CHALMERS

Deterministic and Stochastic Modeling of Insulin Sensitivity

Master's Thesis in Engineering Mathematics and Computational Science

ELÍN ÖSP VILHJÁLMSDÓTTIR

Department of Mathematical Science
Chalmers University of Technology
Gothenburg, Sweden 2013
Master's Thesis 2013

Master Thesis

Deterministic and Stochastic Modeling of Insulin Sensitivity

Elín Ösp Vilhjálmsdóttir

May 2013

Department of Mathematical Sciences
Chalmers University of Technology
Göteborg, Sweden 2013
Examiner: Ziad Taib

Biostatistics
AstraZeneca R&D
Mölndal, Sweden 2013
Supervisor: Ziad Taib

Deterministic and Stochastic Modeling of Insulin Sensitivity.

Elín Ösp Vilhjálmsdóttir

Department of Mathematical Science
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden

Abstract

Diabetes mellitus is a common disease where a person has high blood glucose levels. The disease has two main causes. The first one is inability of the pancreas to produce enough insulin. The second one is the inability of cells to respond to the insulin produced by the pancreas. In type 2 diabetes patients, the body fails to respond to insulin which results in low "insulin sensitivity". In this thesis, measurements from Intra Venous Glucose Tolerance Test (IVGTT) for both healthy subjects and type 2 diabetes patients were used together with Bergman's deterministic minimal model (ODE) to estimate the insulin sensitivity based on a nonlinear mixed effect model. In addition to the IVGTT data some basic covariates were included and tested for significance. Type 2 diabetes patients are shown to be less sensitive to insulin than healthy subjects and thus need larger amount of insulin to lower blood glucose level. A linear regression model from the covariates was used for estimating insulin sensitivity but did not give conclusive results. The covariates were included in the nonlinear mixed effect model to achieve better parameter estimates. By incorporating the covariates the estimated standard deviation for insulin sensitivity decreased substantially. An attempt was made to extend the deterministic minimal model to a stochastic differential equation (SDE) model to improve the performance and to get better parameter estimates.

Acknowledgements

First I want to express my gratitude and many thanks to my supervisor Ziad Taib for his invaluable help, support and guidance during the writing of this thesis. I also want to thank Jan Eriksson from AstraZeneca for his valuable professional opinions on modeling insulin sensitivity and AstraZeneca for providing me with topic and data on which this thesis was based. Finally, I want to thank my best Gísli Már Reynisson, family and friends for always being there for me and having patients during this thesis writing.

Elín Ösp Vilhjálmsdóttir,
Gothenburg, Sweden 13/5/31

Contents

Introduction

Diabetes is a common disease that increasingly affects many individuals worldwide. When studying diabetes it is of interest to observe how an individual responds to insulin in the process of transporting glucose to various tissues. That can be observed by measuring the insulin sensitivity which is the main topic of this thesis. Different individuals have different insulin sensitivity levels and type 2 diabetes patients tend to have substantially lower insulin sensitivity than healthy individuals. To model the interplay between insulin and glucose and be able to estimate the insulin sensitivity, the minimal model with nonlinear mixed effect model is used.

An introduction to diabetes and insulin sensitivity is presented in the first chapter. Insulin sensitivity can be modeled using the minimal model. Description of how to estimate insulin sensitivity by using the deterministic minimal model is illustrated in the second chapter. A nonlinear mixed effects model approach was chosen to estimate the parameters for the minimal model, including the insulin sensitivity parameter. A nonlinear mixed effect models with population approach allows for variation both within groups and between groups and is described in details in chapter three. A dataset from an Intra Venous Glucose Tolerance Test (IVGTT) was used to model the insulin sensitivity. The dataset included measurements both for healthy subjects and type 2 diabetes patients and results from the model are presented in chapter four. In chapter five a linear regression model for insulin sensitivity was developed both for healthy subjects and type 2 diabetes patients based on known covariates before IVGTT test. This was done in order to see if a proper estimate of the insulin sensitivity could be described from known variables since IVGTT test is extensive. Covariates were included in the minimal model in order to get a better estimate of the insulin sensitivity in chapter six. An attempt was made to use a stochastic version of the minimal model to improve the model performance and a SDE version of the minimal model is presented in chapter seven. Unfortunately some computational difficulties occurred so a simple example is presented to show that a SDE version obtains better parameter estimation.

Chapter 1

Background

1.1 Diabetes

Diabetes mellitus is a metabolic disease where a person has high blood glucose level. Diabetes is caused by two main factors, the pancreas not being able to produce enough insulin or the cells not responding to the insulin produced by the pancreas. It is about quantity and quality of insulin that the pancreas produces. There are three main types of diabetes mellitus; [1]

- **Type 1 Diabetes Mellitus:** The body fails to produce insulin at all. People with diabetes type 1 will need to take insulin injections for the rest of their life and ensure proper blood glucose levels by doing regular blood tests.
- **Type 2 Diabetes Mellitus:** The body fails to use insulin properly, the cells in the body do not react to insulin or not enough insulin is produced. This type is characterized by insulin sensitivity and is the most common type. Overweight and obese people have higher risk of developing type 2 diabetes.
- **Gestational Diabetes:** Affects woman's during pregnancy. High levels of glucose is in the blood and the body is unable to produce enough insulin to transport all of the glucose into the cells.

Type 2 diabetes mellitus (T2DM) is the only type of diabetes considered in this thesis. T2DM disease is characterized by insulin sensitivity and loss of β-cell functions resulting in hyperglycemia. The usual process is that the pancreas produces insulin, which moves glucose from the blood into the cells and is there converted into energy. This glucose disposal fails in T2DM patients since either there is not enough insulin or the insulin is not good enough. To explain this process the following factors, insulin sensitivity, glucose efficiency and pancreas responsiveness, are defined. Insulin sensitivity, S_I, is defined as the capability of insulin to increase glucose utilization to peripheral tissue as muscles

and liver. Glucose efficiency, S_G, is defined as the ability of glucose to enhance its own disposal independently of the insulin level. Finally, pancreas responsiveness is defined as the ability of the pancreatic β-cells to secrete insulin in response to glucose stimuli.

1.2 Insulin Sensitivity

Insulin sensitivity indicates how the body responds to insulin or the capability of insulin to increase glucose efficiency to cells or tissues such as muscles and liver as described before. Insulin sensitivity describes how sensitive the body is to the effects of insulin. An insulin sensitive person requires smaller amount of insulin to lower the blood glucose level than someone who has low sensitivity. A person with low sensitivity requires larger amount of insulin either from the pancreas or from injections in order to keep blood glucose stable. Insulin sensitivity varies from individual to individual and every individual needs to be tested to determine how sensitive they are to insulin. Low insulin sensitivity is also referred to as insulin resistance. From a dataset that includes glucose and insulin concentration measurements one can measure and model the insulin sensitivity.

1.3 Aim

The aim of this thesis is to investigate how good estimates of insulin sensitivity can be obtained. The idea is to model insulin sensitivity for healthy subjects and type 2 diabetes patients. More precisely, the aim is to explore an Intra Venous Glucose Tolerance Test (IVGTT) dataset from AstraZeneca which includes measurements for glucose and insulin concentrations in the blood at different time points, both before and after injection of glucose in to the blood.

The type of mathematical model used to model the insulin sensitivity is based on a compartment model thinking, similar to what is commonly used for modeling within Pharmacokinetics (PK). The model is called the minimal model and is a compartment model based on ordinary differential equations (ODEs). Nonlinear mixed effect (NLME) models will be applied to the minimal model to estimate parameters. The dataset also includes basic covariates and the aim is to obtain if including the covariates in the minimal model will give better results. By knowing more about each individual more accurate results should be obtained. An attempt will also be made to make a stochastic version of the deterministic minimal model in order to obtain an even better estimate of the insulin sensitivity and a better performance of the minimal model.

Chapter 2

Modeling Insulin Sensitivity

2.1 Modeling in Pharmacokinetics/Pharmacodynamics

Pharmacokinetics/pharmacodynamics (PK/PD) modeling aims to describe the process that occurs when a drug is injected in the body. PK/PD models are in general semi-mechanistic mathematical and/or statistical models used to describe and/or predict parameters from a dataset from the underlying experiment or trial. The purpose is to use prior knowledge and physiological interpretation of parameters to model a specific process. Pharmacokinetics describe the relationship between the drug inflow and resulting concentration and pharmacodynamics describes drug effects over time and relates the concentrations to drug effects. To describe this process compartment models are often used [2].

Compartment models function in such a way that the body or a part of the body is represented as a compartment. This is a simplification of the body structure and the compartments are used to describe the dynamics of a drug in the body. In the case of modeling insulin sensitivity the glucose or the insulin that is entering and leaving the model is of interest to measure. A remote insulin compartment is introduced to control the amount of glucose and insulin in the model [3]. The compartment model is specified using ordinary differential equations that describe the change in the model and this compartment specification enables statistical test to determine the functional form of the process. The model is shown in Figure 2.1.

Since insulin sensitivity is an important risk factor for development of type 2 diabetes it is of great interest to measure the insulin sensitivity and hopefully be able to predict if a person has high risk of developing type 2 diabetes. To measure the insulin sensitivity it is necessary to assess insulin action in the pancreas and to be able to measure the glucose or insulin entering and leaving the system. To measure this the minimal model is used to describe the behavior of the drug and the part of the body needed [4].

2.2 The Minimal Model

Using the minimal model is one way to measure insulin sensitivity [4]. The model describes the glucose production and disposal with an remote insulin compartment which is insulin dependent. This insulin dependance can be described with one compartment model as well. The structure of the minimal model makes it possible to uniquely identify model parameters that determine a best fit to glucose disappearance during the Intra Venous Glucose Tolerance Test (IVGTT).

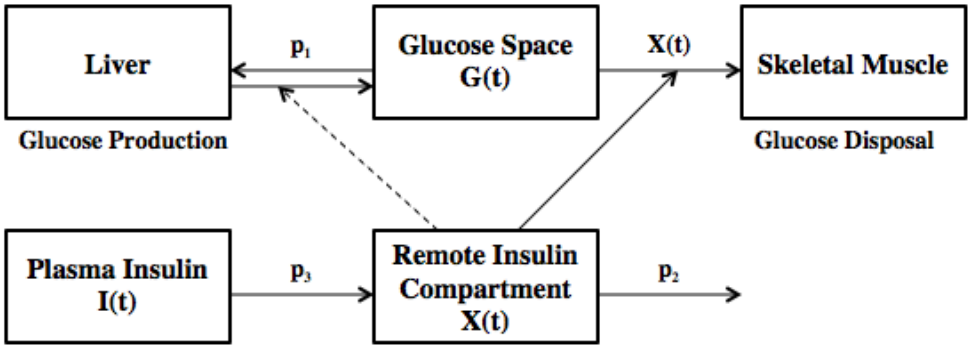


Figure 2.1: The Minimal Model. Glucose leaves and enters the glucose space, $G(t)$, at a rate proportional to the difference between plasma glucose concentration, $G(t)$, and basal plasma concentration, G_b. Glucose also disappears from the glucose space at a rate proportional to insulin concentration in the remote insulin compartment, $X(t)$ [3].

The minimal model in Figure 2.1 can be described with a set of differential equations, both for the glucose space and for the remote insulin compartment with respect to time. The equations are coupled and can be expressed mathematically as,

$$\frac{dG(t)}{dt} = -(p_1 + X(t))G(t) + p_1 G_b, \qquad G(0) = G_0 \tag{2.1}$$

$$\frac{dX(t)}{dt} = -p_2 X(t) + p_3(I(t) - I_b), \qquad X(0) = 0 \tag{2.2}$$

where

$G(t)$: Plasma glucose concentration [μU/ml] at time t.

$I(t)$: Plasma insulin concentration [μU/ml] at time t.

$X(t)$: Insulin concentration in remote compartment at time t.

G_b : Basal plasma glucose concentration, before injection [μU/ml].

I_b : Basal plasma insulin concentration, before injection [μU/ml].

The parameters p_1, p_2, p_3, G_0 and I_0, are unknown parameters in the model and will be uniquely identified from the IVGTT. S_I and S_G will be identified as

S_I : Insulin sensitivity, $S_I = p_3/p_2$

S_G : Glucose effectiveness, $S_G = p_1$

The minimal model is insulin dependent, $I(t)$, which can be described as one compartment model. The model assumes that the insulin enters the compartment at a rate proportional to the product of time passed from glucose injection and the concentration of glucose above threshold h. If the plasma glucose level drops below the threshold h the insulin is not secreted by β-cells any more. The insulin is cleared from the plasma compartment at a rate proportional to its own amount emphasizing the ability of pancreatic cells to control insulin production independently of the glucose concentration. The insulin model can be described by the differential equation

$$\frac{dI(t)}{dt} = -n(I(t) - I_b) + \gamma(G(t) - h)t, \qquad I(0) = I_0 \tag{2.3}$$

where

n : first order decay rate for insulin in plasma.

h : threshold value of glucose above which pancreas cells secrete insulin [mg/dl].

γ : rate of the pancreatic cells release of insulin when glucose concentration above threshold h.

I_b : Basal plasma insulin concentration, before injection [μU/ml].

Pancreas responsiveness is also defined to explain the process, as well as insulin sensitivity and glucose efficiency. Though it is not of interest to analyze the pancreas responsiveness at this moment it follows from the biological meaning of the insulin model parameters that the pancreas first phase responsiveness can be defined as

$$\phi_1 = \frac{I_{max} - I_b}{n(G_0 - G_b)} \tag{2.4}$$

where I_{max} is the maximum insulin response. The pancreas second phase responsiveness can be defined as

$$\phi_2 = \gamma \cdot 10^4 \tag{2.5}$$

Combining both glucose and insulin kinetics the model can then be rewritten in terms of S_G and S_I. The equations 2.1, 2.2 and 2.3 are all coupled together and interact with each other. The following nonlinear system of equations will be used to measure the insulin sensitivity, S_I [5].

$$\frac{dG(t)}{dt} = -(S_G + X(t))G(t) + S_G G_b \tag{2.6}$$

$$\frac{dX(t)}{dt} = -p_2(X(t) + S_I(I(t) - I_b)) \tag{2.7}$$

$$\frac{dI(t)}{dt} = -n(I(t) - I_b) + \gamma(G(t) - h)t \tag{2.8}$$

This is the deterministic minimal model. A stochastic version of the minimal model is introduced and described in chapter 7. The difference between the deterministic and the stochastic minimal model is that a diffusion term and a Wiener process is added to the system which allows for variation and an error in the model which can give better parameter estimation.

