have slight correlations for the weight of the subjects and the rate of elimination as well as the volume of distribution. There may also have been a correlation of gender and parameters for the Warfarin data, but because there was only 5 female subjects of 31 it is unclear if there is a correlation or not.

Populärvetenskaplig sammanfattning

Studien av hur droger, exempelvis läkemedel, absorberas och distribueras i kroppen kallas medicinvetenskap. Apotekare använder medicinvetenskap för att garantera att de ordinerar lämplig dosering av medicin. Processen som används för att bestämma läkemedelskoncentrationen i kroppen när läkemedlet ordineras kallas för ADME; Absorption, fördelning, metabolism och utsöndring. Kroppen betraktas som bestående av delar vilka läkemedel rör sig mellan. För detta projekt såg vi kroppen som en tvåavdelningsmodell och hur kroppen bryter ner drogen är beroende på koncentrationen av drogen.

Två olika data grupper användes för projektet . Den första är Theophyllin som är ett läkemedel som används för andningssjukdomar. De givna uppgifterna bestod av tolv patienter som tog en oral dos av Theophyllin och deras läkemedelskoncentration i blod samlades flera gånger under en period av 24 timmar. Nästa data var från 31 patienter som fick en oral dos av Warfarin. Warfarin är ett antikoaguleringsmedel. Koncentrationen av Warfarin undersöktes under en längre tid, ungefär fem dagar.

Syftet med projektet var att statistiskt utvärdera dynamiken i koncentrationen av drogerna som administrerats till patienterna. När parametrarna skattades och uppskattningarna plottades upptäcktes en intressant förekomst. Stigande koncentrationsvärdena saknades för de flesta warfarinpatienterna och patient 9 från Theophyllindata orsakar en dålig uppskattning av absorptionshastigheten. Det visade sig att data värdena som vi hade var för enkla för en tvåavdelningsmodell och kroppen borde ha betraktats som en avdelningsmodell för uppskattningarna. Patienter som hade stigande koncentrationsvärdet hade också mer exakta uppskattningar. En 95% konfidensintervall och standarderorna beräknades också.

Slutligen plottades korrelationer mellan patienterna och parametrarna och utvärderades visuellt. Theophyllinpatienterna hade inga märkbara korrelationer vilket kan vara ett resultat av det få antal patienter. Warfarin-korrelationsplottorna verkade ha små korrelationer för individens vikt och eliminationshastigheten samt fördelningsvolymen. Det kan också ha varit en korrelation mellan kön och parametrar för Warfarins data, men eftersom det bara fanns fem kvinnor av 31 är det oklart om det finns en korrelation eller inte.

Contents

1	Introduction	5
1.1	Pharmacokinetics	5
2	Theory	6
2.1	Compartmental analysis	6
2.2	Nonlinear Regression Models	8
3	Approach	8
3.1	Optimization	10
3.2	Code and Output	11
3.3	Confidence Intervals	13
4	Results	13
4.1	Theophylline Results	13
4.2	Warfarin Results	15
5	Conclusions	18
5.1	Acknowledgments	20
6	Appendix	21
6.1	Theophylline Results	21
6.2	Theophylline Patients	22
6.3	Theophylline Correlation Plots	24
6.4	Warfarin Results	26
6.5	Warfarin Patients	29
6.6	Warfarin Correlation Plots	34

1 Introduction

Pharmacokinetics is a branch of pharmacology which studies how substances, such as prescription drugs, are absorbed and distributed in the body. The body is viewed as consisting of compartments between which a drug moves and from which elimination occurs. The transfer of the drug between these compartments is represented by rate constants [5]. The way the body reacts to the drug proceeds at a rate that is dependent on the concentration of the drug. There are different ways for a drug to be administered. Some examples include intravenous administration, both single and multiple, and oral dose, both single and multiple. For the analysis here, two different data sets were used, Theophylline and Warfarin. This project focuses on single oral administration, in a two compartment system with rates constant, where the drug is a first order statistic.

The main goal is to consider statistical inference for several models representing the dynamics of the concentration of the drug administered to several subjects. Models will be fitted to available data. This means that model parameters will be estimated statistically for each subject and it will be of interest to check for possible correlations between the estimates and covariates such as gender, age and weight of the subjects. This will be done for both the warfarin and theophylline data sets. Taking into account the fact that measurements have some random measurement "noise", and by assuming this, an additional parameter in the form of variance noise will be estimated.

1.1 Pharmacokinetics

In pharmacology there is pharmacodynamics which is the study of how a drug affects the organism, and pharmacokinetics which is the study of how an organism affects a drug. It can measure anything from pesticides to food additives. The object of pharmacokinetics is to analyze the chemical metabolism and to discover the fate of a chemical from the moment it is administered up to a point at which it completely leaves the body. It is an important field used for pharmacists to ensure that they prescribe the appropriate dosage to achieve optimal treatment and minimal toxicity [4].

The process used to determine the drug concentration in the body when medicine is prescribed is known as ADME; absorption, distribution, metabolism, and excretion. Absorption is the process of a substance entering the blood circulation, distribution means the dispersion of a substance throughout the fluid and tissues of the body and metabolism refers to the compounds breaking down when they enter the body. As metabolism occurs, the initial compound is converted into new compounds called metabolites. Thus, metabolism is the recognition by the body that a foreign substance is present and the irreversible transformation of parent compounds to daughter metabolites [6]. Lastly, excretion refers to the removal of the substance from the body.

A two compartment system is a way to interpret the ADME principle which will be discussed further in section 2.1. Here, when the drug is absorbed into the body it enters the first compartment. Then, it is distributed into the blood and plasma which is the second compartment. During this process the drug is metabolized. The final step is for the drug to leave the body through excretion. The body's reaction to a specific drug proceeds at a rate that is dependent on the concentration of the drug. The pharmacokinetic properties of substances can be affected by both the route of administration and the dose of the drug which can affect the absorption rate.

2 Theory

2.1 Compartmental analysis

A compartment is used as a model to more easily perceive the processes that take place in the interaction between the body and the chemical substance. Inside a compartment the drug is assumed to be uniformly distributed. If there would be a concentration difference inside a compartment, it would be divided into two separate compartments.

The act of the drug entering the body is shown in figure 1 where C_0 represents the central compartment and K_a is the constant rate of absorption having unit t^{-1} , where t represents time. Commonly the elimination occurs in a peripheral compartment, depicted as C_1 , by the liver and kidney breaking down the drug, but the elimination can also occur elsewhere. The constant rate of elimination is represented by K_e , also having unit t^{-1} . Upon entering the body, the drug concentration will decline rapidly in C_0 , then rise to a maximum in C_1 , and then decline as it leaves the system, see figure 1.

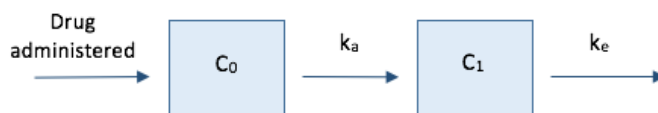


Figure 1: A two compartment system

Compartmental models can be described in terms of a set of linear, first-order, constant-coefficient, ordinary differential equations. A change of concentration in one of the compartments is a linear function of the concentrations in all other compartments. Pharmacokinetic compartmental analysis uses kinetic models to describe and predict the concentration-time curve. The advantage of compartmental analysis is the ability to predict the concentration at any time. The concentration of the drug in the body is used to determine if the dosage regimen is effective.

If the amount of administered drug, X_0 , which for this experiment is in mg, is decreasing at a rate that is proportional to X_0 , then the rate of elimination of the drug is described as:

$$\frac{dX_t}{dt} = -KX_0 \quad (1)$$

where X_t is the quantity of the drug at time t , K is the first order elimination rate constant and the reaction proceeds at a rate that is dependent on the amount of X_0 in the body[5]. This is known as a one-compartment model. Even when the amount of drug administered increases, the body is still able to eliminate the drug accordingly, preventing accumulation. Eventually, if the initial amount increases to a certain extent, all drugs will take on the form of a zero order rate constant. This means the body will only be able to eliminate a certain amount per time. This is the case, for example, with overdoses.

For the single oral dose data, consider the concentration versus time profile. Assuming first-order absorption and first-order elimination for a two-compartment model as in figure 1. The rate of change of the amount of drug in the tissue compartment, C_1 , in the body is given by the differential equation:

$$\frac{dX_t}{dt} = K_a X_o - K_e X_t$$

Following integration, the differential equation becomes:

$$X_t = \frac{X_0 K_a [\exp(-K_e t) - \exp(-K_a t)]}{K_a - K_e}$$

Here, t is the time of the blood sample in hours, and X_0 is the amount of drug administered to the patient. The concentration, C_t of drug in plasma is given by the equation:

$$C_t = \frac{X_0}{V_d}$$

and thus the differential equation becomes:

$$C_t = \frac{X_0 K_a [\exp(-K_e t) - \exp(-K_a t)]}{V_d (K_a - K_e)} \quad (2)$$

V_d is the volume of distribution in liters which gives information on the drugs distribution in the body. It refers to the volume of plasma in which the total amount of drug in the body would be required to be dissolved in order to reflect the drug concentration attained in plasma. This is why, in this experiment, the units of C_t are in mg/liters.

2.2 Nonlinear Regression Models

For concentration measured at n discrete time points $\{t_j\}_{j=1,\dots,n}$, a statistical model for the measurements $\{y_j\}_{j=1,\dots,n}$ can be written as:

$$y_j = f(t_j; \phi) + \epsilon_j, \quad 1 \leq j \leq n$$

Here, f is the structural model which is a parametric function of time and is dependent on ϕ . In our case, $f(t_j; \phi) = C_t$. The vector (t_1, t_2, \dots, t_n) is the vector of sampling times and $\phi = (\phi_1, \phi_2, \dots, \phi_d)$ is a vector of d structural parameters which, for this project, is $\phi = (K_a, K_e, V_d)$. $(\epsilon_1, \epsilon_2, \dots, \epsilon_n)$ is the residual error. With classical nonlinear regressions it is possible to make some assumptions about the error. The errors ϵ_j have mean zero, and are uncorrelated. They have common variance σ^2 and are identically distributed for all t_j . Lastly, the errors ϵ_j are normally distributed [3].

