Longitudinal data analysis of fasting blood sugar level of adult diabetic patients In Jimma University Specialized Hospital 2006: An application of semi-parametric mixed model.

By: Tafere Tilahun

A thesis to be submitted to the Department of Statistics, College of Natural Science, Jimma University for partial fulfillment of the requirements for the degree of Master of Science (MSc) in Biostatistics

November 6, 2014

Jimma, Ethiopia

Longitudinal data analysis of fasting blood sugar level of adult diabetic patients in Jimma University Specialized Hospitalb 2006: An application of semi-parametric mixed model.

By: Tafere Tilahun

email: tafere27@yahoo.com

Advisor: Wondwossen Kassahun(PhD)

Co-Advisor: Belay Birlie (Msc,PhD candidate)

November 6, 2014

Jimma, Ethiopia

STATEMENT OF THE OUTHER

I declare that this thesis is a result of my genuine work and all sources of materials used, for writing it, have been duly acknowledged. I have submitted this thesis to Jimma University for the partial fulfillment of the Degree of Master of Science in Biostatistics. The thesis can be in the university library to be made available to borrowers for reference. I solemnly declare that I have not so far submitted this thesis to any other institution anywhere for that award of any academic degree, diploma or certificate.

Brief quotations from this thesis are allowed without requiring special permission provided that an accurate acknowledgement of the source is made. Requisites for extended quotations for the reproduction of the thesis in whole or in part may be granted by the head of the department of statistics when in his or her judgment the proposed use of the material is for a scholarly interest. In all other instances, however, permission must be obtained from the author.

JIMMA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

DEPARTMENT OF STATISTICS

As thesis research advisors, we hereby recommend that the thesis prepared under our supervision by TAFERE TILAHUN, which is entitled " Longitudinal data analysis of Fasting Blood sugar level of diabetic patients In Jimma University Specialized Hospital in 2006" be accepted in partial fulfillment of the requirement for the degree of master of science (Msc) in Biostatistics.

Wondwossen Kassahun (PhD) .....

Advisor Signature

Belay Birlie (Msc, PhD candidate) .....

Co-advisor Signature

As the members of the board of examiners of Msc thesis open defense examination of TAFERE TILAHUN ANILEY, we certify that we have read and evaluated the thesis and examined the candidate. We recommend that the thesis be accepted as it fulfills the requirements for the degree of Master of Science (Msc) in Biostatistics.

Approve by the board of examiners:

....

Chairman Signature

....

External Examiner Signature

....

Inernal Examiner Signature

Acknowledgement

With most lists, there is an inherent emphasis placed on the first and last slots, as these are the most unique. Regarding my acknowledgements, I have specifically chosen those to thank first and last, in between, I do not have any specific order or agenda. As far as I know, there hasn't been a thesis produced without the dedicated assistance of an advisor and co-advisor, and this thesis is certainly with no exception, it would not have been completed without the supervision and guidance of Wondwossen Kassahun (PhD) and Mr. Belay Birlie (Msc, PhD candidate) starting from title selection to the end of my work. I have a tremendous amount of gratitude towards them for the myriad ways in which they helped me to come up with this kind work. Furthermore, I would like to thank Mr Zelalem Mehaire for his interest and discussions. I would like to tanks all my friends who were besides me in all aspects. Last but not least, I am forever grateful to my brother who supported me through my sorrow and happiness all along the way.