As already mentioned, the main interest of these coupled equations is to identify and estimate the insulin sensitivity parameter, S_I, and explore how the insulin sensitivity behaves for both healthy subjects and type 2 diabetes mellitus (T2DM) patients. S_I is thus our primary focus. Conclusions will be formulated in terms of the insulin sensitivity, S_I, and how it differs between individuals.

Chapter 3

Parameter Estimation

When applying a statistical model to a PK/PD dataset the model must allow variation both within groups and between groups. Mixed effect models allow for such variation and a population based approach of nonlinear mixed effect models will be applied to the glucose and the insulin concentration time data which can provide information about insulin sensitivity [6].

3.1 The Population Approach

To estimate the parameters of the model a dataset from Intra Venous Glucose Tolerance Test (IVGTT) will be used. The population approach allows for simultaneous estimation across individuals and this joint estimation enables a more robust parameter estimation. The theory of population modeling comes from the statistical mixed effects models, where responses of repeated measurements are modeled using fixed and random effects [2]. The fixed effects are used to describe the the population parameters but the random effects are used to account for the population variation on parameters for each individual. In general it is of interest to obtain not only individual parameters but also a quantitative description of the parameter distribution across a population.

Nonlinear mixed effect models will be applied to the minimal model described in chapter 2 as the proposed expectation function is nonlinear [6].

3.2 Nonlinear Mixed Effect Models

A mixed effect model is a statistical model containing both fixed effects and random effects where the fixed effects account for fixed parameters for every individual and the random effects account for individual deviation. Nonlinear mixed effect (NLME) models are useful in describing a nonlinear relationship between a response variable and parameters whereas the parameters estimates are allowed to vary among groups

and the parameter variation is modeled by some underlying distribution. The NLME gives information about variation of parameter values between groups and they are commonly used in PK/PD modeling because their flexible covariance allows for non-constant correlation among observation and unbalanced data. Since in our case it is expected to have both variability within individuals and between individuals, NLME model is a good choise. Using NLME model within the population approach is a way to describe variability within the population [6].

A nonlinear mixed effect model can be defined as

$$y_{ij} = f(\phi_i, x_{ij}) + \epsilon_{ij} \tag{3.1}$$

for the j'th observation from the minimal model on the i'th individual where

- y_{ij} : j'th glucose or insulin concentration observed in individual i.
- x_{ij} : time point of observation j and individual i.
- f : expected glucose $G(t)$/insulin $I(t)$ amount.
- ϕ_i : parameter vector of individual i.
- ϵ_{ij} : $\epsilon_{ij} \sim N(0, \sigma^2\Sigma)$, a noise term.

The population model for individual i can be defined as

$$\phi_i = \beta + Bb_i \tag{3.2}$$

where

- β : vector of fixed population parameters.
- B : matrix to determine which parameters have random effects.
- b_i : $b_i \sim N(0, \sigma^2\Omega)$, $\sigma^2\Omega$ is the covariance matrix of random effects.

For the i'th individual the model can be written as

$$y_i = \eta_i(\phi_i) + \epsilon_i \tag{3.3}$$

where

$$y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{bmatrix}, \eta_i(\phi_i) = \begin{bmatrix} f(\phi_i, x_{i1}) \\ f(\phi_i, x_{i2}) \\ \vdots \\ f(\phi_i, x_{in_i}) \end{bmatrix}, \epsilon_i = \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \vdots \\ \epsilon_{in_i} \end{bmatrix}$$

and $\epsilon_i \sim N(0,\sigma^2 E)$, the noise distribution for individual i. Define for all M individuals

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix}, \phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_M \end{bmatrix}, \eta(\phi) = \begin{bmatrix} \eta_1(\phi_1) \\ \eta_2(\phi_2) \\ \vdots \\ \eta_M(\phi_M) \end{bmatrix}, b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{bmatrix}$$

Then the model is

$$y \quad | \quad b \sim N(\eta(\phi),\sigma^2\Lambda) \tag{3.4}$$

$$\phi = \beta + Bb \tag{3.5}$$

$$b \sim N(\eta(\phi), \sigma^2\tilde{\Omega}) \tag{3.6}$$

where Λ=$diag(E,E,\ldots,E)$, $\tilde{\Omega} = diag(\Omega,\Omega,\ldots,\Omega)$ and $B = diag(B,B,\ldots,B)$.

To estimate the parameters the maximum likelihood function is defined with respect to the marginal distribution of y

$$p(y) = \int p(y|b)p(b)db \tag{3.7}$$

The difficulty in estimating the parameters based on maximum likelihood lies in the fact that exact calculations of the integral is very difficult. This is due to the fact that the expectation function η is nonlinear in b so there is no closed form expression for this density. Instead the conditional distribution of y is approximated for b near $\hat{b}$ by a multivariate normal distribution with expectation that is linear in b. Thus, first linearization is applied to the residual

$$y - \eta(\beta + Bb) \approx y - [\eta(\beta + B\hat{b}) + \hat{Z}b - \hat{Z}\hat{b}] \tag{3.8}$$

where

$$\hat{Z}_i = \left.\frac{\partial \eta_i}{\partial b_i^T}\right|_{\hat{\beta},\hat{b}} = \left(\left.\frac{\partial \eta_i}{\partial \phi^T}\right|_{\hat{\beta},\hat{b}}\right) B_i$$

and

$$\hat{Z} = diag(\hat{Z}_1, \hat{Z}_2, \ldots, \hat{Z}_M) = \left.\frac{\partial \eta}{\partial b^T}\right|_{\hat{\beta},\hat{b}}$$

Then

$$y - \eta(\beta + B\hat{b}) + \hat{Z}b - \hat{Z}\hat{b} \quad | \quad b \sim N(0,\sigma^2\Lambda)$$

and the approximate conditional distribution of y is

$$y \quad | \quad b \sim N(\eta(\beta + B\hat{b}) + \hat{Z}b - \hat{Z}\hat{b}, \sigma^2\Lambda)$$

The marginal distribution of y can then be approximated, from this expression along with the distribution of b, as

$$y \sim N(\eta(\beta + B\hat{b}) - \hat{Z}\hat{b}, \sigma^2\hat{V}) \tag{3.9}$$

where $\hat{V} = \Lambda + \hat{Z}\hat{\Omega}\hat{Z}^T$. The log-likelihood function corresponding to the approximate marginal distribution in equation 3.9 is

$$l_F(\beta, \sigma, b \mid y) = -\frac{1}{2}\log|\sigma^2\hat{V}| - \frac{1}{2}\sigma^{-2}[y - \eta(\beta + B\hat{b}) + \hat{Z}\hat{b}]^T\hat{V}^{-1}[y - \eta(\beta + B\hat{b}) + \hat{Z}\hat{b}] \tag{3.10}$$

The log-likelihood function 3.10 is then optimized in order to get the parameter estimates in the minimal model. The add-on package **nlme** in R was used to get parameter estimation [6]. The **nlme** package uses the first order condition linearization method to maximize the likelihood function [7].

The NLME model will be applied to the minimal model in order to estimate the parameters S_I, p_2 and S_G. It is of interest to not only obtain individual parameters but also a quantitative description of the parameter distribution across the population.

Chapter 4

Modeling Data

4.1 Data

It is important to estimate the parameters from a reliable dataset. A dataset with Intra Venous Glucose Tolerance Test (IVGTT) is used in this thesis to model the insulin sensitivity, S_I. The dataset includes measurements of glucose and insulin concentration which were frequently measured after glucose injection. In the study the usefulness of glucose stimulation for assessment of dynamic cell function was evaluated in young T2DM patients, with different disease duration and treatments, as well as in healthy subjects.

The study included 19 healthy subjects and 46 T2DM patients. The T2DM patients were 15-34 years old, had disease duration two to ten years, classified as having T2DM by the reporting physician and HbA1c $< 10\%$ [1]. The healthy subjects were 25-50 years old, had a BMI 19-40 kg/m^2 and no diabetes. Each individual was measured before injection of the glucose and after 3, 4, 5, 7, 10, 15, 20, 25, 30, 60, 115 and 120 minutes. The amount of glucose injected to both healthy subjects and T2DM patients was 0.3 g/kg. Both insulin and glucose concentration are measured at each time point and μU/ml was the measure unit. The dataset also included gender and BMI for both healthy subjects and T2DM patients as well as disease duration and different treatments for T2DM patients.

The model's parameters were estimated using the above described dataset.

4.2 Parameter Estimation

The parameters were estimated using the dataset described earlier, both for healthy subjects and T2DM patients. The results were as follows.

[1] HbA1c is the amount of glucose that sticks to the red blood cells [8].

4.2.1 Healthy Subjects

An analysis was conducted for the healthy subject data, which is shown in Figure 4.1.

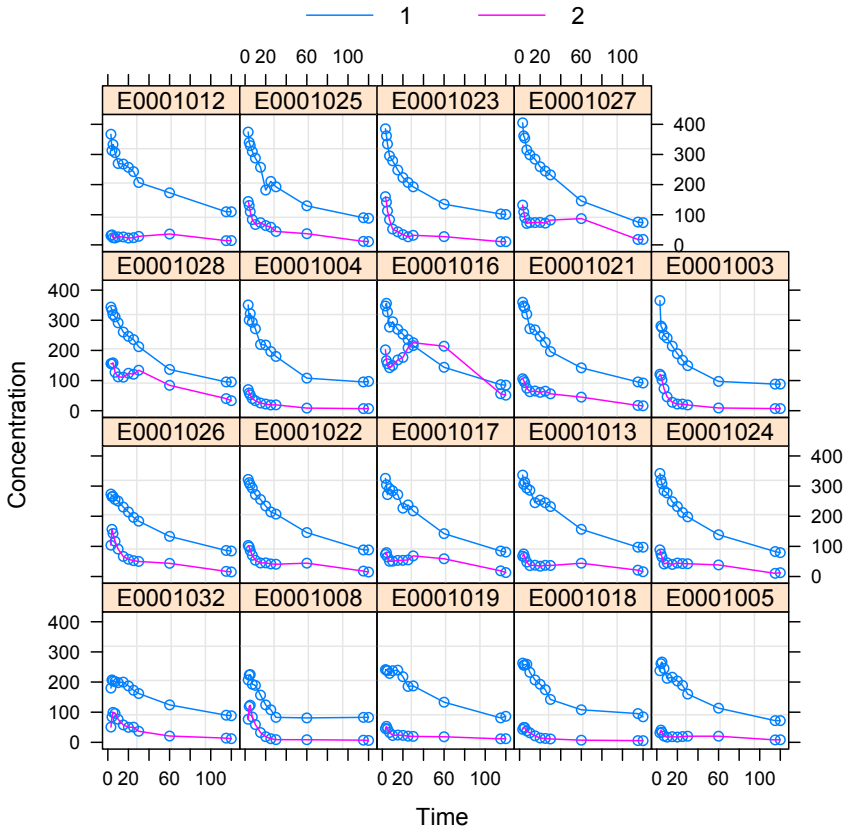


Figure 4.1: Plot of the insulin and glucose concentration for each healthy subject, were the glucose concentration is shown in blue and the insulin concentration is shown in pink [μU/ml].

To start estimating it is important to have good initial values. The initial values were chosen assuming that the reported values were within a normal range for healthy subjects [5]. A convergence of the optimization method used to maximize the log likelihood function depends on the assumed initial values, therefore it is vital that reliable values are used.

Parameter	S_{Gi}	p_{2i}	S_{Ii}	I_{0i}	G_{0i}	n_i	γ_i	h_i
Initial Value	0.014	0.074	0.00035	88	292	0.1	0.0007	90

Table 4.1: Initial values for healthy subjects.

Random effects were assumed for the key parameters, S_g, S_i, I_0 and G_0. The reason for not having random effects for p_2 is that the standard deviation was so small which implies that the parameter is not random. Parameter values for the population estimates, means

and standard deviations, are shown in Table 4.2.

	S_G $[10^{-2}]$ $\sigma_{log}(S_G)$	p_2 $[10^{-1}]$	S_I $[10^{-4}]$ $\sigma_{log}(S_I)$	I_0 $\sigma_{log}(I_0)$	G_0 $\sigma_{log}(G_0)$
Mean	2.096110	8.174785	1.926484	89.64738	306.2856
SD	0.576273		0.2254188	0.1546452	0.5028193

Table 4.2: Parameter values for the population estimates, means and standard deviations for healthy subjects.

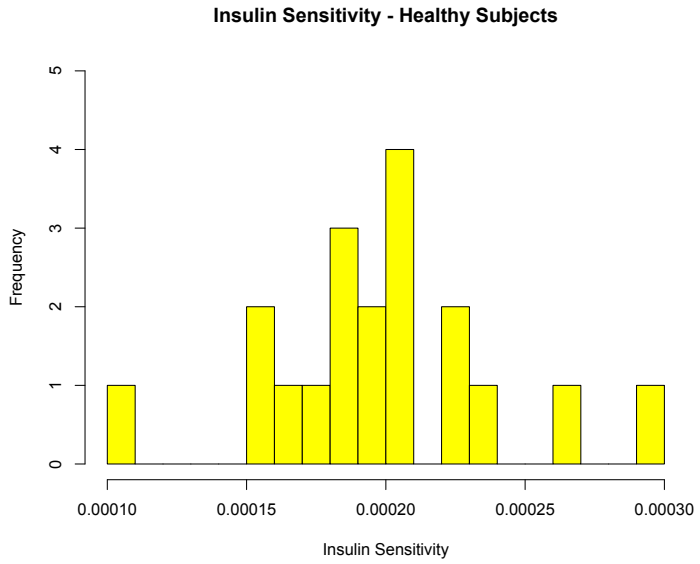
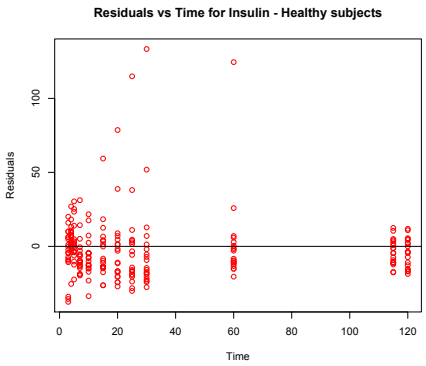
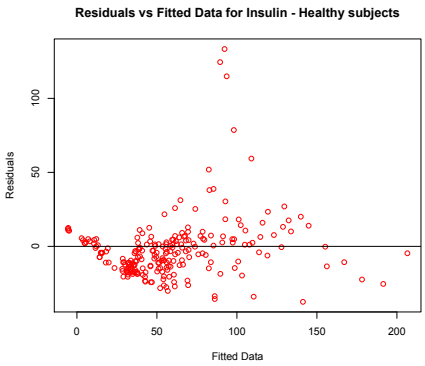


Figure 4.2: Histogram of the values for S_I for healthy subjects.