For $y = y_j$ $1 \leq j \leq n$ the aim is to estimate the models parameters, ψ . Where $\psi = (\phi, \sigma^2)$ is the vector of parameters to be estimated and σ^2 , estimated by $\hat{\sigma}^2$, is calculated by the following equation:

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^n (y_j - C_{t_j})^2}{n - d}$$

Then the confidence intervals of the parameters should be computed in order to evaluate the reliability of the estimates. With the observed data, y_t given, and the parameter model, C_t from (2), the goal is to find the best parameter for the model. This is done using least squares which, for the stated assumptions, on $\{\epsilon_j\}$ also returns maximum likelihood estimates.

3 Approach

The data sets used are for the drug Theophylline which is a drug used for respiratory disease, and Warfarin, an anticoagulant used as a blood thinner. The figures 2 and 3 show a part of the data sets and 4 shows the plot of the concentration values for Patient 1 from the Theophylline data. Notice, at time $t = 0$ the drug is administered orally and the time entries after that are given in hours. Then, after the drug has had some time to appear in the plasma, the first measurement is taken. The concentration measurements are written in mg/l and, for Warfarin, the dose is $1.5mg/kg$ of body weight.

ID	Amount	Time	Concentration	Weight	Sex
1	4.02	0.00	0.00	79.6	M
1	0	0.25	2.84	79.6	M
1	0	0.57	6.57	79.6	M
1	0	1.12	10.5	79.6	M
1	0	2.02	9.66	79.6	M
1	0	3.82	8.58	79.6	M
1	0	5.10	8.36	79.6	M
1	0	7.03	7.47	79.6	M
1	0	9.05	6.89	79.6	M
1	0	12.12	5.94	79.6	M
1	0	24.37	3.28	79.6	M
2	4.4	0.00	0.00	72.4	M
2	0	0.27	1.72	72.4	M
2	0	0.52	7.91	72.4	M
2	0	1.00	8.31	72.4	M
2	0	1.92	8.33	72.4	M
2	0	3.50	6.85	72.4	M
2	0	5.02	6.08	72.4	M
2	0	7.03	5.40	72.4	M
2	0	9.00	4.55	72.4	M
2	0	12.00	3.01	72.4	M
2	0	24.30	0.90	72.4	M
3	4.53	0.00	.	70.5	M
3	.	0.27	4.4	70.5	M
3	.	0.58	6.9	70.5	M
3	.	1.02	8.2	70.5	M

Figure 2: A part of the data set of Theophylline with the subject's weight in kg.

id	time	amt	dv	wt	sex	age
1	0.0	100	.	66.7	1	50
1	24.0	.	9.2	66.7	1	50
1	36.0	.	8.5	66.7	1	50
1	48.0	.	6.4	66.7	1	50
1	72.0	.	4.8	66.7	1	50
1	96.0	.	3.1	66.7	1	50
1	120.0	.	2.5	66.7	1	50
2	0.0	100	.	66.7	1	31
2	0.5	.	0	66.7	1	31
2	2.0	.	8.4	66.7	1	31
2	3.0	.	9.7	66.7	1	31
2	6.0	.	9.8	66.7	1	31
2	12.0	.	11	66.7	1	31
2	24.0	.	8.3	66.7	1	31
2	36.0	.	7.7	66.7	1	31
2	48.0	.	6.3	66.7	1	31
2	72.0	.	4.1	66.7	1	31
2	96.0	.	3	66.7	1	31
2	120.0	.	1.4	66.7	1	31
3	0.0	120	.	80.0	1	40
3	3.0	.	12	80.0	1	40
3	6.0	.	13.2	80.0	1	40
3	9.0	.	14.4	80.0	1	40
3	24.0	.	9.6	80.0	1	40
3	36.0	.	8.2	80.0	1	40
3	48.0	.	7.8	80.0	1	40

Figure 3: A part of the data set of Warfarin, where $sex = 1$ refers to male patients and dv is the concentration in mg/l and the subject's weight in kg.

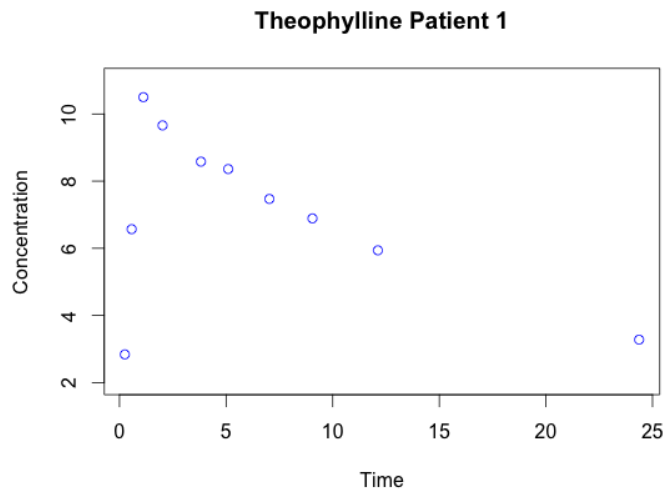


Figure 4: A plot of the concentrations for Patient 1 from the Theophylline data.

Using the concentration values, y_t given from the data, the goal was to estimate the parameters $K_a, K_e,$ and V_d that would minimize:

$$\sum_{j=1}^n (y_j - C_{t_j})^2$$

Using the concentrations and times given, the parameters, K_a, K_e, V_d can be estimated by using a built in function in the R software called Optim.

3.1 Optimization

An optimization problem consists of maximizing a real function by systematically choosing input values from within an allowed set and computing the value function. It is a selection of a best element for some set of available alternatives[2]. Optimization includes finding "best available" values of some objective function given a defined domain.

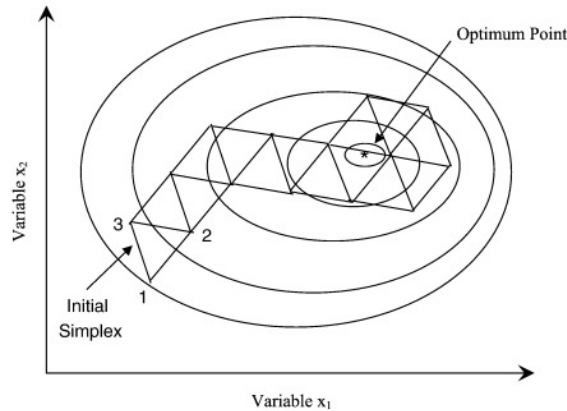


Figure 5: An example of the Nelder-Mead method [1].

In R, optimization is attained through the function Optim. Optim is a general purpose optimization based on Nelder-Mead, quasi-Newton and conjugate-gradient algorithms. For this project, the Nelder-Mead optimization was used. Figure 5 shows a visual of the Nelder-Mead method. The idea is to minimize a function of p parameters, construct a simplex of $p + 1$ vertices, and evaluate the function at each vertex. A simplex is a generalization of a notion of a triangle or tetrahedron for arbitrary dimensions. In the figure 5 the simplex is shown as a triangle. The aim of Nelder-Mead is to walk around the surface until it locates the minimum or maximum in a multidimensional space. It is a heuristic search method that converges to a non-stationary point[2].

It is important to be aware of the possibility that the optimizer could fail. Some examples include, the iteration reaching a singular gradient and being forced to stop,

or it exceeds the maximum number of iterations without converging. These are both quite easy to check.

3.2 Code and Output

As mentioned above, the Optim function in R was used to optimize the parameters according to the least squares criterion, where C_t is as in equation (2). Below, are samples from the code created as well as the output when Optim is run. In the code, X1_0 is the amount of the drug taken orally by the patient, Y1t are the concentrations in the blood taken throughout the process, and t1 are the times that the blood concentration is taken. LS.1 is the function used to obtain the least squares estimate and for the result, Optim is used.

```
### Patient 1 ###
X1_0 <- Amount[1]
Y1t <- CO[2:11]
t1 <- Time[2:11]
data1 = data.frame(t1,Y1t)
LS.1 <- function(data1,param){
  Ka = param[1]
  Ke = param[2]
  Vd = param[3]
  Ct = X1_0*Ka*(exp(-Ke*t1)-exp(-Ka*t1))/(Vd*(Ka-Ke))
  with(data1, sum((Y1t-Ct)^2))
}
result1 <- optim(par=c(1 ,0.1,1), LS.1, data=data1, hessian = TRUE)
result1
```

Figure 6: Code used for Patient 1 from the Theophylline trial.

In figure 7, the output of Optim is shown. First, `par`, is the estimates of the three parameters, K_a , K_e , V_d , in that order. Next, `value`, is the value of the sum of squares for the patient. `Counts`, is the number of iterations that occurred before convergence, and `NA` is shown because the Nelder-Mead method does not use gradients. In Optim, convergence is an integer code. If 0 is shown it means that there was a successful completion and thus convergences has taken place. If the output 10 arises for the convergence value, this means there is degeneracy of the Nelder-Mead simplex. If, instead, the value for convergence is 1, the max iterations has been reached. For Nelder-Mead, the maximum number of iterations is 500. This number can be enlarged by entering an extra "maxit" setting. Here the iteration was less than 500 and convergence occurred, thus, a larger maximum iteration value was omitted. Lastly, `Hessian`, returns the Hessian matrix for the parameters. The Hessian of the least squares is also known as the observed Fisher information matrix because it is a function of observe variables[5]. The Hessian matrix will be important for computing the variance-covariance matrix of the parameter estimates, which is used to compute the standard error and confidence intervals.

```

> result1
$par
[1] 1.77707529 0.05396418 0.36923198

$value
[1] 3.73841

$counts
function gradient
      182      NA

$convergence
[1] 0

$message
NULL

$hessian
      [,1]      [,2]      [,3]
[1,] 24.75416 21.12415 -191.5429
[2,] 21.12415 35580.76660 11978.9855
[3,] -191.54290 11978.98551 7980.2617

```

Figure 7: Output of the Optim function

As mentioned above, there are some important output possibilities to be aware of. The convergence value was already explained, and the result of patient 1 shows that Optim was able to find a local minimum. For the parameter values, par, one would hope that the first value is larger than the second value. This means that the rate that the body absorbs the drug is quicker than the rate that the body eliminates the drug. There are several reasons why this might not be the outcome. This will be discussed further in the results.