Contents

List of Tables

List of Figures

Acronyms

ADA -American Diabetic Associations

AIC - Akaike Information Criterion

BIC - Bayesian Information Criterion

DBP - Diastolic Blood Pressure

edf - effective degrees of freedom

FBS - Fasting blood sugar

GLM - Generalized Linear Model

GEE - Generalized Estimating Equation

GLMM - Generalized Linear Mixed Model

IDDM - Insulin-Dependent Diabetes Mellitus

IDF - International Diabetes Federation

LMM - Linear Mixed Model

mg - milligram

MmHg-millimeter mercury

NIDDM - Non Insulin-Dependent Diabetes Mellitus

SBP - Systolic Blood Pressure

WHO - World health organization

Abstract

Background: Diabetes mellitus is a metabolic disorder of multiple aetiology characterized by chronic hyperglycaemia with disturbances of carbohydrate, fat and protein metabolism resulting from defects in insulin secretion, insulin action, or both. Fasting blood sugar level is a quantity used to detect the diabetes status of patients. There were several factors for the change in their fasting blood sugar level of adult diabetic patients.

Objectives: The main aim of this study is to investigate the fasting blood sugar level of adult diabetic patient's overtime who have been under follow up from 2004 to July 2006 in Jimma University specialized hospital.

Methods: This study used data obtained from a retrospective cohort follow up study from adult diabetic patients who have been under follow up from 2004 to July 2006 in Jimma University Specialized Hospital. All patients included in this study were those diabetic patients whose age is 18 years and above and who have been followed at least two times. This thesis deals with the regression analysis of repeated measurements taken at irregular and possibly subject-specific time points. Therefore, we proposed a Semi-parametric Mixed-effects model (SPMM) using smoothing spline to investigate the evolution of FBS overtime.

Result and Discussion: From 534 adult diabetic patients, 200(35.27%) were females and 367(67.73%) were males. Twenty five percent were type I diabetic and seventy five percent were type II diabetic patients. The mean age of patients at baseline was 45.16 years and that of the mean fasting blood sugar level(FBS) level of patients at baseline was 204 mg/dl. In both linear mixed models and semi-parametric linear mixed models FBS level declines over time, but the predictive ability of the SPMM model is increased. The semi-parametric mixed model further suggested that the decline was non-parametric. Follow up time (p-value=0.000) and weights (p-value=0.000) are the only significant factors for the change of fasting blood sugar level.

Conclusion and Recommendation: Semi-parametric mixed effect model with quadratic time random effects fits the data appropriately to the continuous measurement of fasting blood sugar level in mg/dL of adult diabetic patients.

1 INTRODUCTION

1.1 Background

Diabetes mellitus is a metabolic disorder of multiple aetiology characterized by chronic hyperglycaemia with disturbances of carbohydrate, fat and protein metabolism resulting from defects in insulin secretion, insulin resistance, or both (WHO, 1999). The effects of diabetes mellitus include long-term damage, dysfunction and failure of various organs. Diabetes mellitus may present with characteristic symptoms such as thirst, polyuria, blurring of vision, and weight loss. In its most severe forms, ketoacidosis or a non-ketotic hyperosmolar state may develop and lead to stupor, coma and, in absence of effective treatment, death. Often symptoms are not severe, or may be absent, and consequently hyperglycaemia sufficient to cause pathological and functional changes may be present for a long time before the diagnosis is made. The long-term effects of diabetes mellitus include progressive development of the specific complications of retinopathy with potential blindness, nephropathy that may lead to renal failure, and/or neuropathy with risk of foot ulcers, amputation, Charcot joints, and features of autonomic dysfunction, including sexual dysfunction. People with diabetes are at increased risk of cardiovascular, peripheral vascular and cerebrovascular disease. There are two types of diabetes: these are Type I diabetes or Insulin Dependent Diabetes Mellitus (IDDM) and Type II diabetes or Non-Insulin-Dependent Diabetes Mellitus (NIDDM).

Type 1 diabetes(IDDM) occur as a results of auto-immune beta-cell destruction in the pancreas, characterized by a total absence of insulin production. Type 2 diabetes is characterized by high levels of blood sugar (glucose) resulting from defects in insulin production, insulin action, or both. Over time, high blood glucose levels can lead to serious complications including blindness, kidney failure, nerve damage, cardiovascular disease, heart disease and stroke and even death.