A histogram of the insulin sensitivity values, S_I, for each individual is shown in Figure 4.2. The values vary from 0.0001 to 0.0003 and follow in some way the normal distribution. To further analyze the results residual plots for both glucose and insulin concentration and observed data are plotted.



(a) Residuals vs Time for Insulin.



(b) Residuals vs Fitted Data for Insulin.

(c) Observed Data vs Fitted Data for Insulin.

Figure 4.3: Plots for the Insulin concentration for healthy subjects.

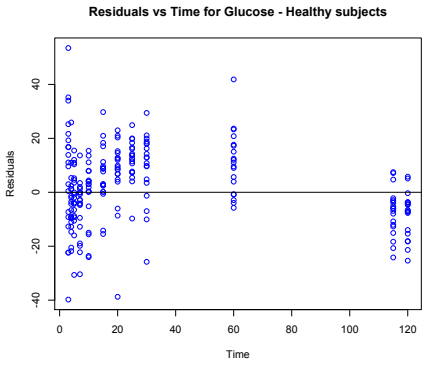
By looking at Figure 4.3 (b) one can see that the residuals form a slight curved pattern and that there are few outliers. Moreover, one can see that the points are always above or below the real values. The reason for the observed curved pattern in Figure 4.3 (b) is that the observed values are either higher or lower than the predicted values at each point in time. Figure 4.3 (c) shows that the most variation is for high values of insulin concentration and a curved pattern is observed for the same reason as in the residual plot.

An Anderson-Darling hypothesis test was used to test if the observed residuals of the insulin concentration were normally distributed. The Anderson-Darling test is a statistical test of whether a given sample of data is drawn from a given probability distribution, in this case normal distribution. Before using the hypothesis test, one can visually inspect from Figure 4.3 (b) in the attempt to determine if the observed residuals are seemingly randomly scattered or if they form a structured pattern, which would imply that they are less likely to follow a normal distribution. The hypothesis are

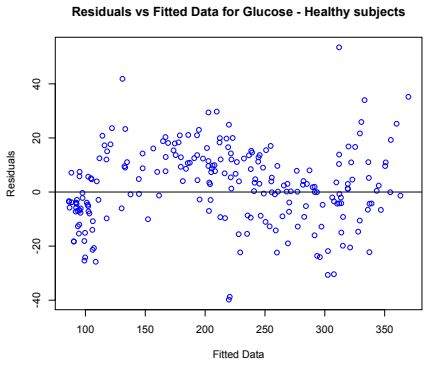
H_0 : The residuals follow a standard normal distribution, $\epsilon \sim N(0,1)$.

H_A : The residuals do not follow the standard normal distribution.

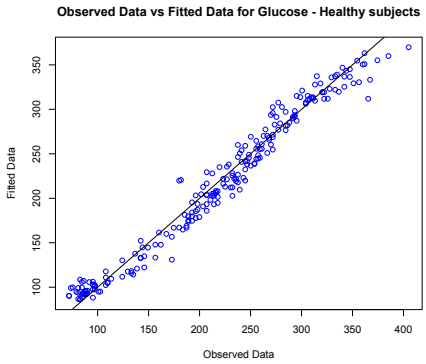
The resulting p-value was $2.2 \cdot 10^{-16}$ with the test statistic 10.5738. Since the p-value is so small it is not possible to reject the null hypothesis, H_0, i.e the residuals are standard normally distributed.



(a) Residuals vs Time for Glucose.



(b) Residuals vs Fitted Data for Glucose.



(c) Observed Data vs Fitted Data for Glucose.

Figure 4.4: Plots for the Glucose concentration for healthy subjects.

From Figure 4.4 (b) one can see that the residuals form a obvious curved pattern. The curve pattern occurs for the same reason as for the insulin concentration but since it is far more noticeable and stronger the residuals are not likely to bee standard normally distributed. Figure 4.4 (c) has an even better fit than the insulin concentration but still has the curved pattern.

The Anderson-Darling test was performed again to test if the residuals follow a standard normal distribution and one can visually inspect from Figure 4.4 (b) in the attempt to determine if the observed residuals are seemingly randomly scattered or if they form a structured pattern as before. The hypothesis were the same as for the insulin

concentration. The resulting p-value was 0.4603 with the test statistic 0.3539. For significance level 0.05 the null hypothesis, H_0, is rejected and the residuals are less likely to follow the standard normal distribution.

4.2.2 T2DM Patients

The same analysis was conducted for the T2DM patients data, which is shown in Figure 4.5.

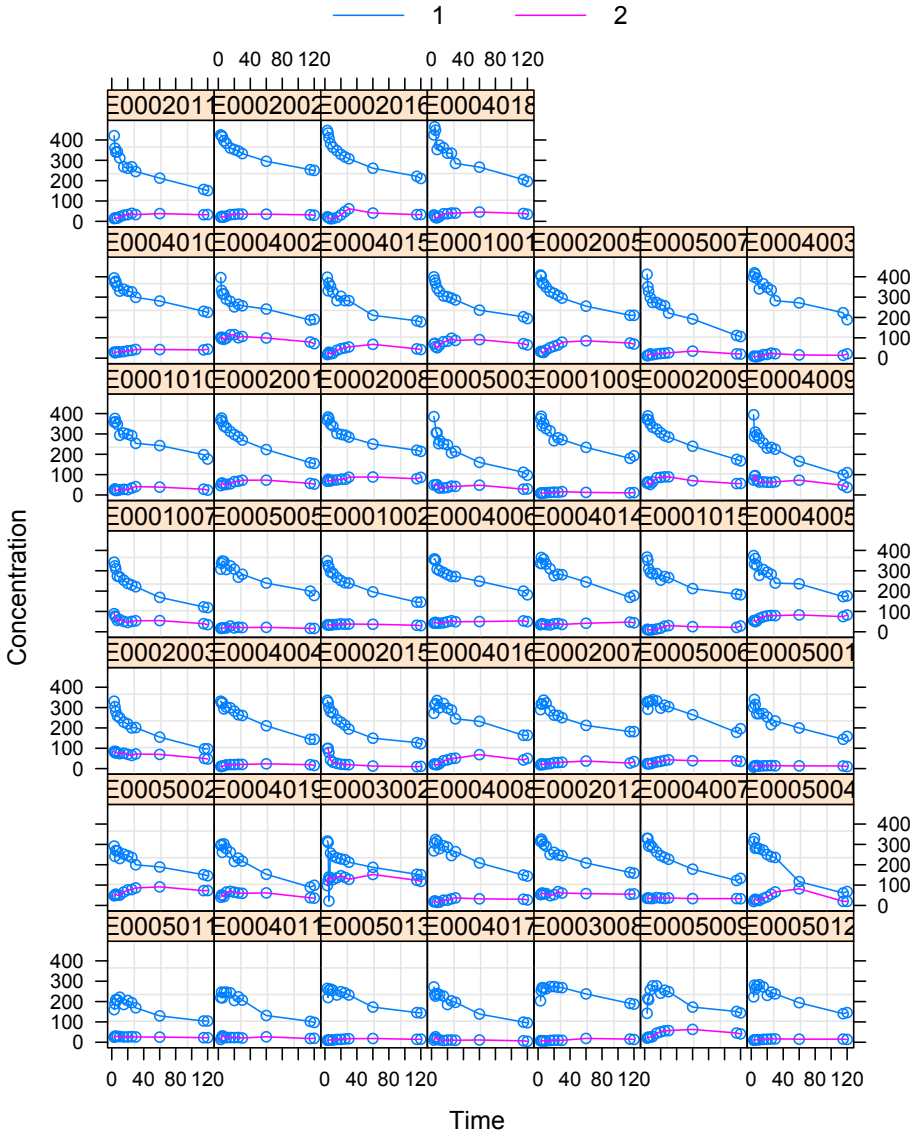


Figure 4.5: Plot of the insulin and glucose concentration for each T2DM patient, were the glucose concentration is shown in blue and the insulin concentration is shown in pink [μU/ml].

The same initial values were used for the T2DM patients as for the healthy subjects and

are shown in Table 4.1. To have the T2DM patients comparable to healthy subjects it is suitable to have the same initial values. Random effects were put on the key parameters, S_g, S_i, I_0 and G_0. Parameter value for the population estimates, means and standard deviations, are shown in Table 4.3.

	S_G $[10^{-2}]$ $\sigma_{log}(S_G)$	p_2	S_I $[10^{-5}]$ $\sigma_{log}(S_I)$	I_0 $\sigma_{log}(I_0)$	G_0 $\sigma_{log}(G_0)$
Mean	1.759684	459780.7	2.78773	44.57617	342.5570
SD	0.2322227		2.403692	0.6410483	0.1276996

Table 4.3: Parameter values for the population estimates, means and standard deviations for T2DM patients.

The value of p_2 explodes since the value of S_I is much smaller than for the healthy subjects. From Table 4.3 it can bee seen that the estimated standard deviation for S_I is much higher for the T2DM patients then for healthy subjects. Each individual is really different and they vary a lot.

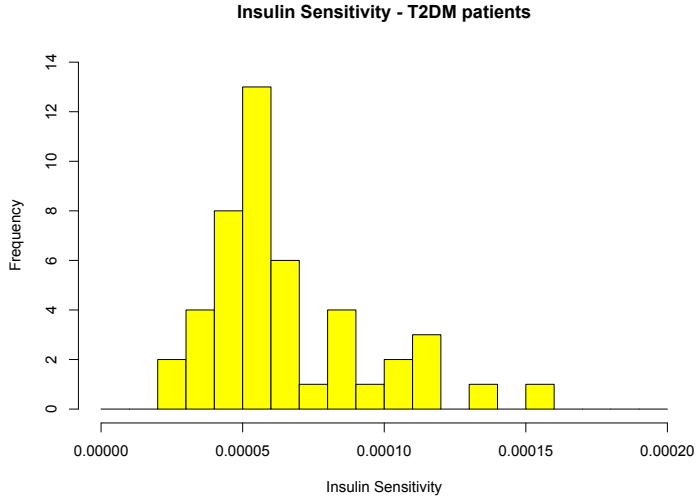
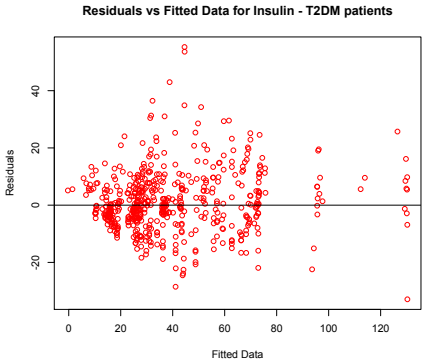


Figure 4.6: Histogram of the values for S_I for T2DM patients.

The histogram of the insulin sensitivity values, S_I for each individual is shown in Figure 4.6. The values vary from 0.000001 to 0.00017. Here the values are much lower than for healthy subjects, T2DM patients are seemingly less sensitive to insulin than healthy people. The values do not seem to follow the normal distribution as clearly as for the healthy subjects. The T2DM patients are thus more unpredictable then the healthy subjects.

To further analyze the results, residual plots for both glucose and insulin concentration and observed data versus fitted data were plotted as for the healthy patients.

(a) Residuals vs Time for Insulin.



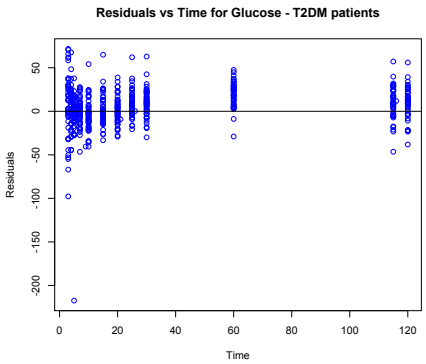
(b) Residuals vs Fitted Data for Insulin.

(c) Observed Data vs Fitted Data for Insulin.

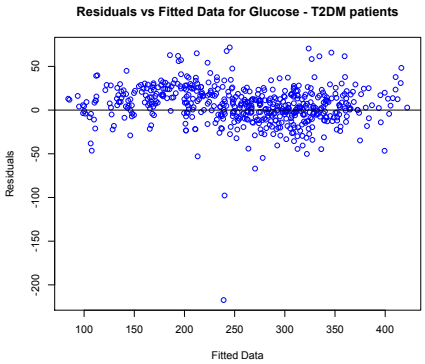
Figure 4.7: Plots for the Insulin concentration for T2DM patients.

The residuals are more randomly spread in Figure 4.7 (b) for T2DM patients than for healthy subject. The values vary more since the estimated standard deviation for T2DM patients is higher than for healthy subjects as can be seen in Table 4.3. Figure 4.7 (c) shows that there is much difference between the observed values and the fitted values.

An Anderson Darling test was used to test if the residuals follow a normal distribution as for the healthy subjects. Before using the hypothesis test, one can visually inspect from Figure 4.7 (b) in the attempt to determine if the observed residuals are seemingly randomly scattered or if they form a structured pattern, which would imply that they are less likely to follow a normal distribution. The resulting p-value was $2.181 \cdot 10^{-12}$ with the test statistic 5.005. For all significance levels the null hypothesis, H_0, cannot be rejected and the residuals thus follow the standard normal distribution.



(a) Residuals vs Time for Glucose.



(b) Residuals vs Fitted Data for Glucose.

(c) Observed Data vs Fitted Data for Glucose.

Figure 4.8: Plots for the Glucose concentration for T2DM patients.

From Figure 4.8 (b) one can see that the residuals form a slight curved pattern. As before the reason is because the model depends on ordinary differential equations and the observed values are either higher or lower than the predicted values at each point in time. Figure 4.8 (c) shows one outliner but otherwise there is a good fit.

As before an Anderson Darling test was used to test if the residuals follow a normal distribution as for the healthy subjects. Before using the hypothesis test, one can visually inspect from Figure 4.8 (b) in the attempt to determine if the observed residuals are seemingly randomly scattered or if they form a structured pattern, which would imply that they are less likely to follow a normal distribution. The resulting p-value was $2.181 \cdot 10^{-12}$ with the test statistic 5.005. For all significance levels the null hypothesis, H_0, cannot be rejected and the residuals thus follow the standard normal distribution.