Figure 8 shows the result when the starting parameters were changed. A sensitivity analysis is then performed in order to check if it is possible to find a different local minimum. This is done by setting in different initial values into Optim and evaluating the resulting parameters. If completely different outcomes arise, it can be of interest to plot the values and see which one is closest to the data values. Figure 7 and figure 8 are from the Theophylline data. During the sensitivity analysis, the Theophylline patients had more consistent parameter estimates than the Warfarin Patients.

```

> result1 <- optim(par=c(1.3 ,1,4), LS.1, data=data1, hessian = TRUE)
> result1
$par
[1] 1.77764459 0.05396654 0.36928016

```

Figure 8: Alternate starting points when using Optim for Patient 1 from the Theophylline trial.

3.3 Confidence Intervals

Let ψ_k for $k = 1, 2, 3$ be the k -th component of the parameter vector ψ and $\hat{\psi}_k$ be its estimate. Here, $k = 1$ refers to K_a , $k = 2$ refers to K_e and $k = 3$ refers to V_d . This estimate, $\hat{\psi}_k$, is the k -th component of the least squares equation. For a maximum likelihood estimator;

$$\hat{\psi} \sim N(\psi, I^{-1}(\hat{\psi}))$$

Here, I^{-1} refers to the Fisher information matrix which is estimated by the inverted hessian matrix, H^{-1} . An estimate of the variance is then the k -th element on the diagonal of H^{-1} , here denoted with $H_{kk}^{-1}(\hat{\psi})$:

$$\widehat{Var}(\hat{\psi}_k) = H_{kk}^{-1}(\hat{\psi})$$

And the standard error is given by:

$$\widehat{s.e.}(\psi_k) = \sqrt{H_{kk}^{-1}(\hat{\psi}_k)}$$

The Hessian matrix is a square matrix of second order partial derivatives of the log likelihood function used to describe the local curvature of a function of multiple variables. The variance is calculated by taking the inverse of the Hessian matrix previously obtained from Optim. The inverse of the Hessian matrix is the covariance matrix. It is therefore important to insure that there are no negative values on the diagonal of the inverted Hessian matrix.

4 Results

4.1 Theophylline Results

Theophylline data contained samples that were tested more frequently over a shorter period of time, 24 hours. Because of this, it was easier to find estimated parameters to fit the data. As it is shown in the appendix6.2, the estimates fit the data reasonably accurately. As the plots show, the curve is more steep in the beginning of the measured time where absorption occurs. Then once the drug has been metabolized, the elimination process begins to come into effect and the concentration of Theophylline increases

quickly in the first few hours before it begins to slowly decrease. This results in the estimated curve not being as accurate in the beginning, and once the drug begins to be eliminated, the estimated curve becomes much more accurate.

Once the parameters were estimated it was possible to calculate their confidence intervals and the standard errors. The goal is for the confidence intervals to be very narrow. This result can be seen in the the 95% confidence intervals in the appendix, table 4. As mentioned above, it was more challenging to estimate the rate of absorption compared to the rate of elimination. The challenge of estimating the rate of absorption is the result of the lack of data points before reaching maximum concentration. because, for most of the patients in the Theophylline data there were only two or three points to use for the estimate of the rate of absorption, it gave a less accurate approximation. Most of the outcomes had narrow confidence intervals with the largest interval for the K_e parameter having a span of 0.1068935 and, for V_d , a span of 0.2828266.

The plotted correlations can also be found in the appendix 6.3. The three parameters were plotted against the Patients weight and sex respectively. By a simple visual inspection of the plots, no correlations were found between any of the above mentioned. This could have been a result of the lack of patients in the study.

An interesting result that was discovered from the analysis was with Patient number 9. Even the plot shows an indication as to why the concentrations could result in a more challenging estimate. Although the rate of elimination and the volume of distribution had small and accurate confidence intervals, there was a problem with the rate of absorption. This could be due to the fact that there were no values in the data that could be used to estimate the rate of absorption. For the patient, its estimates were $K_a = 8.862$, $K_e = 0.0866$, and $V_d = 0.377$ and the confidence interval for the rate of absorption was (0.59 , 17.135).

One approach to determine the cause of the large standard error was to perform a sensitivity analysis. It proved challenging to find initial values that produced estimates of interest. They would either result in the same estimated parameters or occasionally they would result in negative values on the diagonal of the covariance matrix. After several trials a different result was finally discovered. The estimated values were (0.086539245 , 8.872736848, 0.003680695), and the standard errors were (0.0115968278 1.0632046889 0.0001395465). These new standard errors are better for the rate of absorption and the volume of distribution but a worse value for the rate of elimination. Because the new estimates had such a large rate of elimination compared to absorption, the original values were kept and used for the tables in the appendix 6.1.

4.2 Warfarin Results

Compared to the Theophylline data, the Warfarin data contained multiple faults. When first evaluating the data, it was clear that some values were inaccurate. The data from Patient 10 was not included, and some of the patients had several concentration measurements at the same time positions. An example of this is shown for Patient 8 in the table 1.

Patient	Time	Concentration
8	3.0	12.7
8	3.0	8
8	6.0	12.7
8	6.0	11.5
8	9.0	12.9
8	9.0	11.4
8	12.0	11.4
8	12.0	11.0
8	24.0	9.1

Table 1: Example of problematic data entries for Warfarin.

There were similar occurrences for some of the other patients as well. It insinuates that there is a mistake here because the drug is only administered once. It would be impossible for the concentration of the drug in the body to increase, decrease and increase again without more drug being administered. Because of this, it seemed appropriate to remove the first value at time 3.0 and the second values at time 6.0, 9.0 and 12.0 from the data set. This was done for the other patients with similar data entries as well. If, however, the data appears to be incorrect but it could not be proven how, it was left as it was. An example of this was Patient 7 who had a concentration of 15 mg/l at hour 6 and 9. It would seem unlikely that the body would not eliminate any amount of the drug in 3 hours, but both values were kept in the data set.

Once the data was altered, the parameters were estimated using Optim in the same way as for the Theophylline patients. Unlike the previous drug, the results from Optim were found to be more difficult to understand. Originally, the criteria for a successful result were that; the rate of absorption was greater than the rate of elimination, all the rates were positive, and that the plots were as accurate as possible. When a sensitivity analysis was performed, the results were more temperamental than for the Theophylline trial.

An example of the temperamentality was observed with Patient 3. For all Patients from both data sets, the starting parameter values entered into Optim were (1, 0.1, 1).

Then the sensitivity analysis was performed by changing the starting parameters and evaluating the outcome. The results from the sensitivity analysis were quite different from the Warfarin patients and further evaluations were necessary. Figure 9 shows the result of Optim when the usual starting parameters were used. As it can be seen, the value for the rate of absorption, K_a is smaller than the rate of elimination, K_e . This would mean that the body does not have time to absorb any of the Warfarin before the elimination process occurs. Because of this peculiar result, the sensitivity analysis was performed to check if different starting values would give a better result.

```
> result.3 <- optim(par=c(1, 0.1, 1), LS3.t, data=data.3)
> result.3
$par
[1] 0.01400213 0.62254021 0.18467259

$value
[1] 3.879482
```

Figure 9: Optim result with starting parameters (1, 0.1, 1).

```
> result.3 <- optim(par=c(2, 0.1, 1), LS3.t, data=data.3)
> result.3
$par
[1] 0.6247028 0.0139925 8.2157699

$value
[1] 3.879504
```

Figure 10: Optim result with starting parameters (2, 0.1, 1).

Figure 10 shows the result of Optim when the starting parameters were changed to (2, 0.1, 1). This outcome appears to be better than the previous result because now the rate of absorption is larger than the rate of elimination. Now it was left to decide which result was the better one. This was decided by plotting both of the values and seeing which was the closest approximation to the Patients values. When this was done, it showed that both of the resulting values created this exact same curve.

This was an interesting result. The fact that both of the plots with different estimates created the same curve suggests that we are attempting to fit a model that is "too high" for the data. This means that the data is not "rich enough" to estimate both the values K_a and K_e . This is because, unlike for the Theophylline data, here the initial "rising" curve is not present. A solution to evaluate this could have been to instead use equation (1) and see if this gave a better fit to the data.

The other approach performed to determine the better estimates was to check the inverse of the Hessian matrix for each patient. Although the inverted Hessian matrix values were reasonable for Patient 3, other patients resulted in very absurd values. One patient in particular that had challenging values was Patient 25. Figures 11-13 show the outputs when the initial parameters were changed.


```
> solve(result.25$hessian)
      [,1]      [,2]      [,3]
[1,] -7.635513e-05  0.0001767342 -0.01204410
[2,]  1.767342e-04 -0.0003665940  0.02571583
[3,] -1.204410e-02  0.0257158280 -1.77743348
```

Figure 11: Inverse Hessian matrix from starting parameters (1, 0.1, 1) put in Optim.

```
> solve(result.25$hessian)
      [,1]      [,2]      [,3]
[1,]  3.748404e+06  2.123842e-01 -3.457125e+03
[2,]  2.123842e-01  2.699299e-06 -9.781843e-04
[3,] -3.457125e+03 -9.781843e-04  3.486098e+00
```

Figure 12: Inverse Hessian matrix from starting parameters (0.5, 0.1, 2) put in Optim.

```
> solve(result.25$hessian)
      [,1]      [,2]      [,3]
[1,]  1.009058e-05 -0.0001219416  0.001763296
[2,] -1.219416e-04  0.0022279872 -0.030135233
[3,]  1.763296e-03 -0.0301352328  0.413410330
```

Figure 13: Inverse Hessian matrix from starting parameters (0.11, 0.1, 2) put in Optim.

The first initial parameters used for Patient 25 were (1, 0.1 1), the same as for all of the other patients. Figure 11 shows the inverted Hessian matrix when using those initial parameters. The result were negative values along the diagonal. Because the inverse of the Hessian matrix estimates the covariance matrix, it is not possible for the diagonal to have negative values. This would mean that there is a negative variance and it would be impossible to obtain the standard error. Because of this, different initial parameter values were tested. Figure 12 and Figure 13 depict the inverted matrix for different initial values. For Figure 12, the initial values were (0.5, 0.1, 2) and the resulting parameter estimates were (10.54, 0.0015, 6.935). As the figure shows, the inverted Hessian matrix had a very large variance for the K_a estimate. The starting values were then changed to (0.11, 0.1, 2) which gave the estimates (0.01657, 0.09983, 1.1762) and Figure 13 shows the inverted Hessian matrix. This result appears to be better because the variance for the estimated rate of absorption is much smaller than that for the values in figure 12.