Excess body weight or obesity is associated with insulin resistance (Hhans et al., 2006) and is overwhelmingly associated with incidence of type 2 diabetes (Colditz G et al., 1990; Colditz G et al.,1995; Perry IJ et al., 1995; Vanderpump MPJ et al., 1996; Wilson P et al., 2007). In ad-

dition to obesity several lifestyle, environmental and demographic factors have been associated with diabetes, the main ones being: physical activity, genetic causes, pregnancy, environmental factors like what you eat and how active you are, combined with genetic causes affect risk of developing diabetes.

Many people with type 2 diabetes have a family member with either type 2 diabetes or other medical problems associated with diabetes, such as high cholesterol level, high blood pressure or obesity. The lifetime risk of developing type 2 diabetes is 5 to 10 times higher in first-degree relatives (sister, brother, son, daughter) of a person with diabetes compared with a person with no family history of diabetes. The risk of developing type 2 diabetes is also greater as we get older. A person is considered to be diabetic if he or she has fasting blood sugar level of 126 mg/dL (7.0 mmol/L) or higher (http://www.uptodate.com/contents/diabetes-mellitus-type-2-overview-beyond-the-basics). There are risk factors for the decrease or increase of fasting blood sugar level of patients. These are overweight, family history, Age, type of diabetes, blood pressure, sex and etc. These factors will also affect the disease progression (http://www.who.int/diabetes/facts/en/ accessed on Dec 5, 2011). Therefore in this study, the effect of these factors on the level of blood sugar of patients was investigated.

The global prevalence of Diabetes has been persistently rising for the last few decades and it is being recognized as a world wide epidemic (Shaw and Sicree Zimmet, 2010). According to predictions from the World Health Organization (WHO, 2013) developing countries bear the highest share of the diabetes epidemic. Currently, more than 80% of people with diabetes live in low- and middle income countries (WHO, 2013). In 2013, the global prevalence of diabetes was estimated at 8.3%. Diabetes does not spare Africa. Although the current estimated prevalence in Africa is relatively lower (3.8%), the region is expected to experience the highest increase in its diabetes prevalence in the next two decades. The highest prevalence of diabetes in the Africa Region is in the island of Runion (16.3%), followed by Seychelles (12.4%), Botswana (11.1%) and Gabon (10.6%). Some of Africa's most populous countries also have the highest number of people with diabetes, with Nigeria having the largest number (3.0 million), followed by South Africa (1.9 million), Ethiopia (1.4 million), and Kenya (769,000). The top six countries with

the highest number of people with diabetes make up just over half of the total number in the region (http://www.idf.org/diabetesatlas/5e/africa). Ethiopia, which is one of the developing nations, is at a risk of increased diabetes incidence. In Ethiopia, the number of deaths attributed to diabetes reached over 23,869.00 in 2012 (IDF, 2012, Diabetes atlas 5th edition).

In many scientific disciplines, studies that predict or forecast what will happen in the future have contributed to our understanding of the world. The value of scientific studies that provide models to inform strategies that can modify and possibly mitigate future events is of importance to society. In the field of epidemiology, prediction models are underrepresented and the concept of risk prediction is overshadowed by the estimation of relative risk measures to clarify etiological perspectives of disease. Etiological models use the same estimation procedures as most predictive modeling (i.e., regression) in order to quantify the relative risk associated with a particular exposure on an outcome. Though regression is often used for both purposes, the way in which the model is constructed will differ due to the goals of the model (Laura Rosella et al., 2009).

Longitudinal studies represent enormous advantages over cross-sectional studies in terms of providing foundations for causal inference. There are several different approaches to the analysis of longitudinal data. Generalized mixed-effects regression models are now quite widely used for analysis of longitudinal data. These models can be applied to both normally distributed continuous outcomes as well as categorical outcomes and other non-normally distributed outcomes such as counts that have a Poisson distribution. Mixed-effects regression models are quite robust to missing data and irregularly spaced measurement occasions and can easily handle both time-invariant and time-varying covariates. As such, they are among the most general of the methods for analysis of longitudinal data. The advantage is that missing data are ignorable if the missing responses can be explained either by covariates in the model or by the available responses from a given subject. The disadvantage is that full-likelihood methods are more computationally complex than qausi-likelihood methods, such as generalized estimating equations (GEE) (Jie Shen, 2011).