4.2.3 Comparing Healthy Subjects and T2DM Patients

From Figures 4.3, 4.4, 4.7 and 4.8 one can conclude that the minimal model performs better for glucose concentration than insulin concentration both for healthy subjects and T2DM patients. Also better parameter estimates are obtained for healthy subjects than for T2DM patients as was expected since the estimated standard deviation is substantially lower for healthy subjects than for T2DM patients. An evident difference can be seen in healthy subjects and T2DM patients, the insulin sensitivity is much lower for T2DM patients which confirms that T2DM patients need more insulin or glucose to maintain the appropriate level needed. The curved pattern that forms on the residuals can be seen more obvious for the healthy subjects than for T2DM patients. One explanation that the T2DM patients vary much more in how sensitive they are to insulin. The healthy subjects are more similar to each other than the T2DM patients.

4.3 Grouping Data

The dataset was divided into groups and tested if there was a significant difference between groups. For both healthy subjects and T2DM patients BMI and gender were tested for significance and for T2DM patients the significant difference for different treatments and disease duration was also tested. The groups were tested using ANOVA and pairwise comparison to see if there was a significant difference between all groups and pairs of groups respectively.

4.3.1 BMI

For both healthy subjects and T2DM patients it is of interest to see if insulin sensitivity is different dependent on BMI. Four BMI groups were created, normal weight with BMI 18-25, overweight with BMI 25-30, obese with BMI 30-35 and extremely obese with BMI >35.

Healthy Subjects

First consider the four BMI groups created for the healthy subjects. Box plots of the insulin sensitivity for each BMI group are shown in Figure 4.9.

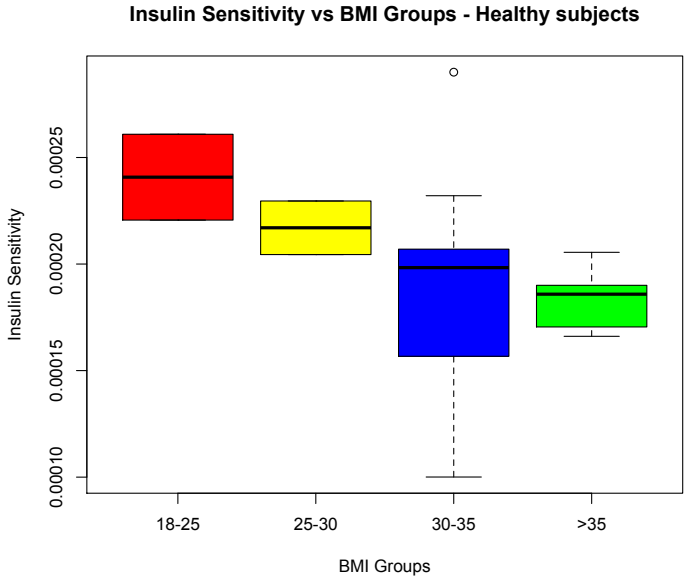


Figure 4.9: Box plots of insulin sensitivity for BMI groups, healthy subjects.

The mean of insulin sensitivity for each BMI group was calculated and is shown in Table 4.4.

	# **individuals**	**Mean** $[10^{-4}]$
Normal weight: BMI 18-25	2	2.407681
Overweigh: BMI 25-30	2	2.17021
Obese: BMI 30-35	9	1.916665
Extremely Obese: BMI >35	6	1.839679

Table 4.4: The means of insulin sensitivity for different BMI groups, healthy subjects.

Analysis of variance (ANOVA) test was performed with the hypothesis:

H_0 : The means of the groups are equal, $\mu_1 = \mu_2 = \mu_3 = \mu_4$.

H_A : The means of the groups are not equal, i.e $\mu_i \neq \mu_j$ for at least one choice of i and j.

Studying the output of the ANOVA test the F-statistic is 3.603 with p-value equal to 0.0748. With significance level 0.05 the null hypothesis, H_0, cannot be rejected i.e. the means for BMI groups are equal.

T2DM patients

Next consider the four BMI groups created for the T2DM patients. Box plots of the insulin sensitivity for each BMI group are shown in Figure 4.10.

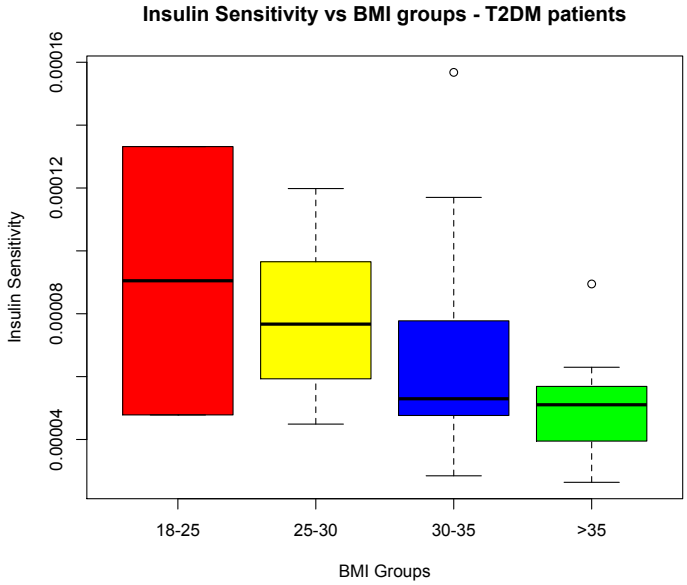


Figure 4.10: Box plots of insulin sensitivity for BMI groups, T2DM patients.

The mean of insulin sensitivity for each BMI group was calculated and is shown in Table 4.5.

	# individuals	**Mean** $[10^{-5}]$
Normal weight: BMI 18-25	2	9.049418
Overweight: BMI 25-30	11	7.857276
Obese: BMI 30-35	20	6.588098
Extremely Obese: BMI >35	13	4.978387

Table 4.5: The means of insulin sensitivity for different BMI groups, T2DM patients.

Analysis of variance (ANOVA) test was performed with the same hypothesis as for healthy subjects. Studying the output of the ANOVA test the F-statistic is 8.757 with p-value equal to 0.0495. With significance level 0.05 the null hypothesis, H_0, can be rejected, i.e. the means are not equal. The ANOVA test is only testing if there is difference between BMI groups as a whole but does not tell which BMI groups differ form one and other. It was obtained that the means differ significantly across the insulin sensitivity with significance level 0.05 but not which pairs are significantly different from each other. To test which pairs differ significantly a pairwise comparison using t-test was conducted.

	BMI 18-25	BMI 25-30	BMI 30-35
BMI 25-30	0.575	-	-
BMI 30-35	0.233	0.224	-
BMI >35	0.057	0.014	0.107

Table 4.6: P-values for pairwise comparison of different BMI groups using t-test, T2DM patients.

As can be seen from Table 4.6 only overweight differs significantly from extremely obese with significance level 0.05. If significance level were to be 0.1 normal weight would also differ significantly from extremely obese.

4.3.2 Gender

For both healthy subjects and T2DM patients it is of interest to see if insulin sensitivity is different dependent on gender.

Healthy Subjects

For the healthy subjects insulin sensitivity was measured for 9 males and 10 females. Box plots of the insulin sensitivity for each gender are shown in Figure 4.11.

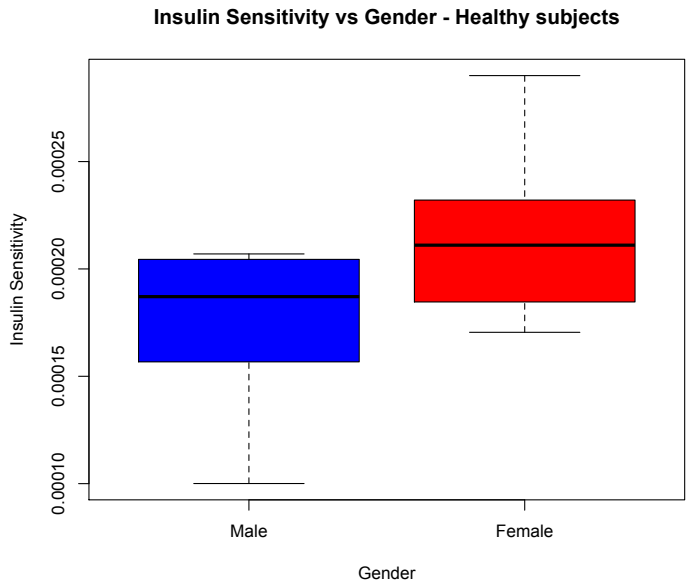


Figure 4.11: Box plots of insulin sensitivity for gender, healthy subjects.

The mean of insulin sensitivity for each gender was calculated and is shown in Table 4.7.

	# **individuals**	**Mean** $[10^{-4}]$
Male	9	1.755752
Female	10	2.164208

Table 4.7: The means of insulin sensitivity for male and female, healthy subjects.

Analysis of variance (ANOVA) test was performed with the hypothesis:

$$H_0 \quad : \quad \text{The means of the groups are equal, } \mu_1 = \mu_2.$$
$$H_A \quad : \quad \text{The means of the groups are not equal, } \mu_1 \neq \mu_2.$$

Studying the output of the ANOVA test, which in fact is the same as t-test in this case, the F-statistic is 5.881 with a p-value equal to 0.0267. With significance level 0.05 the null hypothesis, H_0, can be rejected, i.e. the means for males and females are not equal.

T2DM patients

For the T2DM patients insulin sensitivity was measured for 29 males and 17 females. Box plots of the insulin sensitivity for each gender are shown in Figure 4.12.

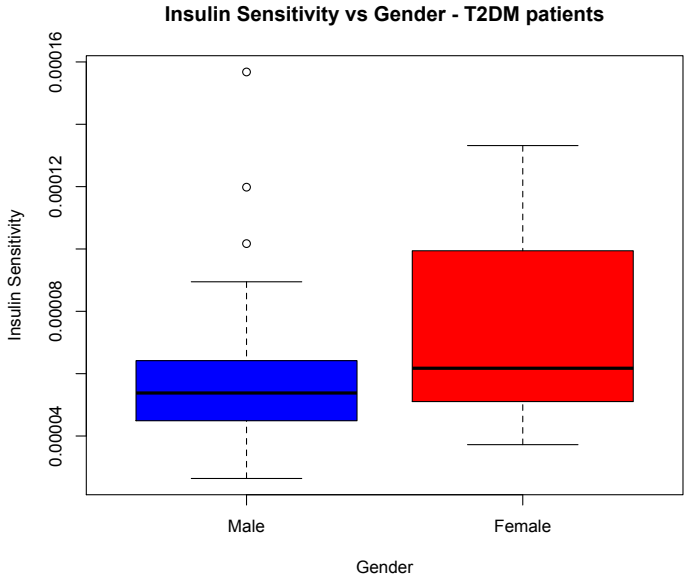


Figure 4.12: Box plots of insulin sensitivity for gender, T2DM patients.

The mean of insulin sensitivity for each gender was calculated and is shown in Table 4.8.

	# **individuals**	**Mean** $[10^{-5}]$
Male	29	6.108149
Female	17	7.286679

Table 4.8: The means of insulin sensitivity for male and female, T2DM patients.

Analysis of variance (ANOVA) test was performed with the hypothesis:

$$H_0 \quad : \quad \text{The means of the groups are equal, } \mu_1 = \mu_2.$$
$$H_A \quad : \quad \text{The means of the groups are not equal, } \mu_1 \neq \mu_2.$$

Studying the output of the ANOVA test the F-statistic is 1.798 with a p-value equal to 0.187. With significance level 0.05 the null hypothesis, H_0, cannot be rejected. There is not significant difference between males and females.

4.3.3 Treatment

For the T2DM patients it is of interest to test if insulin sensitivity is different between treatments. Four groups were created based on different treatments, diet, tablets, tablets+insulin and insulin. Box plots of the insulin sensitivity for each treatment group are shown in Figure 4.13.

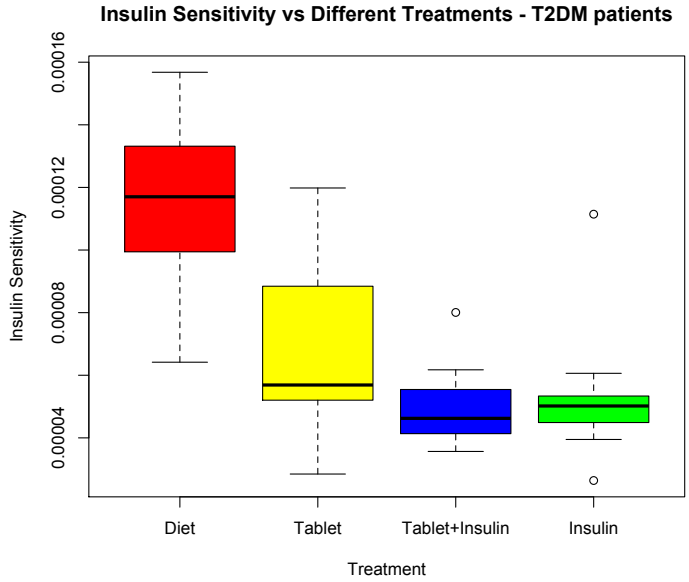


Figure 4.13: Box plot of insulin sensitivity for different treatment, T2DM patients.

The mean of insulin sensitivity for each treatment group was calculated and is shown in Table 4.9.

	# **individuals**	**Mean** $[10^{-5}]$
Diet	5	11.41128
Tablets	21	6.678019
Tablets+Insulin	11	5.00363
Insulin	9	5.408348

Table 4.9: The means of insulin sensitivity for different treatment, T2DM patients.

Analysis of variance (ANOVA) test was performed with the hypothesis:

H_0 : The means of the groups are equal, $\mu_1 = \mu_2 = \mu_3 = \mu_4$.

H_A : The means of the groups are not equal, i.e $\mu_i \neq \mu_j$ for at least one choice of i and j.

Studying the output of the ANOVA test the F-statistic is 6.336 with a p-value equal to 0.0155. With significance level 0.05 the null hypothesis, H_0, can be rejected i.e the means for different treatment are not equal. The ANOVA test is only testing if there are differences between treatments as a whole but does not tell which treatment differ from one other. It was obtained that the means differ significantly across the insulin sensitivity but not which pairs are significantly different from each other. To test which pairs differ significantly a pairwise comparison using t-test was conducted.

	Diet	**Tablets**	**Tablets+Insulin**
Tablets	0.00017	-	-
Tablets+Insulin	$6.3 \cdot 10^{-6}$	0.05742	-
Insulin	$3.0 \cdot 10^{-5}$	0.17368	0.69774

Table 4.10: P-values for pairwise comparison of different treatments using t-test, T2DM patients.

As can be seen from Table 4.10 the Diet group differs significantly from every other group. With significance level 0.05 pairs of other groups do not differ significantly but if the significance level were to be 0.1 Tablets and Tablets+Insulin would differ significantly. That means that the null hypothesis, H_0, can be rejected for the Diet group paired with any other group, i.e. the mean of the Diet group is not the same as in any other group but for the means of the other groups compared to one and other the null hypothesis, H_0, cannot be rejected.