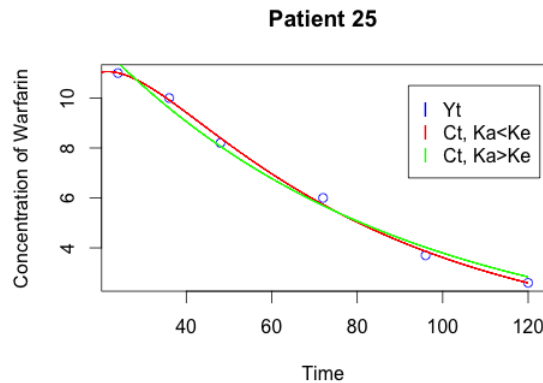


Figure 14: Plotting both of the estimated values.

In order to see if one result is better than the other, both the estimated values were also plotted in the same plot along with the the data values. Figure 14 shows the plot of all three values. It is difficult to see if one of the curves is more accurate than the other. The only thing that seems clear is that the value for the rate of absorption plays a huge part in determining the shape of the curve. This is because they are the most challenging to estimate because of the lack of data. Any change in the estimate of the rate of absorption will change the outcome of the curve. The deciding factor in determining which estimates to use was the smaller variance from the inverse Hessian matrix, and $(0.01657, 0.09983, 1.1762)$ were the values chosen for the estimates of the parameters. This same approach was used for all other patients and their values were saved in the tables in Figure 6.4 in the appendix.

The 95% confidence intervals, which can also be found in the appendix 7, were then calculated using the values from the estimated parameters and standard errors. Although some of the standard errors are larger than desired, these saved results were the smallest values found without increasing the other parameters standard errors or creating a negative value along the diagonal of the covariance matrix. Because of the lack of data depicting the rise of concentration, it was challenging to obtain smaller standard errors than those saved.

Lastly the correlations were evaluated and the plots can be found in Appendix, 6.6. As was the case for the Theophylline data, it was difficult to visually notice any strong correlations. Some possible correlations that were noticed, however, were between weight and K_e and V_d respectively. There appears to be a slight positive correlation between an increase in weight and an increase in the rate of elimination and the volume of distribution. It is also unclear whether there is a correlation between gender and the estimated parameters. This is due to the fact that of the 31 patients only 5 of them were female. When looking at the correlation plots, the points representing female patients appear to be more closely clumped together and the male patients appear more spread out. But it is difficult to see if this is solely due to the lack of female patients.

5 Conclusions

The aim of this thesis was to consider statistical inference models representing the dynamics of the concentration of Warfarin and Theophylline administered to several subjects. The models were fitted to the available data from both trials, where dynamics for the concentrations were as in equation (2). The models were fitted by using the built in function `Optim` in the program R. The estimated model parameters that were returned were then evaluated. The values were plotted and their standard errors and confidence intervals were calculated.

Estimating the parameters proved quite easy for the Theophylline data. Every subjects rate of absorption were greater than the rate of elimination and almost every individual had very small standard error values. The one exception was Patient 9 with a slightly larger standard error. The reason was because of the lack of data for the rise of the concentration. Upon evaluation, it was noticed that the first value for Patient 9 was much larger than the other subjects. After 30 minutes, the concentration for the other patients was between 1.5 and 4.5 mg/l while for Patient 9 the concentration was 7.37. Because it was unsure how the concentrations rose, it was difficult to estimate the rate of absorption.

Next, the Warfarin data was evaluated. Estimating the parameters using Optim was much more challenging, compared to Theophylline. At first glance, the data seemed to have several inflections. Some subjects had multiple concentration entries saved at the same time and some patients went several hours with the exact same concentration of Warfarin in their bodies. This could have aided in the strange results that occurred.

After the parameters were estimated the correlations were plotted. When rising concentration entries existed, the plots for the Theophylline patients appeared to be fairly accurate. As for the correlation plots, there does not seem to be any type of correlation between the estimated parameters and age, or sex. As mentioned earlier, this could be because of the small number of subjects.

Unlike the Theophylline correlations, the plots for the Warfarin patients appeared to have slight correlations. The most noticeable were for weight and rate of elimination and volume of distribution, respectively. Because of the lack of female patients, it was difficult to see if there was a correlation between gender. Although from evaluating the correlation plots, it seems as though the female patients are more concentrated than the men.

When estimating the parameters it was difficult to ensure that Optim returned the best results. Unlike for the Theophylline data, a sensitivity analysis did not suffice. It was also increasingly important to check the values along the diagonal of the inverted Hessian matrix. Occasionally the values were negative, and for most of the patients, if incorrect values were entered into Optim, the inverted Hessian would have extremely large values along the diagonal. Some values even reached the millions. When there were multiple parameter estimates that seemed plausible for a patient, the deciding factor would be to plot all of the curves and see which one was closest to the data points.

An over parameterization was the cause for the strange results. Because of the lack of rising concentration data, it was very difficult to estimate the rate of absorption. The model used for the data was more complex than needed. If equation (1) had been used instead and evaluated the data using a one compartment model, the rate, K , might

have been estimated more accurately. This seems to have been the cause for both Patient 9 from the Theophylline trial and for most of the Warfarin Patients.

5.1 Acknowledgments

I would like to give a special thank you to my supervisor Umberto Picchini for all of his help. He has shown tremendous patience with me throughout the process of this project. It has been a very enjoyable experience being taught by him.

6 Appendix

6.1 Theophylline Results

Patient	K_a	K_e	V_d
1	1.7775724	0.05394508	0.3692675
2	1.9427244	0.10165800	0.4403454
3	2.4535463	0.08142548	0.4858308
4	1.1715348	0.08746410	0.4275969
5	1.4712349	0.08843534	0.4930644
6	1.1637094	0.09952963	0.5137981
7	0.6797272	0.10224700	0.5046092
8	1.3755595	0.09195504	0.5052709
9	8.8619315	0.08662512	0.3773504
10	0.6955062	0.07396610	0.4386186
11	3.8490528	0.09812300	0.5834105
12	0.8328807	0.10557680	0.3977868

Table 2: Theophylline Estimated Parameters.

Patient	K_a	K_e	V_d
1	0.2605278	0.008810771	0.02061797
2	0.3376771	0.016298891	0.03302997
3	0.5102277	0.013522762	0.03442040
4	0.1993916	0.015716553	0.03510647
5	0.2039570	0.012838531	0.03206065
6	0.2775899	0.024262548	0.06065350
7	0.1703963	0.027268755	0.07214965
8	0.2829847	0.018676485	0.04791871
9	4.2209571	0.013862441	0.02166746
10	0.1202974	0.013956915	0.03905449
11	0.8920121	0.016531949	0.04244057
12	0.1394723	0.018029372	0.03607481

Table 3: Standard Error Results.

Patient	K_a -lower	K_a -upper	K_e -lower	K_e -upper	V_d -lower	V_d -upper
1	1.2669380	2.288207	0.03667597	0.07121419	0.3288563	0.4096787
2	1.2808773	2.604572	0.06971217	0.13360383	0.3756067	0.5050841
3	1.4535000	3.453593	0.05492087	0.10793009	0.4183668	0.5532947
4	0.7807273	1.562342	0.05665966	0.11826854	0.3587882	0.4964056
5	1.0714793	1.870991	0.06327182	0.11359886	0.4302255	0.5559033
6	0.6196332	1.707786	0.05197504	0.14708422	0.3949173	0.6326790
7	0.3457504	1.013704	0.04880024	0.15569376	0.3631959	0.6460225
8	0.8209095	1.930210	0.05534913	0.12856095	0.4113503	0.5991916
9	0.5888556	17.135007	0.05945474	0.11379550	0.3348822	0.4198187
10	0.4597234	0.931289	0.04661055	0.10132165	0.3620718	0.5151654
11	2.1007091	5.597396	0.06572038	0.13052562	0.5002270	0.6665940
12	0.5595149	1.106246	0.07023923	0.14091437	0.3270802	0.4684934