In many longitudinal studies, repeated measurements of the response variable are collected at irregular and possibly subject-specific time points. Parametric regression methods for analyzing such data have been well developed by Liard and Ware (1982) and Liard and Zeger (1986) among others, summarized by Diggle, Liang, and Zeger (1994). Although these methods are highly useful, they require parametric specification for the conditional mean of the response variable. However, this assumption is too restrictive and may not be fulfilled in many studies. To avoid this, Moyeed and Diggle (1994) and Zeger and Diggle (1994) proposed a semi-parametric model that relates the response at time t to the vector of covariates.

Semi-parametric regression is concerned with the flexible incorporation of non-linear functional relationships in regression analyses. Semi-parametric mixed models (SPMMs; Diggle et al., 2002; Zhang et al., 1998) are useful extensions to linear mixed models (Lard and Ware, 1982; Verbeke and Molenberghs, 2000; Diggle et al., 2002) and provide a flexible framework for analyzing longitudinal data. An SPMM uses parametric fixed effects to represent covariate effects that have a known parametric relationship with the response variable, unknown smooth function for covariates that have a complex non-linear relation with the response variable, and modeling, the within-subject correlation using random effects and stochastic processes.

This research is motivated by the existence a longitudinal follow up study of diabetic patients in Jimma university specialized Hospital, where patients fasting blood sugar level is routinely measured along with many individual factors, such as, age, sex, participants' physical characteristics, bodyweight, systolic blood pressure, diastolic blood pressure etc. Notable features of this data are the presence of unequal numbers of unevenly spaced and intermittently missing observations for each individual. Zelalem et al., 2014 analyzed this data using liner mixed model assuming a quadratic relationship between fasting blood sugar level and observation time and their result indicated that on average the Fasting blood sugar level of adult diabetic patients decreased over time quadratically and then gradually it becomes rebound and stable. However, modeling such complex nonlinear relationship with parametric function using higher order polynomials is not ideal due to the known undesirable properties of higher order polynomial models. Therefore, in this study we propose a data driven approach using SPMM to model the

association of the fasting blood sugar level and covariates, and to investigate the evolution FBS level over time in Diabetic patients followed in Jimma University specialized hospital during September 2004 to June 2006. The methodology is designed to take advantage of the extra information about the typical curve available with irregular observation times. The reminder of this thesis is organized as follow; Section 1.2 describes about the problem statement, chapter 2 describes review of related literature, chapter 3 describes some general methods for longitudinal data analysis, with emphasis on semi-parametric mixed-effect modeling , chapter four describes the Result and discussion of the study and chapter five describes conclusion and recommendations.

1.2 Problem Statement

Diabetes mellitus is one of the chronic diseases, which is a growing public health problem in both developed and developing countries causing severe and costly complications, including blindness, kidney and heart diseases, strokes, nerve damage and amputations. Uncontrolled diabetes can complicate pregnancy, and birth defects are more common in babies born to women with diabetes. In Ethiopia, the number of diabetic people has been increasing over the years since it attained independence (Ethiopian Diabetic Association, 1993). For such longitudinally collected data often parametric models such as a linear mixed model is used. This models are known to be parsimonious and efficient when the models are correctly specified Using LMM by imposing a parametric function to model the mean change of FBS level over time may be too restrictive and more flexible modeling approach using non-parametric methods is needed. Therefore we propose a semi-parametric mixed modeling approach to investigate the evolution of FBS level over time by then the study aimed to be able to address the following research questions:

- Do groups have a similar over all FBS profile over time?
- What is the average evolution of fasting blood sugar level of diabetic patients?
- What are the characteristics of the deviations of each patient from a typical curve? The deviations are potentially useful in identifying subjects who are progressing more slowly with the hope of finding factors associated with their longer life periods. The individual curves can also be used in counseling about the disease progression.