4.3.4 Disease Duration

For the T2DM patients it is of interest to see if insulin sensitivity is different dependent on disease duration. Three groups were created for disease duration, less than 5 years, 5-8 years and 8-10 years. Box plots of the insulin sensitivity for each disease duration group are shown in Figure 4.14.

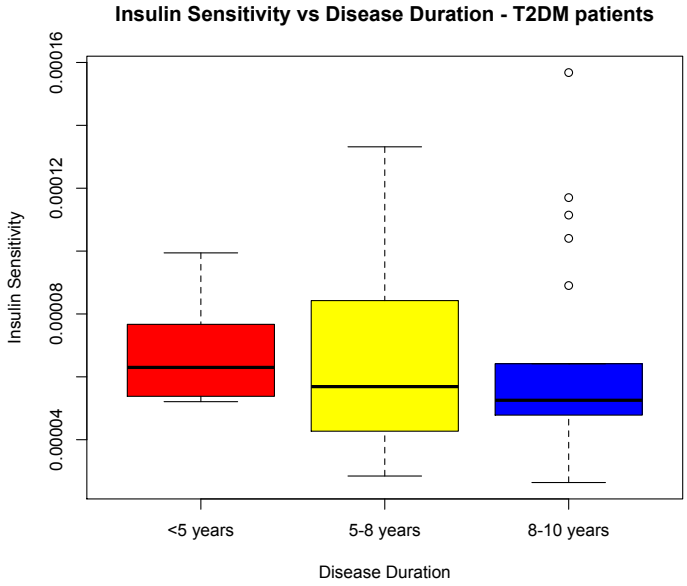


Figure 4.14: Box plots of insulin sensitivity for disease duration, T2DM patients.

The mean of insulin sensitivity for each disease duration group was calculated and is shown in Table 4.11.

	# individuals	**Mean** $[10^{-5}]$
<5 years	6	6.800414
5-8 years	19	6.562657
8-10 years	21	6.453186

Table 4.11: The means of insulin sensitivity for disease duration, T2DM patients.

Analysis of variance (ANOVA) test was performed with the hypothesis:

H_0 : The means of the groups are equal, $\mu_1 = \mu_2 = \mu_3$.

H_A : The means of the groups are not equal, i.e $\mu_i \neq \mu_j$ for at least one choice of i and j.

Studying the output of the ANOVA test the F-statistic is 0.033 with a p-value equal to 0.968. With significance level 0.05 the null hypothesis, H_0, cannot be rejected.

Chapter 5

A Linear Regression Model

A Linear regression is an approach to model the relationship between a dependent variable and a vector of regressors. In this case the dependent variable is the insulin sensitivity, S_I, and the regressors are some covariates. The relationship between the dependent variable, S_I, and the regressors is assumed to be linear and the data is modeled using linear prediction functions and unknown model parameters are estimated. The model also includes an error variable, ϵ_i, which is an unobserved random variable that adds noise to the linear relationship between the dependent variable and the regressors.

5.1 A Linear Regression Model

A linear regression model of insulin sensitivity was created with all appropriate variables for both healthy subjects and T2DM patients. The model included all variables that are known with the aim to make some conclusions about insulin sensitivity from the known variables.

5.1.1 Healthy Subjects

A linear regression model for insulin sensitivity was created for healthy subjects. The known predictor variables for healthy subjects are BMI and gender. The linear regression model is defined as

$$S_I = \beta_0 + \beta_1(\text{BMI}_i) + \beta_2(\text{Gender}_i) + \epsilon_i \tag{5.1}$$

where ϵ_i is the error, $\epsilon_i \sim N(0,s^2)$.

A t-test was performed to see if the parameters are significant.

Coefficient	Estimate	Std. Error	t-value	p-value
BMI	$-2.919 \cdot 10^{-6}$	$1.824 \cdot 10^{-6}$	-1.600	0.12910
Gender	$3.370 \cdot 10^{-5}$	$1.673 \cdot 10^{-5}$	2.014	0.06110

Table 5.1: T-test for linear regression model, healthy subjects.

Studying the output of the t-test, the p-value for the BMI coefficient, β_1, is 0.12910 and for the Gender coefficient, β_2, the p-value is 0.06110. With significance level 0.05 neither coefficients is significant but if significance level is 0.1 the Gender coefficient is significant. R^2_{adjusted} was also calculated for the model and was 0.2795. That means that the linear regression model only explains 27.95% of the data and the rest is covered by the error. Analysis of variance (ANOVA) test was performed to see if the variables were significant. The hypothesis are

$$\begin{aligned} H_0 &: \beta_0 = \beta_1 = \beta_2 = 0. \\ H_A &: \beta_0 \neq \beta_1 \neq \beta_2 \neq 0. \end{aligned}$$

The F-statistic is 4.491 with p-value 0.02832. For significance level 0.05 the null hypothesis, H_0, can be rejected i.e. the variable values are not the same. After the model is fitted residuals diagnosis was done and the following was investigated: if errors are uncorrelated, outliers, constant error variance, symmetric errors and if the linear model is adequate.

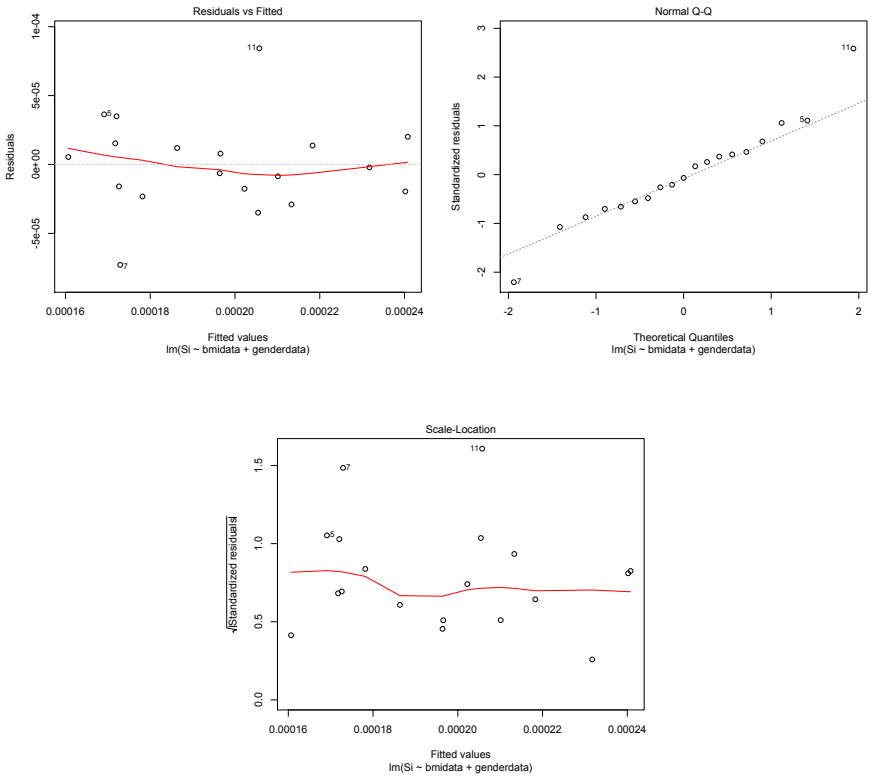


Figure 5.1: Residual diagnosis of the linear regression model for healthy subjects.

Figure 5.1 (a) and (c) show that the residuals tend to have almost constant variance and although there is some variation, there seems to be no correlation between residuals and the fitted values and the residuals do not form a pattern. Figure 5.1 (b) shows that the residuals seem to follow a normal distribution except for two outliers. The parameter estimation is not conclusive as was expected, the insulin sensitivity estimate is mostly covered by the error term. More information about each individual is needed to estimate the insulin sensitivity.

5.1.2 T2DM Patients

A Linear regression model of insulin sensitivity was created for T2DM patients. The known predictor variables for T2DM patients are BMI, disease duration, different treatments and gender. The linear regression model is defined as

$$S_I = \beta_0 + \beta_1(\text{BMI}_i) + \beta_2(\text{Duration}_i) + \beta_3(\text{Treatment}_i) + \beta_4(\text{Gender}_i) + \epsilon_i \qquad (5.2)$$

where ϵ_i is the error.

A t-test was performed to see if the model is significant.

Coefficient	Estimate	Std. Error	t-value	p-value
BMI	$-2.029 \cdot 10^{-6}$	$7.364 \cdot 10^{-7}$	-2.756	0.00869
Duration	$1.091 \cdot 10^{-7}$	$1.699 \cdot 10^{-6}$	0.064	0.94911
Treatment	$-1.074 \cdot 10^{-5}$	$4.464 \cdot 10^{-6}$	-2.405	0.02075
Gender	$2.366 \cdot 10^{-6}$	$8.386 \cdot 10^{-6}$	0.282	0.77928

Table 5.2: T-test for linear regression model, T2DM patients.

Studying the output of the t-test, the p-value for the BMI coefficients, β_1, is 0.00869, Duration coefficient, β_2, 0.94911, Treatment coefficient, β_3, 0.02075 and Gender coefficient, β_4, 0.77928. With significance level 0.05 BMI coefficient and Treatment coefficient are significant but not Duration coefficient and Gender coefficient. R^2_{adjusted} was also calculated for the model and was 0.2005. That means that the linear regression model only explains 20.05% of the data and the rest is covered by the error. ANOVA test was performed to see if the parameters were significant. The hypothesis are the same as for the healthy subjects. The F-statistic was 2.985 with p-value 0.0221. For significance level 0.05 the null hypothesis, H_0, can be rejected, i.e the parameter values are not the same. After the model is fitted residuals diagnosis was performed and analyzed as for the healthy subjects.

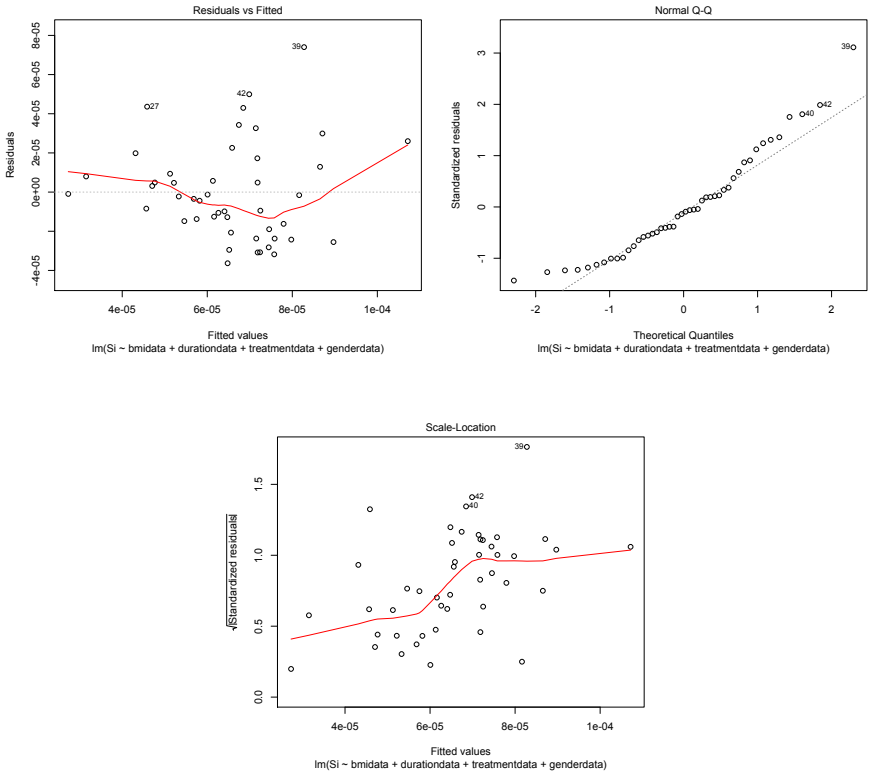


Figure 5.2: Residual diagnosis of the linear regression model for T2DM patients.

By looking at Figure 5.2 (b) the residuals are approximately normally distributed. Figure 5.2 (a) and (c) show that the residuals to not tend to have constant variance, there is a big variation between the insulin sensitivity for each individual. There seems to be no correlation between the residuals and the fitted values. The linear regression model is not a good fit so an improvement was done by only including the significant parameters in order to try to get a conclusive model.

5.2 An Improved Linear Regression Model

An improved version of the linear regression model for T2DM patients is proposed. Since the parameters β_2 and β_4 were not significant the model was redefined as

$$S_I = \beta_0 + \beta_1(\text{BMI}_i) + \beta_2(\text{Treatment}_i) + \epsilon_i \tag{5.3}$$

T-test was performed to see if the model is significant.

Coefficient	**Estimate**	**Std. Error**	**t-value**	**p-value**
BMI	$-2.066 \cdot 10^{-6}$	$7.088 \cdot 10^{-7}$	-2.915	0.00563
Treatment	$-1.107 \cdot 10^{-5}$	$4.180 \cdot 10^{-6}$	-2.648	0.01126

Table 5.3: T-test for improved linear regression model, T2DM patients.

Studying the output of the t-test the p-value for the BMI coefficient, β_1, is 0.00563 and the Treatment coefficient, β_2, is 0.01126. With significance level 0.05 both of the coefficient are significant. R^2_{adjusted} was also calculated for the model and was 0.2361. That means that the linear regression model explains 23.61% of the data and the rest is covered by the error. That is a better result than for the first linear regression model so this model explains the insulin sensitivity more accurate. ANOVA test was performed to see if the parameters were significant. The hypothesis are the same as for the healthy subjects. The F-statistic was 7.955 with p-value 0.001149. For significance level 0.05 the null hypothesis, H_0, can be rejected, i.e the parameter values are not the same. After the model is fitted residuals diagnosis was done and investigated as before.