Table 4: 95% Confidence Intervals

6.2 Theophylline Patients

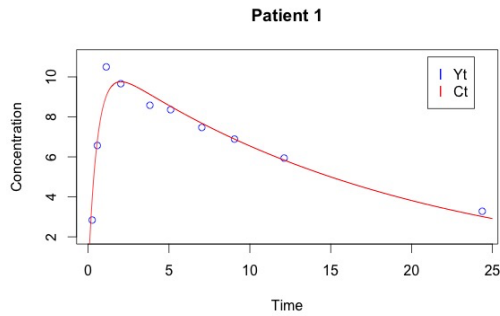


Figure 15

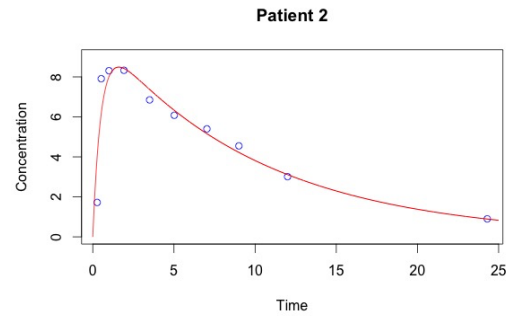


Figure 16

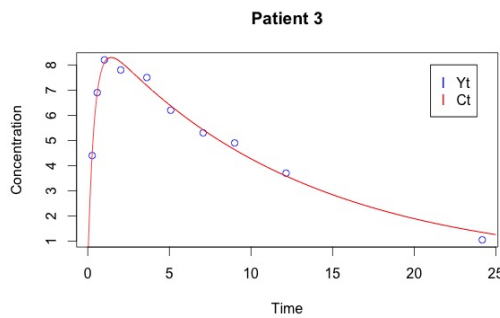


Figure 17

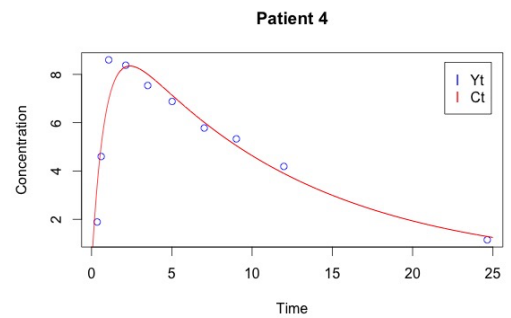
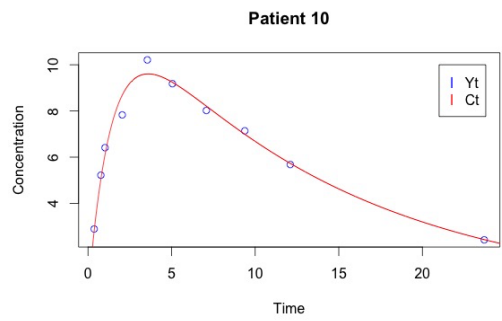
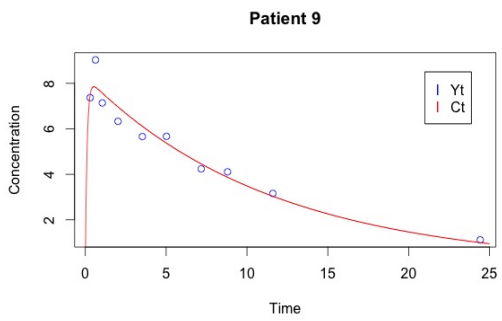
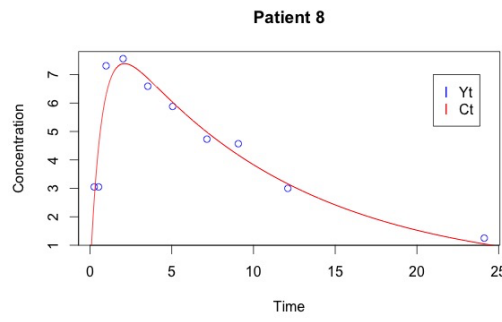
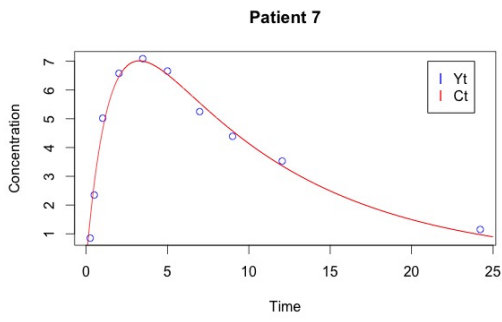
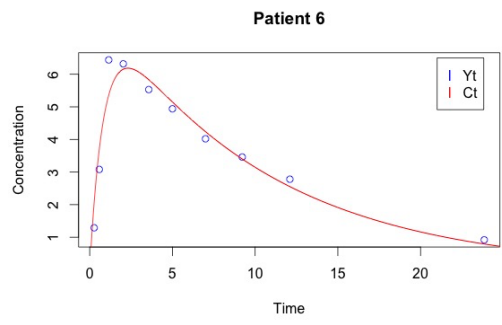
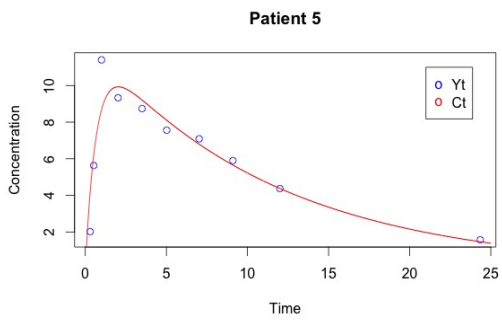


Figure 18



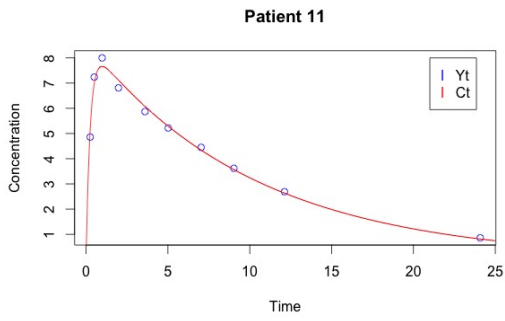


Figure 25

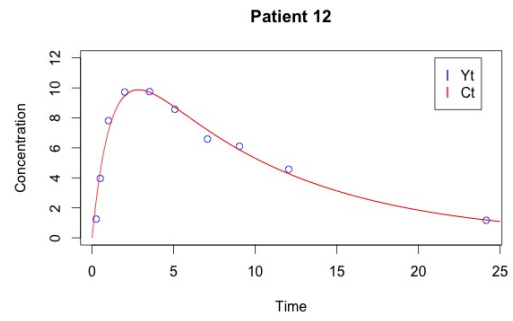


Figure 26

6.3 Theophylline Correlation Plots

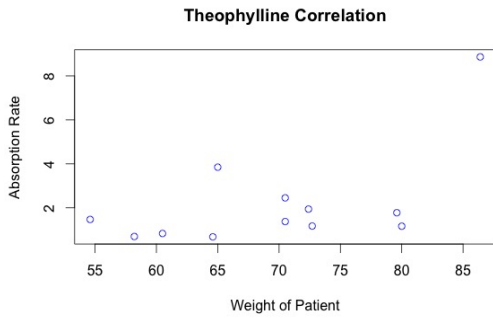


Figure 27: Correlation of rate of absorption and weight of patients

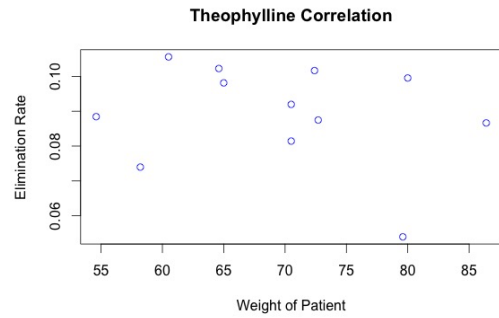


Figure 28: Correlation of rate of elimination and weight of patients

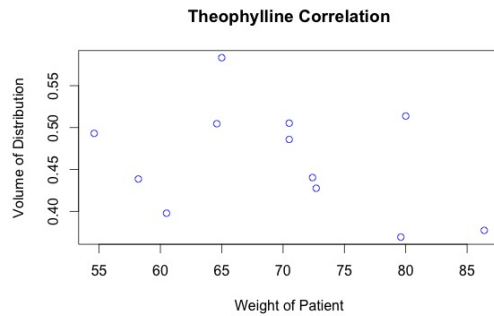


Figure 29: Correlation of variable of distribution and weight of patients

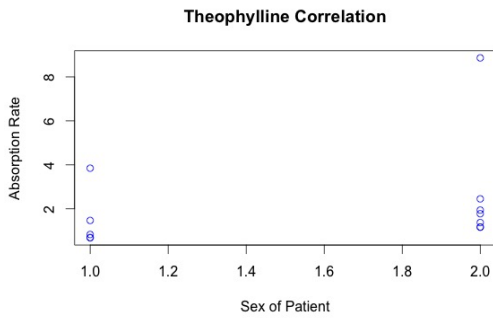


Figure 30: Correlation of rate of absorption and sex of patients

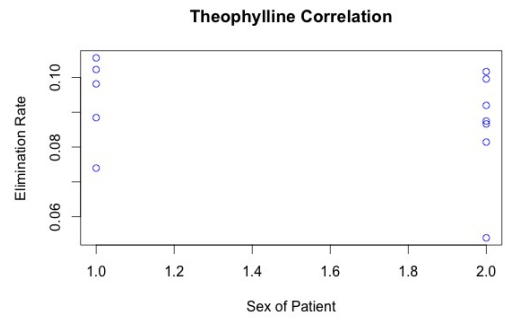


Figure 31: Correlation of rate of elimination and sex of patients

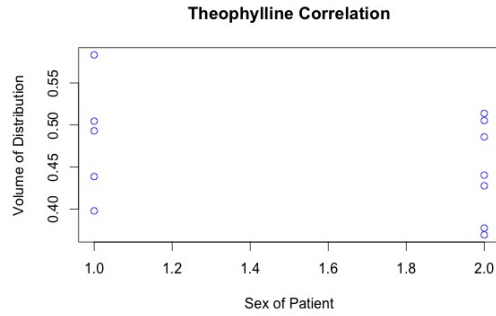


Figure 32: Correlation of variable of distribution and sex of patients

6.4 Warfarin Results

Patient	K_a	K_e	V_d
1	0.13267702	0.01446303	7.50849843
2	0.01574023	0.45836077	0.27680724
3	0.01398917	0.62400211	0.18412751
4	0.98868210	0.01452000	4.94290560
5	0.01581992	0.44091934	0.40900544
6	0.04249780	0.85868670	0.28201590
7	0.02527076	2.74939088	0.06941347
8	0.22072066	0.01622726	4.79098130
9	0.08402325	0.01192151	8.51144165
11	0.02068321	0.73728081	0.19292812
12	0.02437440	0.15404780	0.92786300
13	0.02257397	0.48867486	0.38773475
14	0.02930772	0.53260490	0.30553906
15	0.01500893	0.77133285	0.12731195
16	0.17711971	0.01691729	6.00949389
17	0.13779535	0.01372055	11.75917513
18	0.03051123	0.05903605	3.79955762
19	0.02064515	0.20299897	0.99489492
20	0.01360843	0.25598296	0.37507983
21	0.01645334	0.17475980	0.89369526
22	0.02454555	0.05168202	2.34878236
23	2.47485657	0.01520814	7.81099979
24	0.02206191	0.03182613	4.47792417
25	0.01656874	0.09987269	1.17554411
26	0.13521980	0.01465556	8.90243016
27	0.01706494	0.20505741	0.62131525
28	0.01601209	0.35004986	0.39939177
29	0.01440479	0.17920212	1.13687813
30	0.01669954	0.30977080	0.41659123
31	0.01489507	0.16658854	0.90373794
32	0.01406604	0.14077535	0.78947091