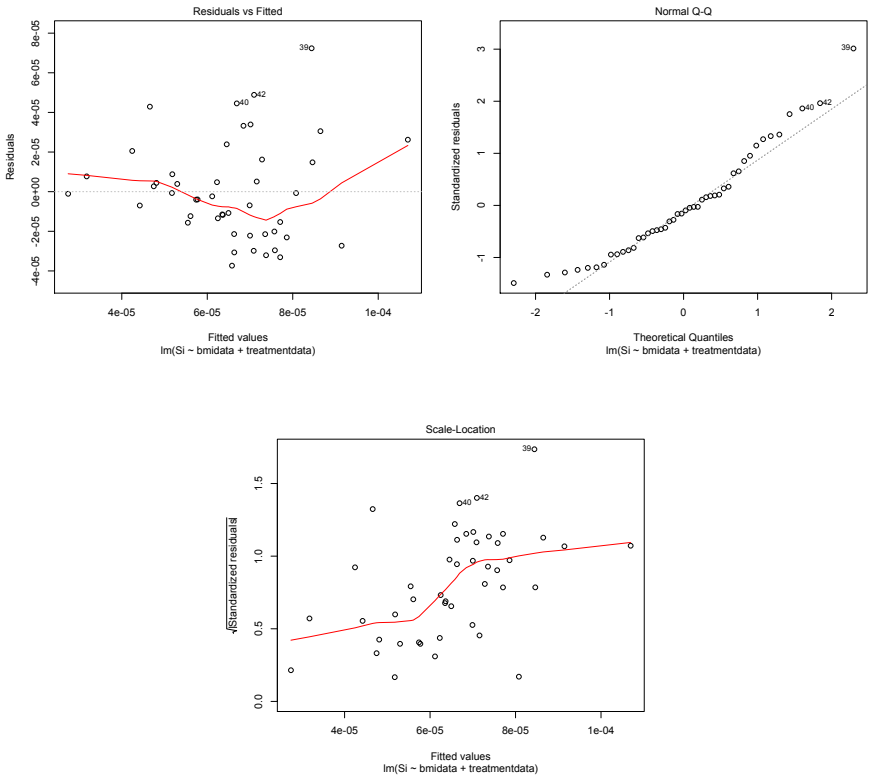


Figure 5.3: Residual diagnosis of the improved linear regression model for T2DM patients.

Figure 5.3 (a) and (c) imply that the residuals for the improved linear regression model have more constant variance than before but not completely constant. The residuals do not seem to be correlated to the fitted values and forms no pattern which makes the model more adequate then original linear regression model. From Figure 5.3 (b) it can be seen that the residuals do not strictly follow the normal distribution but are closer to follow it than before. The linear regression model represents insulin sensitivity better than before but the fit is not good enough for the model to be conclusive. More information about each individual is needed to estimate the insulin sensitivity.

Chapter 6

Covariates in the Minimal Model

6.1 Covariates in Nonlinear Mixed Effect Models

As described in Chapter 3, nonlinear mixed effect (NLME) models are useful when there is both variability within individuals and between individuals. In addition, NLME models can be very useful in describing nonlinear relationship between a response variable and parameters and are covariates in the data grouped according to a classification factor. Some of the parameters are allowed to vary with groups through the random effects and covariates are common to all groups. By associating common random effects to observations in the same group, NLME flexibly represents the covariance structure induced by the grouping of the data. As described before the random effects account for individual deviation in the parameters among groups but these deviations can be explained in some way by difference in covariate values among groups [9].

Including covariates in the NLME model to explain the group variation often leads to simplification in the random effects and allows for better understanding of the mechanism producing the response.

A nonlinear mixed effect model is defined as

$$y_{ij} = f(\phi_{ij}, x_{ij}) + \epsilon_{ij} \tag{6.1}$$

as in chapter 3. The population model for the individuals is now defined as

$$\phi_{ij} = A_{ij}\beta + B_{ij}b_i \tag{6.2}$$

for the j'th observation from the minimal model and the i'th individual where the covariates are included through the A_{ij} matrix. The most promising covariates are included in the model by adding the corresponding columns to A_{ij}, with resulting estimated fixed effects being tested for significance. Thus,

A_{ij} : matrix to determine which parameters have fixed effects.

β : vector of fixed population parameters.

B_{ij} : matrix to determine which parameters have random effects.

b_i : $b_i \sim N(0,\sigma^2\Omega)$, $\sigma^2\Omega$ is the covariance matrix of random effects.

To estimate the parameters the NLME models are used as before. The description of the NLME models are in chapter 4. The maximum likelihood function with respect to the marginal distribution of y is defined as in equation 3.7 and the log likelihood function is optimized.

6.2 Modeling Data with Covariates

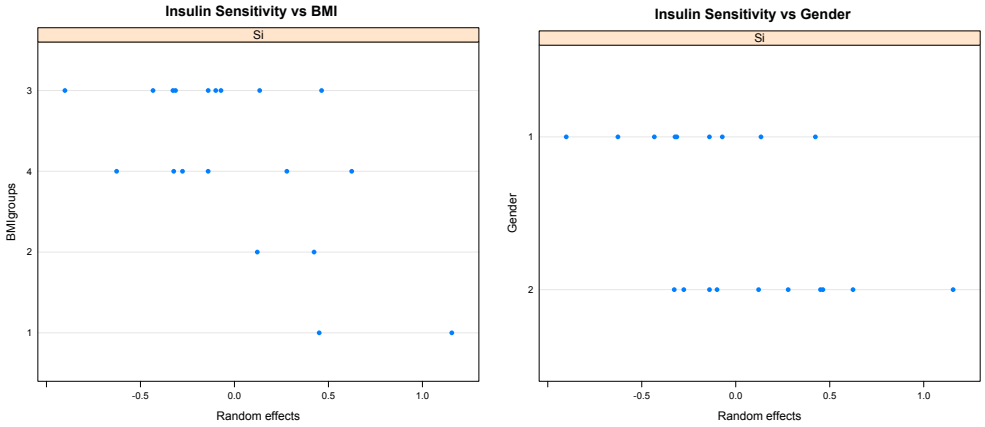
Including covariates in the minimal model can hopefully give a better estimate of the insulin sensitivity parameter. The same dataset is used and the parameters were estimated both for healthy subjects and T2DM patients. Initial values for both healthy subjects and T2DM patients are shown in Table 4.1. The model is first solved with no covariates to evaluate which covariates should be incorporate. Plots of the estimated random effects versus candidate covariates are used to identify interesting patterns and which covariates should be included in the model. Since the random effects accommodate individual variation from the population mean, plotting the estimated random effects against the candidate covariates provides useful information. A systematic pattern in given random effects with respect to candidate covariate would indicate that the covariate should be included in the model [9].

The only random effect parameter of interest is insulin sensitivity, S_I, and thus is only considered which candidate covariates should be included in the model with respect to insulin sensitivity.

6.2.1 Healthy Subjects

For healthy subjects the candidate covariates are BMI and gender. To explain the plots, the numbers on the vertical axis represent:

BMI	**Gender**
1 : Normal weight, 18-25	1 : Male
2 : Overweight, 25-30	2 : Female
3 : Obese, 30-35	
4 : Extremely obese >35	



(a) Plot of insulin sensitivity vs BMI. **(b)** Plot of insulin sensitivity vs Gender.

Figure 6.1: Plots of insulin sensitivity, S_I, against candidate covariates for healthy subjects, BMI and gender.

Both Figure 6.1 (a) and Figure 6.1 (b) show a systematic pattern. The relationship between insulin sensitivity and BMI is such that, the higher the BMI the lower is the insulin sensitivity is. Between insulin sensitivity and gender the insulin sensitivity tends to be lower for males but higher for females. From the candidate covariates both BMI and gender will be included in the model. It is interesting to note that both when the data was grouped and a linear regression model was created gender and BMI were significant parameters with significance level 0.1 for healthy subject.

The model was updated and both candidate covariates included as a covariates in the minimal model for healthy subjects. Same initial values were used and initial value for BMI and gender was set to zero. By including covariates the standard deviation for insulin sensitivity went from 0.2254188 to 0.1395728, indicating that variation in insulin sensitivity for healthy subjects can be explained in some way by BMI and gender.

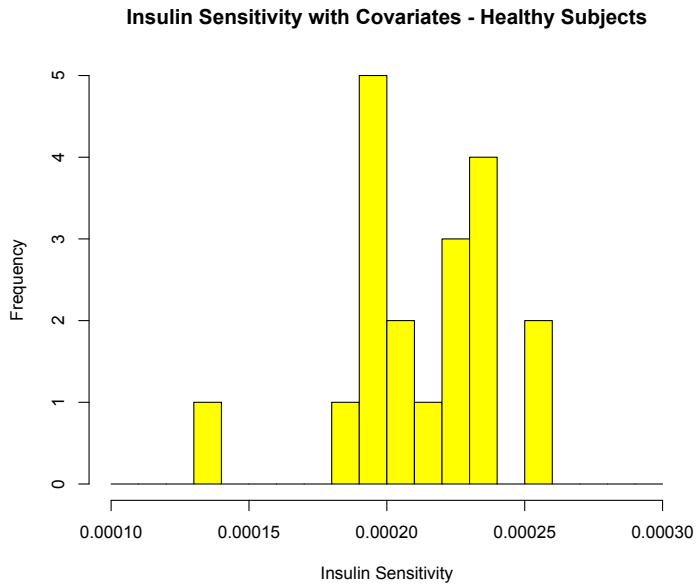
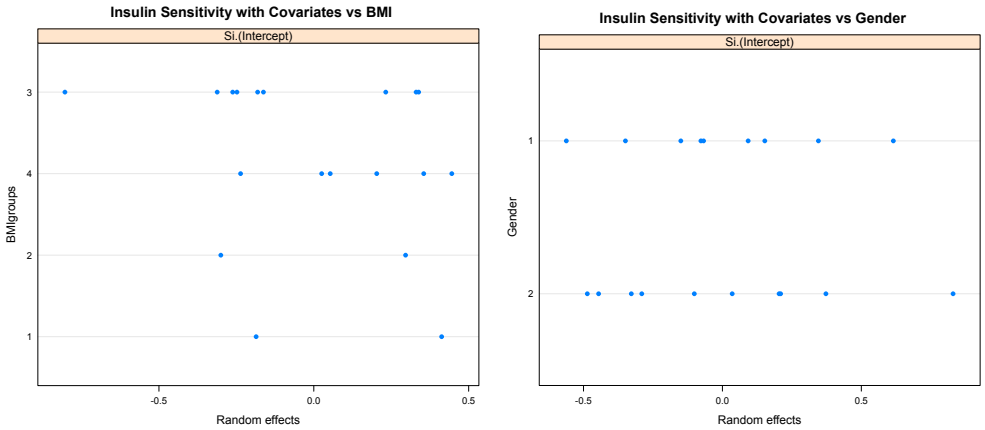


Figure 6.2: Histogram of insulin sensitivity after including candidate covariates, healthy subjects.

From Figure 6.2 it can be seen that the variation of insulin sensitivity values is less than in Figure 4.2. Since the standard deviation is smaller when including covariates the minimal model gives better estimate of insulin sensitivity when covariates are included.



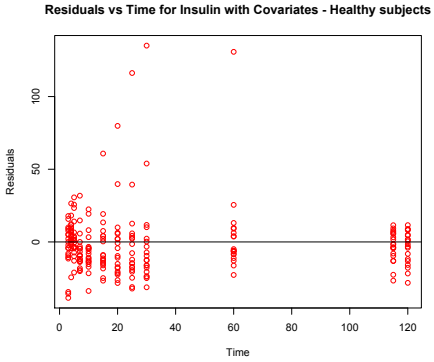
(a) Plot of insulin sensitivity vs BMI after including candidate covariates.

(b) Plot of insulin sensitivity vs Gender after including candidate covariates.

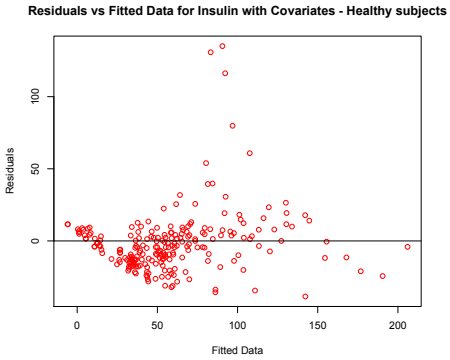
Figure 6.3: Plot of insulin sensitivity against covariates that were included for healthy subjects, BMI and gender.

After covariates were included in the model, less systematic pattern occurs for insulin sensitivity versus both BMI and gender as expected. The ratio between a random effects standard deviation and an absolute value of the corresponding fixed effect gives an idea of the relative group variability for the coefficient and is useful in testing which random

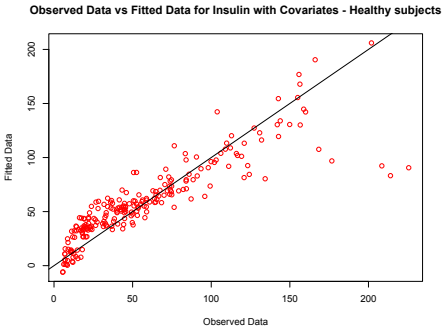
effects should be tested for deletion from the model. Insulin sensitivity is only tested in this case since that is the only parameter of interest. For the minimal model with BMI and gender as covariates this ratio is 0.01% which is too small and will result in highly significant p-value, indicating that the insulin sensitivity parameter needs to be included as a random effect in the minimal model although BMI and gender are included as covariates.



(a) Residuals vs Time for Insulin with Covariates.

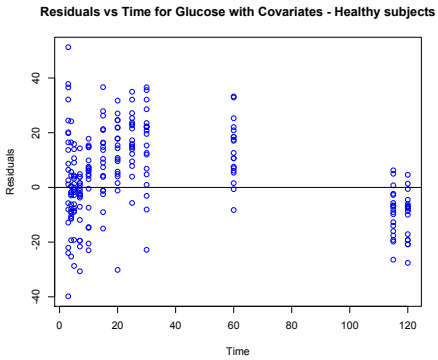


(b) Residuals vs Fitted Data for Insulin with Covariates.

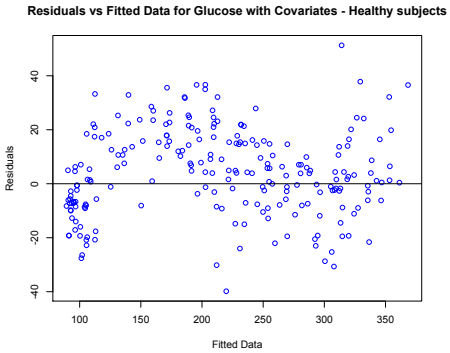


(c) Observed Data vs Fitted Data for Insulin with Covariates.

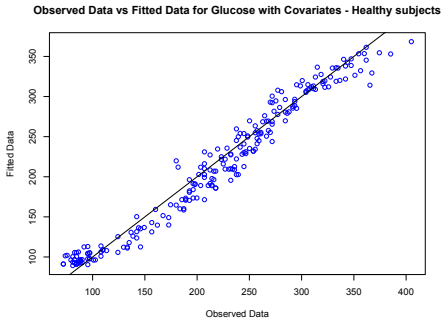
Figure 6.4: Plots for the Insulin concentration for healthy subjects with candidate covariates included.



(a) Residuals vs Time for Glucose with Covariates.



(b) Residuals vs Fitted Data for Glucose with Covariates.



(c) Observed Data vs Fitted Data for Glucose with Covariates.

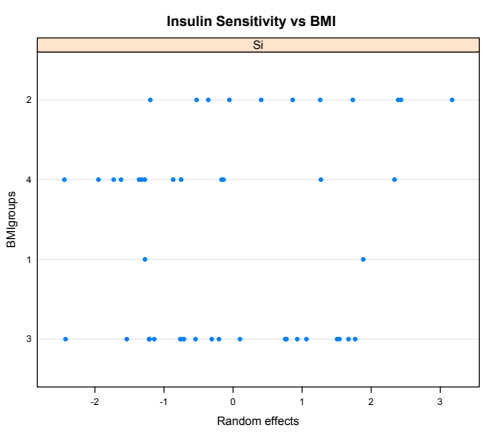
Figure 6.5: Plots for the Glucose concentration for healthy subjects with candidate covariates included.

Comparing Figure 6.4 with Figure 4.3 and Figure 6.5 with Figure 4.4 it is evident that both the residual plots and the observed data versus fitted data plot do not change much by including covariates. That implies that by including covariates does not fix the curved pattern in the residuals. The curved pattern occurs because the model is explained by ordinary differential equations so to decrease the curve pattern further improvements have do be made.

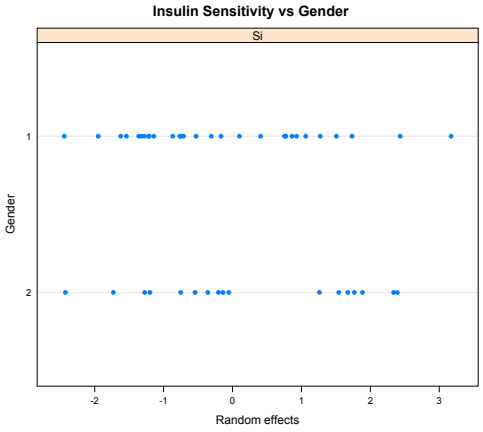
6.2.2 T2DM Patients

For T2DM patients the candidate covariates are BMI, gender, different treatment and disease duration. To explain the plots, the numbers on the vertical axis represent:

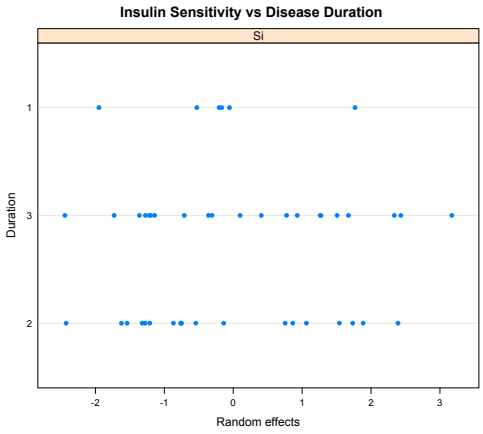
BMI	**Treatments**	**Disease Duration**	**Gender**
1 : Normal weight, 18-25	0 : Diet	1 : <5 years	1 : Male
2 : Overweight, 25-30	1 : Tablets	2 : 5-8 years	2 : Female
3 : Obese, 30-35	2 : Tablets+Insulin	3 : 8-10 years	
4 : Extremely obese >35	3 : Insulin		



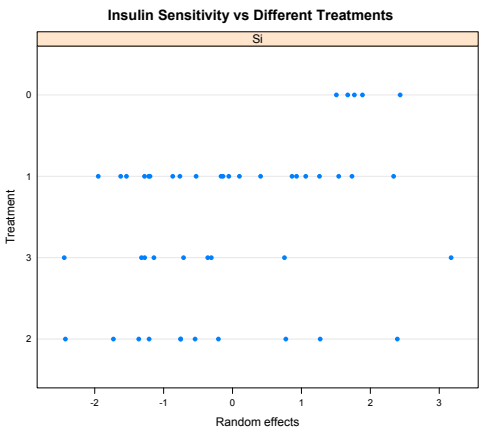
(a) Plot of insulin sensitivity vs BMI.



(b) Plot of insulin sensitivity vs Gender.



(c) Plot of insulin sensitivity vs Disease Duration.



(d) Plot of insulin sensitivity vs Different Treatments.

Figure 6.6: Plots of insulin sensitivity, S_I, against candidate covariates for T2DM patients, BMI, gender, disease duration and different treatments.

Figure 6.8 (a) shows a slight systematic pattern. It seems like if the higher BMI is, the lower insulin sensitivity is. Figure 6.8 (b) and Figure 6.8 (c) show no systematic pattern, which indicates that neither gender nor disease duration should be included in the model. Figure 6.8 (d) shows a systemic pattern, T2DM patients on diet tend to have higher insulin sensitivity than patients on other treatments. From candidate covariates both BMI and different treatments are included in the model. It is interesting to note again that both when the data was grouped and a linear regression model was created BMI and different treatment were significant but not gender and disease duration for T2DM patients.

The model was updated and BMI and treatments were included in the minimal model for T2DM patients. Same initial values were used as before and initial value for both BMI and different treatment was set to zero. By including BMI and different treatments as covariates the standard deviation for insulin sensitivity went from 2.403692 to 1.215193, indicating that variation in insulin sensitivity for T2DM patients can be explain in some way by BMI and different treatments.

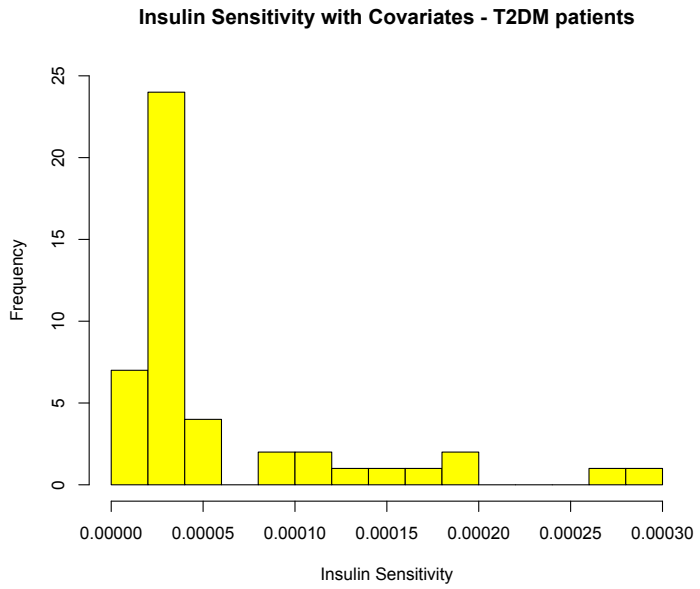
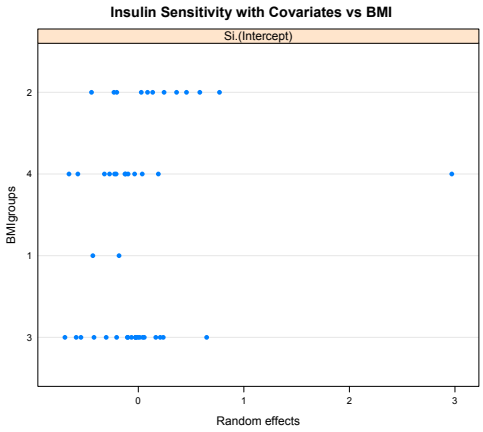
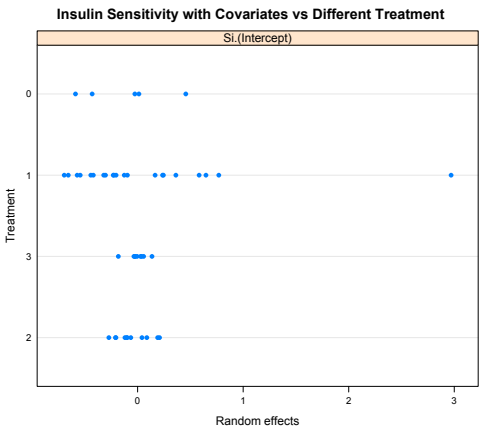


Figure 6.7: Histogram of insulin sensitivity after including candidate covariates, T2DM patients.

From Figure 6.7 it is evident that the variation of insulin sensitivity values is less than in Figure 4.6. Since the standard deviation is smaller when including covariates the minimal model gives better estimate of insulin sensitivity when covariates are included. There are two individuals that tend to have much higher insulin sensitivity than other individuals which shows how divergent the T2DM patients are.



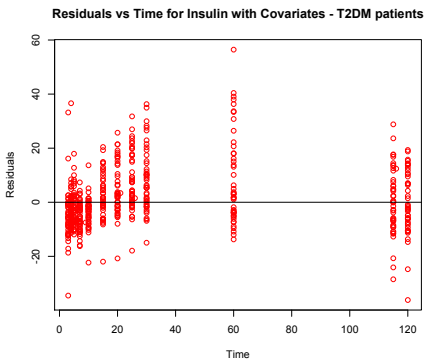
(a) Plot of insulin sensitivity vs BMI after including covariates.



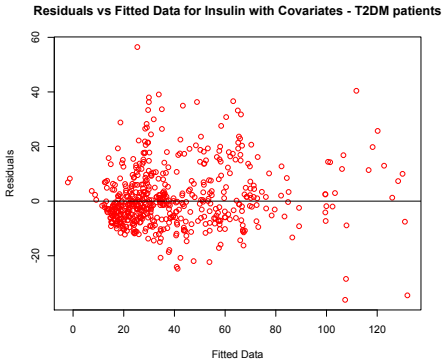
(b) Plot of insulin sensitivity vs Different Treatments after including covariates.

Figure 6.8: Plots of insulin sensitivity, S_I, against covariates that were included for T2DM patients, BMI and different treatments.

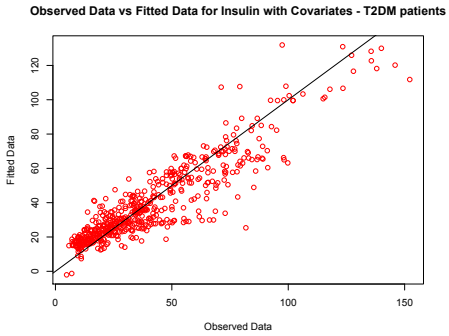
After covariates were included in the model the systematic pattern is much less than before for insulin sensitivity versus both BMI and different treatments. The ratio between a random effects standard deviation and an absolute value of the corresponding fixed effect was used to see if deletion of the insulin sensitivity parameter was plausible. The ratio was 0.2% which will also result in highly significant p-value. The insulin sensitivity parameter needs to be included as a random effect in the minimal model although covariates are added.



(a) Residuals vs Time for Insulin with Covariates.

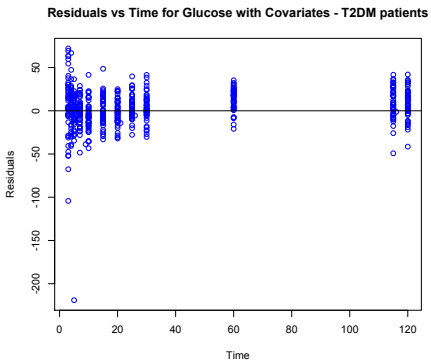


(b) Residuals vs Fitted Data for Insulin with Covariates.

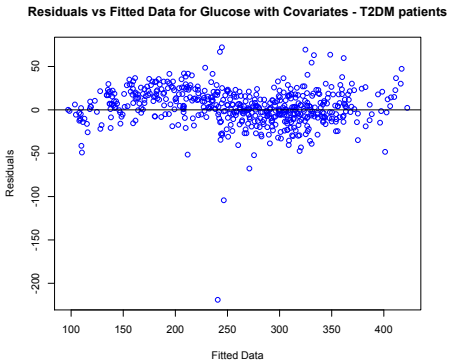


(c) Observed Data vs Fitted Data for Insulin with Covariates.

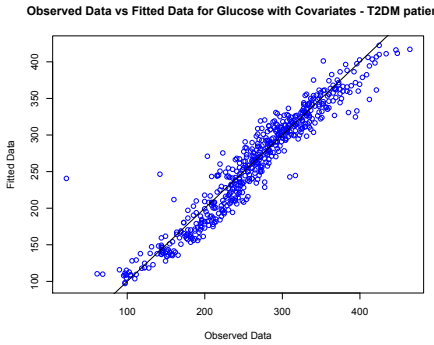
Figure 6.9: Plots for the Insulin concentration for T2DM patients with candidate covariates included.



(a) Residuals vs Time for Glucose with Covariates.



(b) Residuals vs Fitted Data for Glucose with Covariates.



(c) Observed Data vs Fitted Data for Glucose with Covariates.

Figure 6.10: Plots for the Glucose concentration for T2DM patients with candidate covariates included.

As before the by comparing Figure 6.9 with Figure 4.7 and Figure 6.10 with Figure 4.8 it is evident that both the residual plots and the observed versus fitted data plot do not change much by including covariates. That implies again that by including covariates does not fix the curved pattern in the residuals and further improvements have to be implemented to fix the curved pattern.

Chapter 7

Stochastic Minimal Model

7.1 Stochastic Modeling in Nonlinear Mixed Effect Models

When using nonlinear mixed effects model in PK/PD the general approach is to use ordinary differential equations (ODEs). Extending to stochastic differential equations (SDEs) is achieved by adding an additional Wiener noise component. This additional noise allows handling of autocorrelated residuals originating from natural variation or systematic model error. Autocorrelated residuals are often partly ignored in PK/PD modeling although it violates the hypothesis for many standard statistical tests [10]. Since a curved pattern occurs in the residual plots for the deterministic minimal model a stochastic approach could decrease this curve pattern and give more accurate results.

Nonlinear mixed effect models are described in details in chapter 3. As before a nonlinear mixed effect model is defined to describe data with the structure

$$y_{ij}, \quad i = 1, \ldots, M, \quad j = 1,2$$

where y_{ij} is the j'th observation (glucose or insulin concentration) for the i'th individual at time t_{ij} and M is the number of individuals. In a mixed effect model the variation is split into variation between individuals and within individual, which is modeled by a mixed effect model and a population model respectively.

The add on package **PSM** in R is used in order to estimate parameters but the package provides functions for estimation of linear and nonlinear mixed effect models using SDEs. The package allows for any multivariate nonlinear time-variant model to be specified, handles multidimensional input, covariates and missing observations [11].