Table 5: Parameter Estimates, Warfarin

Patient	K_a	K_e	V_d
1	0.102470687	0.003034765	0.78530163
2	0.001667233	0.065640686	0.04926811
3	0.131515734	0.001281673	0.34165428
4	0.427804410	0.001456905	0.19450467
5	0.002406643	0.232079551	0.23406812
6	0.004615725	0.283952605	0.10296644
7	0.001735418	0.576380507	0.01473012
8	0.023932047	0.001554710	0.24209810
9	0.080925483	0.003875417	1.34169575
11	0.001598209	0.087167999	0.02754853
12	0.002700895	0.016743412	0.13279457
13	0.002358758	0.060700616	0.06172156
14	0.003131389	0.059281157	0.04550940
15	0.001396915	0.115969239	0.02267628
16	0.229989077	0.002799703	0.51330700
17	0.167821283	0.003658398	1.41608648
18	0.018261850	0.047744280	3.41065740
19	0.004494952	0.303764669	1.64282557
20	0.001889300	0.235373510	0.36458366
21	0.003181763	0.090054543	0.50879257
22	0.010747209	0.028664206	1.48366040
23	0.002351269	0.559196603	1.46886731
24	0.001786526	0.321204237	1.01170782
25	0.003153909	0.046779447	0.63656811
26	0.102257260	0.002733161	0.80588864
27	0.001931345	1.547194900	1.03853965
28	0.002151445	1.559955720	0.89201763
29	0.002482546	0.139036568	0.96369050
30	0.002310450	0.636186760	0.90455399
31	0.002077210	0.844886746	0.83494065
32	0.002904217	0.132124179	0.81471777

Table 6: Standard Errors Warfarin

Patient	K_a -lower	K_a -upper	K_e -lower	K_e -upper	V_d -lower	V_d -upper
1	-0.068165526	0.33351957	0.008514890	0.02041117	5.96930723	9.04768963
2	0.012472454	0.01900801	0.329705025	0.58701652	0.18024175	0.37337273
3	-0.243781668	0.27176001	0.621490031	0.62651419	-0.48551487	0.85376989
4	0.150185456	1.82717874	0.011664466	0.01737553	4.56167645	5.32413475
5	0.011102900	0.02053694	-0.013956580	0.89579526	-0.04976807	0.86777895
6	0.033450979	0.05154462	0.302139594	1.41523381	0.08020168	0.48383012
7	0.021869341	0.02867218	1.619685087	3.87909667	0.04054243	0.09828451
8	0.173813847	0.26762747	0.013180029	0.01927449	4.31646902	5.26549358
9	-0.074590696	0.24263720	0.004325692	0.01951733	5.88171797	11.14116533
11	0.017550720	0.02381570	0.566431531	0.90813009	0.13893300	0.24692324
12	0.019080645	0.02966815	0.121230712	0.18686489	0.66758565	1.18814035
13	0.017950805	0.02719713	0.369701653	0.60764807	0.26676050	0.50870900
14	0.023170197	0.03544524	0.416413833	0.64879597	0.21634063	0.39473749
15	0.012270977	0.01774688	0.544033142	0.99863256	0.08286645	0.17175745
16	-0.273658881	0.62789830	0.011429872	0.02240471	5.00341218	7.01557560
17	-0.191134364	0.46672506	0.006550089	0.02089101	8.98364563	14.53470463
18	-0.005281997	0.06630446	-0.034542738	0.15261484	-2.88533088	10.48444612
19	0.011835045	0.02945526	-0.392379781	0.79837772	-2.22504320	4.21483304
20	0.009905402	0.01731146	-0.205349119	0.71731504	-0.33950414	1.08966380
21	0.010217084	0.02268960	-0.001747104	0.35126670	-0.10353817	1.89092869
22	0.003481020	0.04561008	-0.00449824	0.10786386	-0.55919203	5.25675675
23	2.470248083	2.47946506	-1.080817203	1.11123348	4.93201986	10.68997972
24	0.018560319	0.02556350	-0.597734175	0.66138644	2.49497684	6.46087150
25	0.010387079	0.02275040	0.008184973	0.19156041	-0.07212939	2.42321761
26	-0.065204430	0.33564403	0.009298564	0.02001256	7.32288842	10.48197190
27	0.013279503	0.02085038	-2.827444594	3.23755941	-1.41422246	2.65685296
28	0.011795259	0.02022892	-2.707463352	3.40756307	-1.34896278	2.14774632
29	0.009539000	0.01927058	-0.093309554	0.45171379	-0.75195524	3.02571150
30	0.012171058	0.02122802	-0.937155249	1.55669685	-1.35633458	2.18951704
31	0.010823737	0.01896640	-1.489389482	1.82256656	-0.73274573	2.54022161
32	0.008373774	0.01975831	-0.118188041	0.39973874	-0.80737593	2.38631775