7.2 Stochastic Minimal Model using NLME

The mixed effect model can be written in the form of a state space model that consists of two parts. The continuous state equation defining the dynamics of the system, in this case the minimal model, and a set of discrete measurement equations, which defines a functional relationship between the states of the system and the obtained measurements. The nonlinear form of the ODE state space model equations are defined as

$$dx_t = f(x_t,u_t,t,\phi_i)dt \tag{7.1}$$

$$y_{ij} = g(x_{ij},u_{ij},t_{ij},\phi_i) + e_{ij} \tag{7.2}$$

but the nonlinear form of the SDE state space model equations are defined as

$$dx_t = f(x_t,u_t,t,\phi_i)dt + \sigma(u_t,t,\phi_i)d\omega_t \tag{7.3}$$

$$y_{ij} = g(x_{ij},u_{ij},t_{ij},\phi_i) + e_{ij} \tag{7.4}$$

where

t_{ij} : continuous time variable.
x_t : states of the model at time t.
u_t : optional inputs at time t.
σ : diffusion term.
ω_t : standard Wiener process such that $\omega_{t_2} - \omega_{t_1} \in N(0,I|t_2 - t_1|)$.
ϕ_i : individual model parameter.
e_{ij} : Gaussian white noise measurement error, $e_{ij} \in N(0,S(\phi_i))$.

The difference between equation 7.1 and equation 7.3 is that the diffusion part $\sigma(u_t,t,\phi_i)$ has been added. Equation 7.3 represents the dynamic system, the minimal model, and equation 7.4 represents the nonlinear mixed effect model [11].

The population model for the individuals is defined as before

$$\phi_i = \beta + Bb_i \tag{7.5}$$

The state concept is essential to the understanding of the model setup. In this case the state vector represents the glucose concentration, the remote insulin concentration and the insulin concentration or $x_t = [G(t), X(t), I(t)]$, and is only observable through measurement noise. Equation 7.4 defines the actual relation between measurement and states.

The minimal model is redefined as a set of stochastic differential equations by the

following nonlinear coupled system of equations

$$dG(t) = (-(S_G + X(t))G(t) + S_G G_b)dt + \sigma_1 d\omega_1 \tag{7.6}$$

$$dX(t) = (-p_2(X(t) + S_I(I(t) - I_b)))dt \tag{7.7}$$

$$dI(t) = (-n(I(t) - I_b) + \gamma(G(t) - h)t)dt + \sigma_2 d\omega_2 \tag{7.8}$$

where σ_1 and σ_2 are diffusion constants and ω_1 and ω_2 standard independent Wiener processes. Diffusion constants are only added to the glucose and insulin concentration since the measurements are based on those variables. To estimate the parameters good initial values are needed as before. Same initial values were used as for the ODE minimal model and are shown in Table 4.1. The initial values for the diffusion constants was calculated as the mean of the residuals from the deterministic minimal model since the goal of using SDE version is to decrease the curved pattern that forms in the residuals as can be seen in Figure 4.4 (b) for example.

The model is nonlinear and it is more complicated to set a nonlinear model up for the **PSM** package than a linear model. Unfortunately, it was really time consuming and some computational difficulties occurred during the process. The optimization procedure proposed was unable to converge. Instead a simple linear model was set up to show that the stochastic approach should give a better estimate of the insulin sensitivity.

7.2.1 Simple Example

The simplest compartment model is the one that only includes one compartment. The same dataset as in the previous chapters was used to solve the model but only glucose concentration was included and only healthy subjects. The model assumes that the body constantly produces glucose that is distributed into the blood and at the same time the body is absorbing an amount of blood glucose that is proportional to the concentration in the compartment $G(t)$.

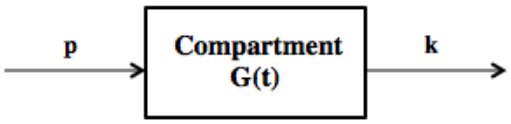


Figure 7.1: The simple compartment model included a dose input, p, one compartment, $G(t)$, and disposal rate output, k.

The model can be described by one ordinary differential equation

$$dG(t) = (-kG(t) + p)dt, \quad G(0) = G_0 \tag{7.9}$$

where

$G(t)$: Glucose concentration at time t [mM].

k : a constant, when multiplied with $G(t)$, gives rate of glucose disposal at time t.

p : constant rate of glucose production in the body.

To estimate the parameters k and p, mixed effect model will be used. The mixed effect model for this compartment model can be defined as

$$y_i(t_{ij}) = G(t_{ij}) + \epsilon_{ij} \tag{7.10}$$

The initial estimates used for the model can be seen in Table 7.1. These initial estimates were published as good initial estimates for this model [4].

Parameter	k	p	G_0
Initial Value	0.0294	0.136	17.4

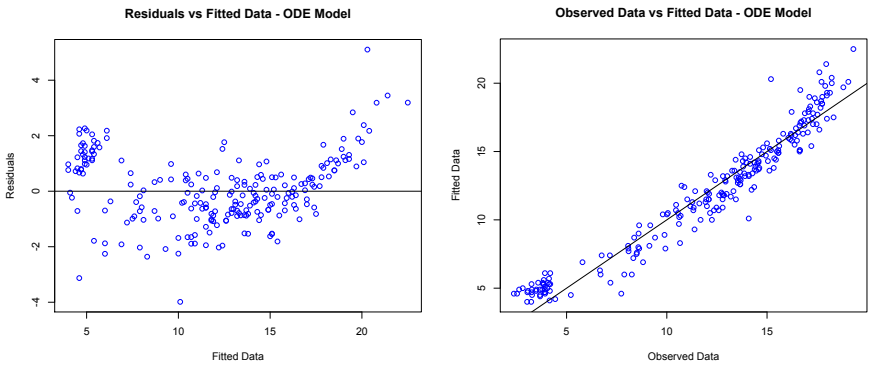
Table 7.1: Initial values for the simple example.

The parameter values for the population estimates, means and standard deviations, are shown in Table 7.2.

	k $\sigma(k)$	p $\sigma(p)$	G_0 $\sigma(G_0)$
Mean	0.013136	0.13598	17.3984
SD	0.00201	0.017559	0.85829

Table 7.2: Parameter values for the population estimates, means and standard deviations for the simple example.

It is of interest is to see how the residuals are and if the curved pattern appears on the residual and the observed data versus fitted data plots.



(a) Residuals vs Fitted Data for Glucose.

(b) Observed Data vs Fitted Data for Glucose.

Figure 7.2: Plots for the Glucose concentration for the ODE model.

From Figure 7.2 the residuals form a curved pattern for the same reason described in chapter 4, the ODE model makes the best line trough the observed points and is thus always above or below the observed point. A SDE version of the model allows for variation of this line, the Wiener process adds an error to the line which makes the line vary. Anderson-Darling test was performed to see if the residuals follow a standard normal distribution. The hypothesis were the same as for the deterministic minimal model. The p-value was 0.2039 with the test statistic 0.5024. For significance level 0.05 the null hypothesis, H_0, is rejected and the residuals do not follow the standard normal distribution.

When expanding the model to stochastic differential equation a diffusion term is added. The model can be defined as

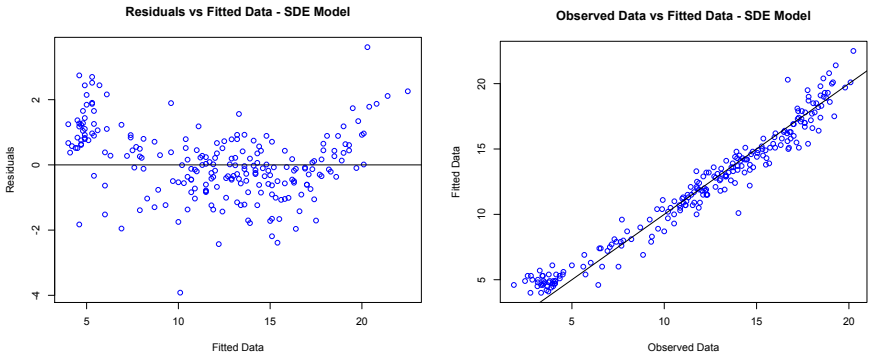
$$dG(t) = (-kG(t) + p)dt + \sigma d\omega, \quad G(0) = G_0 \tag{7.11}$$

where σ is the diffusion constant and ω a standard Wiener process. The same initial values were used as for the ODE model and are shown in Table 7.1 and initial value for the diffusion constant was calculated as the mean of the residuals from the deterministic minimal model. The parameter values for the population estimates, means and standard deviations, are shown in Table 7.3.

	k $\sigma(k)$	p $\sigma(p)$	G_0 $\sigma(G_0)$
Mean	0.0147267	0.13567	17.3984
SD	0.00132	0.011539	0.85829

Table 7.3: Parameter values for the population estimates, means and standard deviations for the simple example.

The standard deviations for k and p have decreased when the stochastic version is applied. The implies that an improved parameter estimations is obtained. The residual were plotted again to see if the curve pattern had decreased with the SDE version.



(a) Residuals vs Fitted Data for Glucose.

(b) Observed Data vs Fitted Data for Glucose.

Figure 7.3: Plots for the Glucose concentration for the SDE model.

By looking at Figure 7.3 the curved pattern has decreased which implies that the stochastic version of the model gives better results. Anderson-Darling test was performed again to see if the residuals follow a standard normal distribution. The hypothesis were the same as before. The p-value was 0.0981 with the test statistic 0.6075. For all significance level 0.05 the null hypothesis, H_0, is rejected and the residuals do not follow a standard normal distribution but for significance level 0.1 the null hypothesis, H_0, is not rejected and the residuals do follow the standard normal distribution.

7.2.2 Comparing ODE and SDE Models

From the simple example it can be established that the stochastic version allows for variation in the model and the curved pattern that forms in the residuals decreases. The model is more accurate and gives better result by including the diffusion term and allows for a systematic model error as was expected.

Chapter 8

Conclusions and Discussion

Insulin sensitivity stands for the capability of insulin to increase glucose utilization to peripheral tissue such as muscles and liver. Insulin sensitivity is thus an important factor when scrutinizing how an individual responds to insulin. As well as looking at each individual, it is interesting to look at the whole population and groups within the population and distinguish difference in insulin sensitivity. When attempting to improve the models that are used to estimate insulin sensitivity it is plausible to include covariates and allow for model deviation and random errors.

T2DM patients clearly had lower insulin sensitivity than healthy subjects. That means that T2DM patients are less sensitive to insulin than healthy subjects and require larger amount of insulin to lower blood glucose levels. It is possible to draw the conclusion that T2DM patients are less sensitive to insulin than healthy subjects and it is easer to predict insulin sensitivity for healthy subjects than T2DM since the standard deviation of insulin sensitivity is extensively smaller for healthy subject. The deterministic minimal model gave better results for healthy subjects than T2DM patients in general but included a curved pattern in the residuals which indicates that the model can be improved. The dataset was divided into groups and tested if there were a significant difference between the groups. The groups tested were BMI and gender for all subjects and in addition for T2DM patients different treatment and disease duration were tested. For significance level 0.05 there was a significant difference in insulin sensitivity with respect to gender for healthy subjects and in BMI and different treatment for T2DM patients. A linear regression model for insulin sensitivity was created based on the same covariates that were tested in the groups. For both healthy subjects and T2DM the model did not explain the insulin sensitivity well.

Covariates were included in the deterministic minimal model and a better result occurred by including them. The standard deviation for insulin sensitivity decreased substantially indicating that insulin sensitivity can be explained in parts by the covariates. However the curved pattern still occurred in the residuals. An attempt was made to decrease

or even eliminate the curved pattern by defining a stochastic version of the minimal model. Unfortunately, it was really time consuming and some computational difficulties occurred but a simple example supported that a SDE version decreases the curve pattern and gives better results.

8.1 Future Work

Since computational difficulties occurred during the implementation of the stochastic version of the minimal model it is of interest to continue solving that model. The parameter estimation procedure did not converge and may potentially be because the **PSM** package requires very specific setup. There is a possibility that a small error is in the setup which results in the model not converging. To improve the SDE model even more covariates can be included and since the diffusion term is an constant it can be allowed to vary dependent on each individual. In addition to improving the model, another dataset can be used to see if the parameter estimations changes or if different results are obtained.

Bibliography

[1] K. G. M. M. Alberti, P. Zimmet, Definition, diagnosis and classification of diabetes mellitus and its complications. part 1: diagnosis and classification of diabetes mellitus. provisional report of a who consultation, Diabetic medicine 15 (7) (1998) 539–553.

[2] S. Klim, H. Madsen, N. R. Kristensen, Predictive tools for designing new insulins and treatment regimens, Ph.D. thesis, Technical University of DenmarkDanmarks Tekniske Universitet, OthersAndre (2009).

[3] R. Muniyappa, S. Lee, H. Chen, M. J. Quon, Current approaches for assessing insulin sensitivity and resistance in vivo: advantages, limitations, and appropriate usage, American Journal of Physiology-Endocrinology And Metabolism 294 (1) (2008) E15–E26.

[4] R. N. Bergman, Y. Z. Ider, C. R. Bowden, C. Cobelli, Quantitative estimation of insulin sensitivity., American Journal of Physiology-Endocrinology And Metabolism 236 (6) (1979) E667.

[5] A. Makroglou, J. Li, Y. Kuang, Mathematical models and software tools for the glucose-insulin regulatory system and diabetes: an overview, Applied Numerical Mathematics 56 (3) (2006) 559–573.

[6] M. J. Lindstrom, D. M. Bates, Nonlinear mixed effects models for repeated measures data, Biometrics (1990) 673–687.

[7] J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, R development core team. nlme: Linear and nonlinear mixed effects models, 2012, URL http://CRAN. R-project. org/package= nlme. R package version 3–1.

[8] C. L. Rohlfing, H.-M. Wiedmeyer, R. R. Little, J. D. England, A. Tennill, D. E. Goldstein, Defining the relationship between plasma glucose and hba1c analysis of glucose profiles and hba1c in the diabetes control and complications trial, Diabetes care 25 (2) (2002) 275–278.

[9] J. C. Pinheiro, Model building using covariates in nonlinear mixed-effects models, Journal de la Société française de statistique 143 (1-2) (2002) 79–101.

[10] S. Klim, S. B. Mortensen, N. R. Kristensen, R. V. Overgaard, H. Madsen, Population stochastic modelling (psm)-an r package for mixed-effects models based on stochastic differential equations, Computer methods and programs in biomedicine 94 (3) (2009) 279–289.

[11] S. Klim, S. B. Mortensen, N. R. Kristensen, R. V. Overgaard, H. Madsen, Population stochastic modelling (psm)-an r package for mixed-effects models based on stochastic differential equations, Computer methods and programs in biomedicine 94 (3) (2009) 279–289.