Table 7: 95% Confidence Intervals

6.5 Warfarin Patients

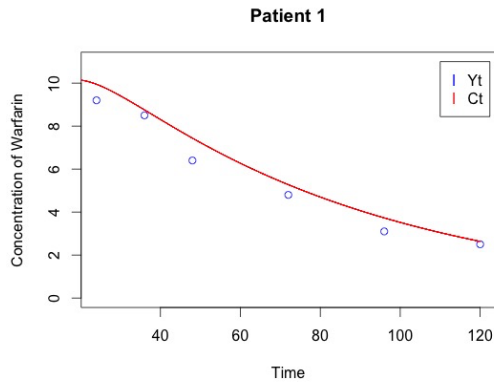


Figure 33

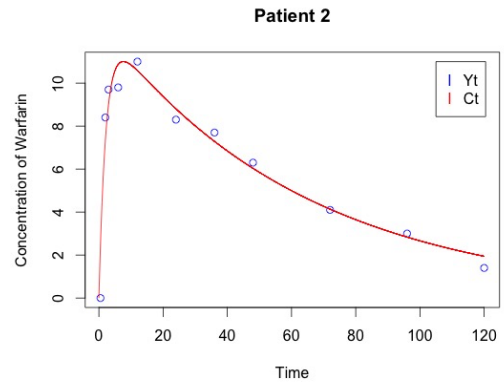


Figure 34

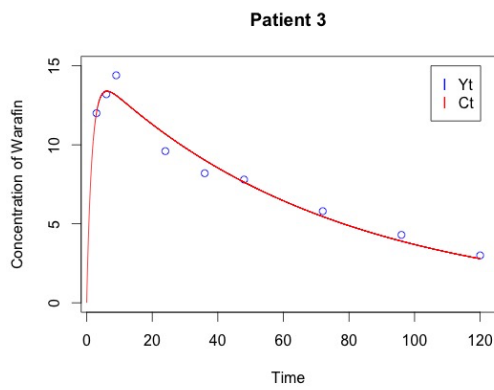


Figure 35

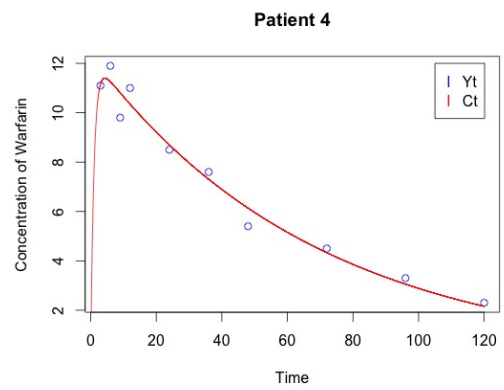


Figure 36

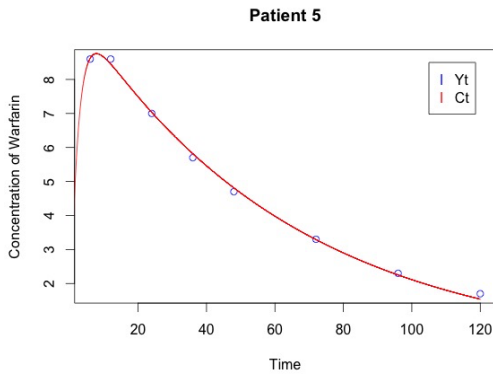


Figure 37

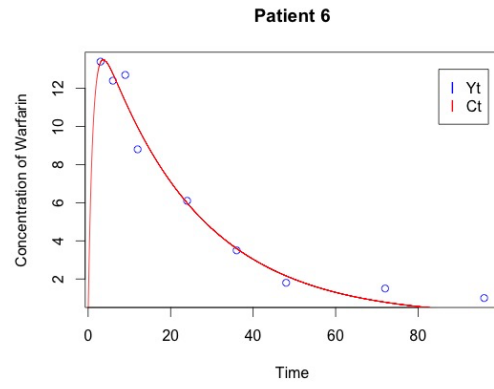


Figure 38

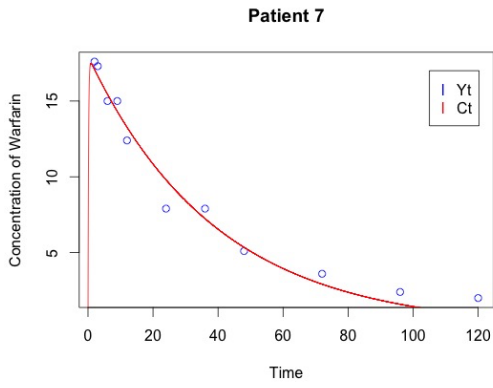


Figure 39

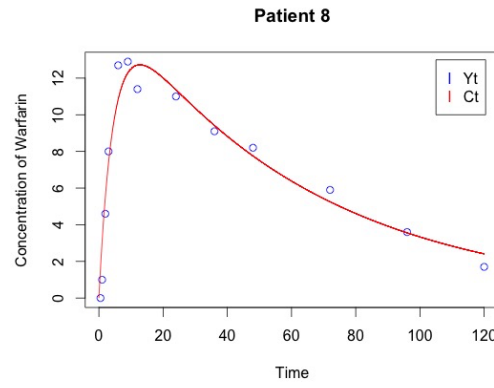


Figure 40

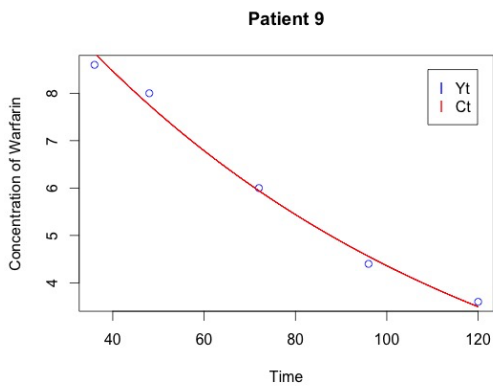


Figure 41

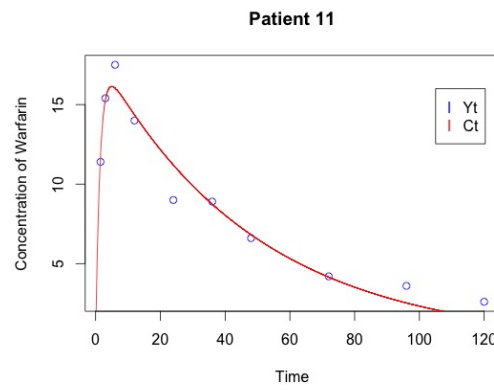


Figure 42

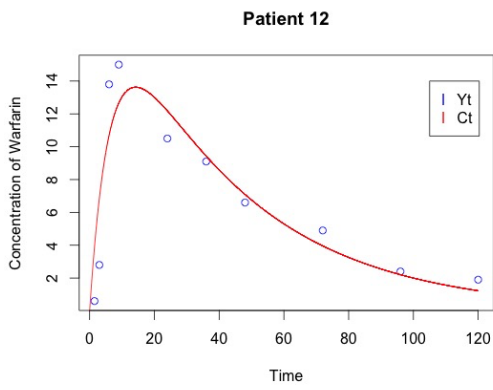


Figure 43

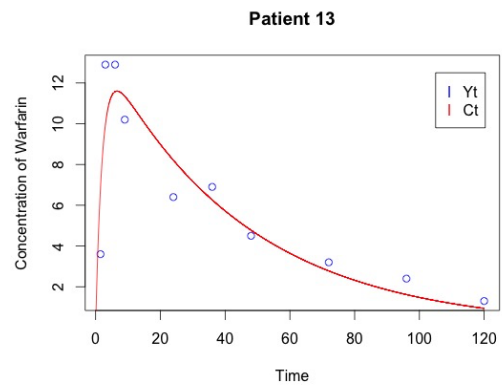


Figure 44

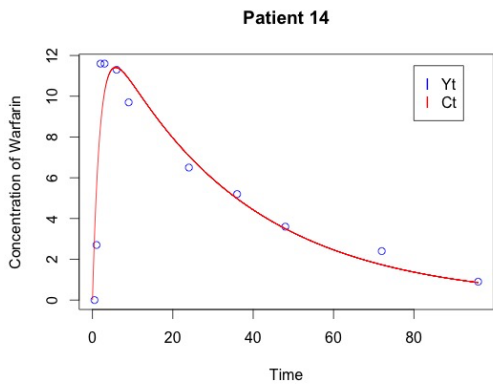


Figure 45

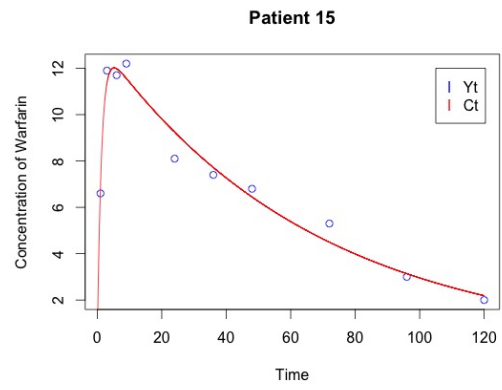


Figure 46

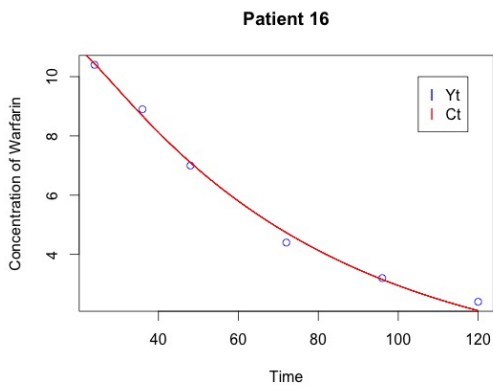


Figure 47

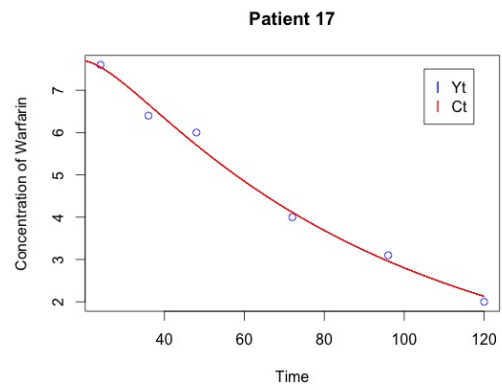


Figure 48

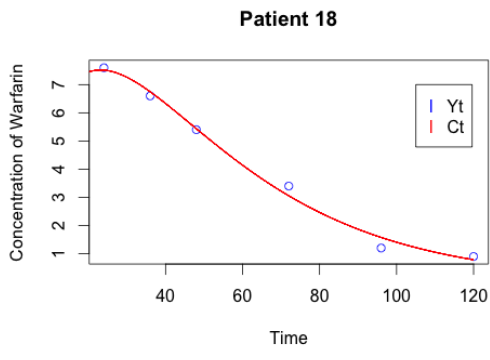


Figure 49

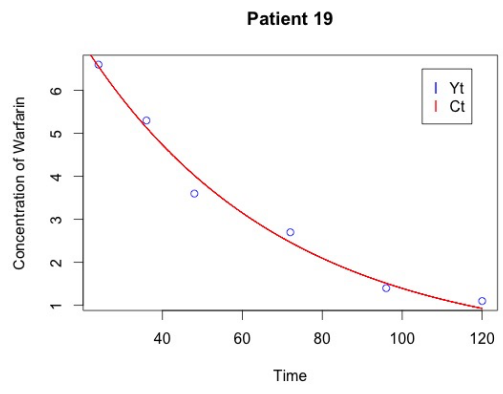


Figure 50

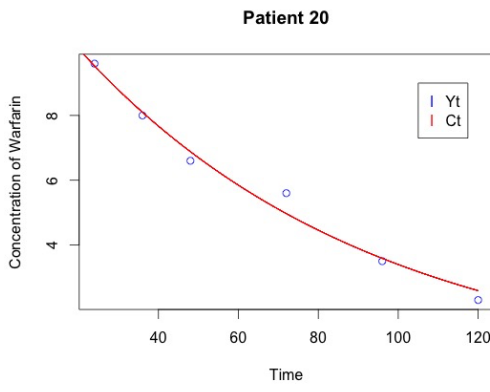


Figure 51

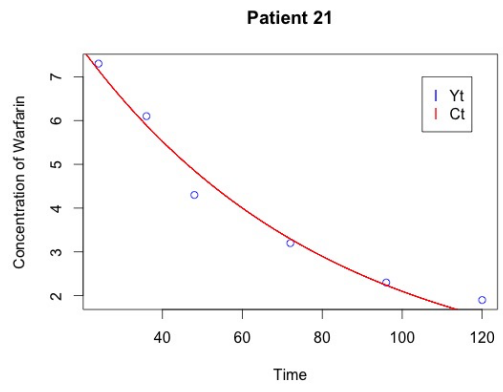


Figure 52

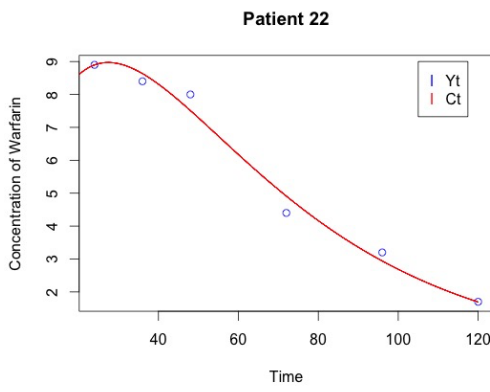


Figure 53

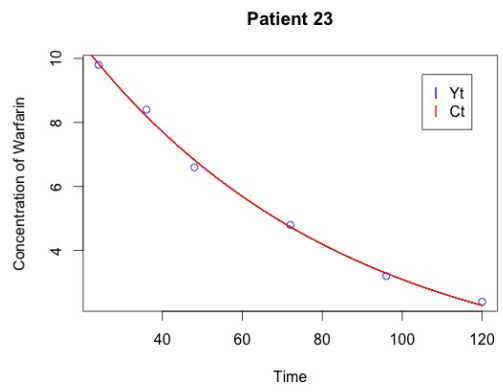


Figure 54

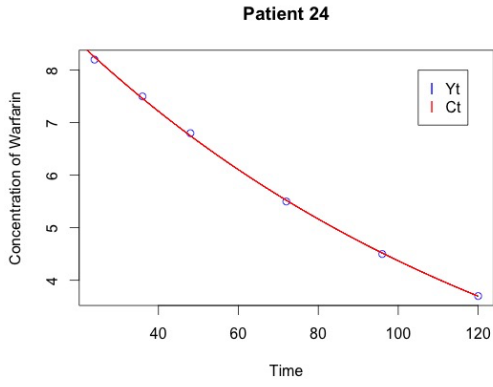


Figure 55

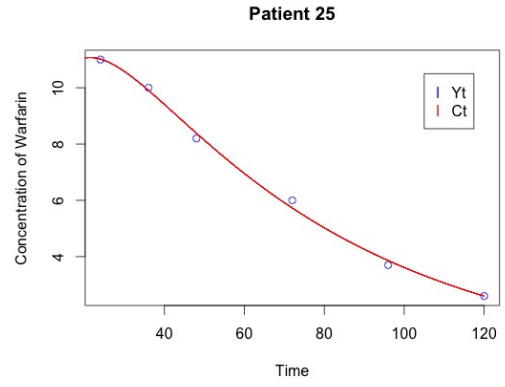


Figure 56

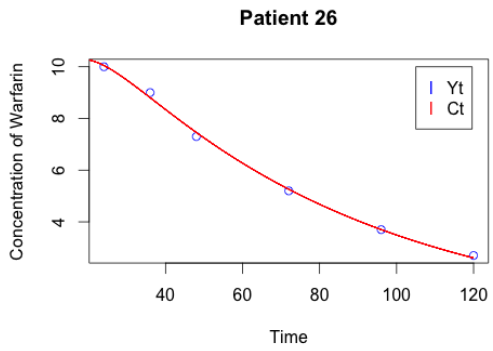


Figure 57

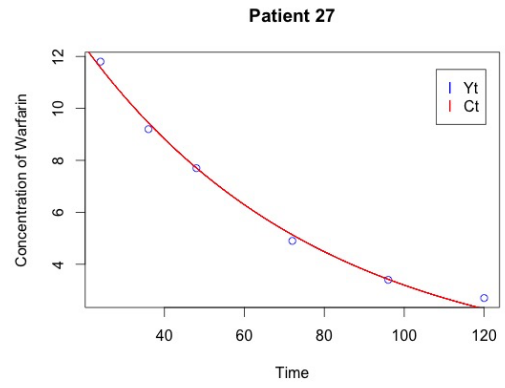


Figure 58

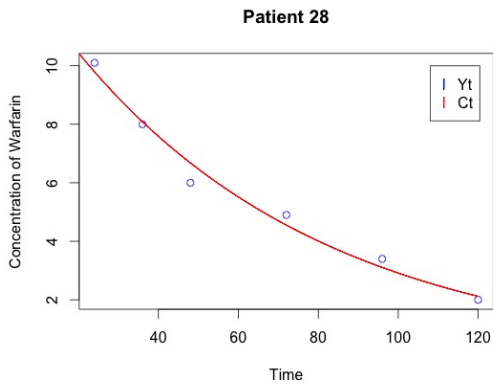


Figure 59

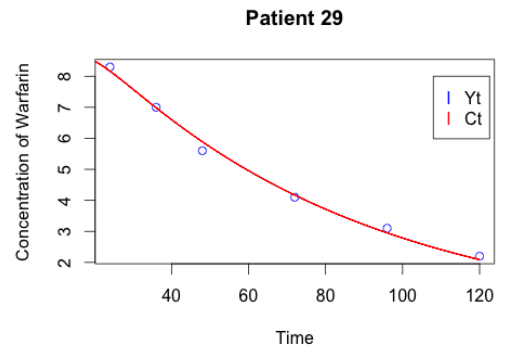


Figure 60

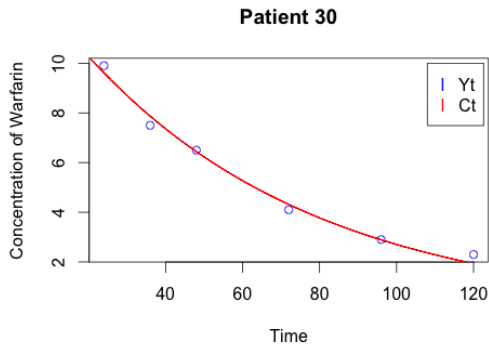


Figure 61

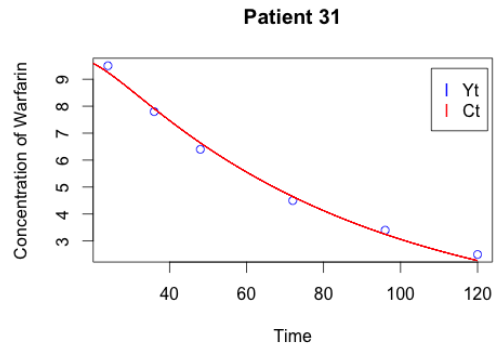


Figure 62

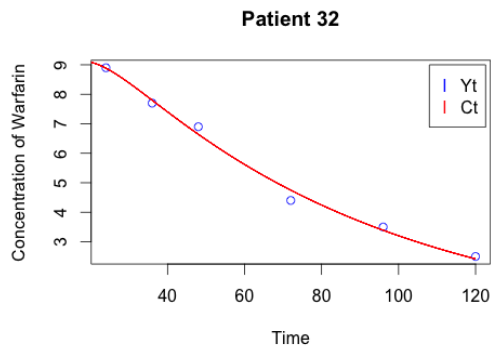


Figure 63

6.6 Warfarin Correlation Plots

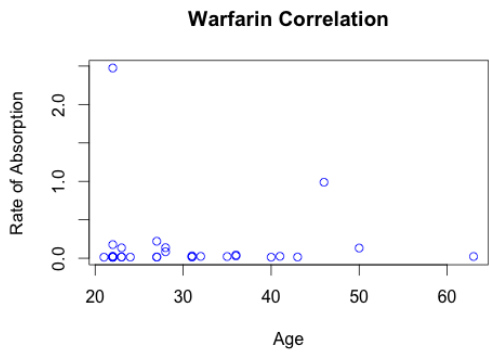


Figure 64

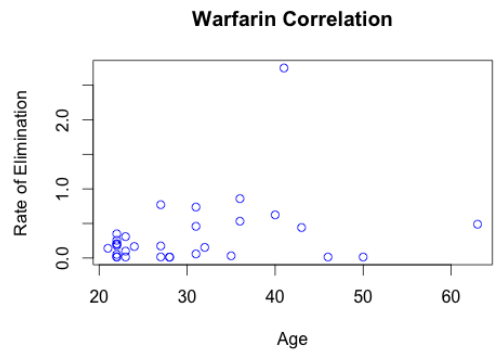


Figure 65

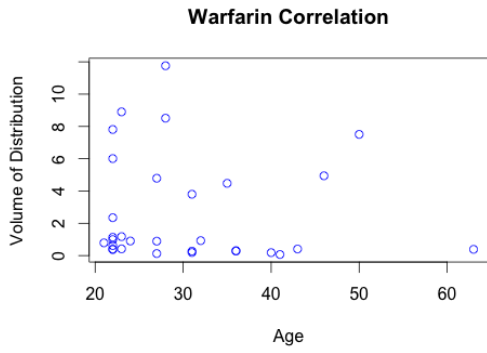


Figure 66

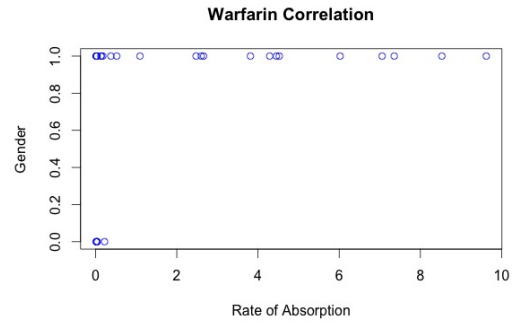


Figure 67

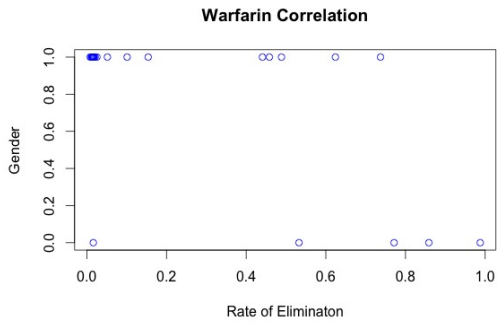


Figure 68

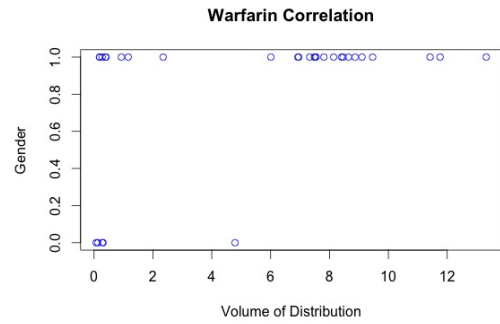


Figure 69

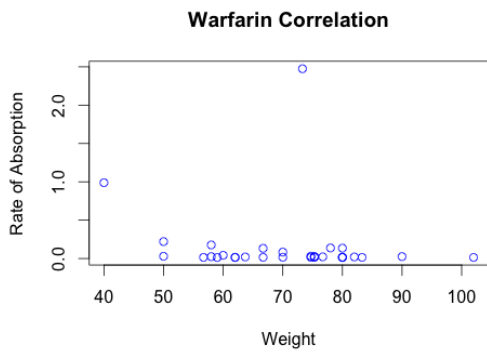


Figure 70

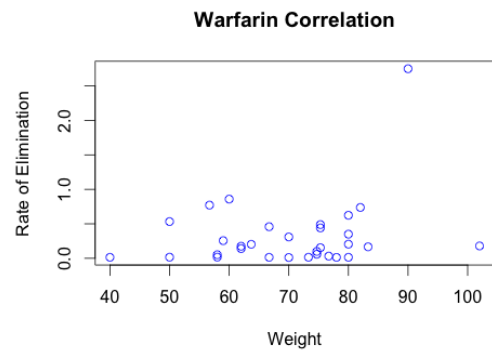


Figure 71

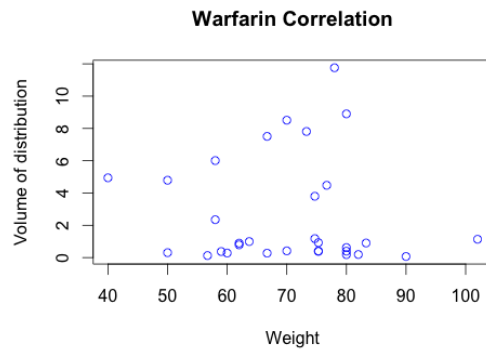


Figure 72

References

- [1] Marcos Almeida Bezerra et al. “Simplex optimization: A tutorial approach and recent applications in analytical chemistry”. In: *Microchemical Journal* (2015).
- [2] Catherine Dalzell. *Optimization in R*. 2013. URL: <https://www.ibm.com/developerworks/library/ba-optimR-john-nash/>.
- [3] Marie Davidian and David Giltinan. *Nonlinear Models for Repeated Measurement Data*. Chapman and hall, 1995.
- [4] Soraya Dhillon and Andrzej Kostrzewski. “Clinical Pharmacokinetics”. In: *Pharmaceutical Press* (2006).
- [5] Marc Lavielle. *Mixed Effects Models for the Population Approach*. Chapman and Hall/CRC, 2014.
- [6] *Pharmacokinetics*. wikipedia. 2017. URL: <https://en.wikipedia.org/wiki/Pharmacokinetics>.