1.3 Objectives of the study

1.3.1 General objective

The general objective of this study is modeling FBS level of adult diabetic patients who have been underfollow-up fro September 2004 to June 2006 in JUSH through the application of Semi-arametric linear mixed effect model.

1.3.2 Specific objectives

- To explore the individual and average change of fasting blood sugar level among adult diabetic patients using a data driven semi-parametric mixed model.
- To estimate the rate of change of FBS level of diabetic function and then to examine for existence of a feature such as bumps and dips in regression curves.
- To fit an appropriate statistical model for the mean evolution of FBS level and identify the potential factors among adult diabetic patients.

1.4 Significance of the study

Diabetes mellitus is a multi system disorder. Its complications can involve any organ in the body. Some of these complications include visual impairment, cardiovascular disease, limb and brain damage, impotence, kidney failure, urinary tract and stroke. Because of its chronic nature, the severity of its complications and the means required to control them, diabetes is a costly disease, not only for the affected individual and his/her family, but also for the health authorities. Death from diabetes or its complications leads to loss of production, which is likely to be more costly to the government than the direct health care costs. However, if proper interventions/ treatment is given, the disease can be controlled and substantially reduce the risk of developing these complications and slow their progression. It is therefore, hoped that the findings of this study would provide an optimal method to handle diabetic cases, which will help the society and the government at large to reduce on the strain and costs. And the study would be an extension of existing models for accurate testing of the fasting blood sugar level and also open room for further research from the mathematical concepts developed.

1.5 Scope of the study

This study was intended to compare and contrast the efficiency of the parameter estimate of model fitted using linear mixed effect and semi-parametric linear mixed effect method, because the data obtained for this study was a measurement of fasting blood sugar level of adult diabetic patients in Jimma Specialized hospital who were under follow up from 2004 to June 2006. Patients were under follow up in irregular manner resulting unbalanced longitudinal data. Therefore this study also tried to specify the correct longitudinal model to analyze the progression of the fasting blood sugar level of diabetic patients over time.

2 LITERATURE REVIEW

2.1 Risk factors of diabetes mellitus

Diabetes mellitus is a complex, chronic disease. It is a condition characterized by an elevation of the level of glucose in the blood. Insulin, a hormone produced by the pancreas, controls the blood glucose level by regulating the production and storage of glucose. All people with diabetes mellitus can be categorized according to clinical stage, and this is achievable in all circumstances. The stage of glycaemia may change over time depending on the extent of the underlying disease processes. The disease process may be present but may not have progressed far enough to cause hyperglycaemia. The aetiological classification reflects the fact that the defect or process which may lead to diabetes may be identifiable at any stage in the development of diabetes - even at the stage of normoglycaemia. Thus the presence of islet cell antibodies in a normoglycaemic individual makes it likely that the person has the Type 1 autoimmune process (Kim M et al., 2006).

Glucose is a simple sugar, which is the body's prime source of energy. The digestive process turns the carbohydrates of a meal eaten into this glucose which is then distributed throughout the body via the bloodstream, thus, "blood sugar". Insulin, the hormone secreted by the pancreas, is what maintains the proper levels of blood sugar. However, when the pancreas fails to produce enough insulin to create a proper release of glycogen from the liver to the bloodstream the result is high blood sugar, or diabetes mellitus. Subjects with diabetes mellitus have blood glucose level of greater than or equal to 180mg/100ml (10mmol/l) of blood (Olive D. Buhule et al.,2007).

Certain conditions or characteristics can increase a person's risk for developing diabetes. Some of these risk factors, such as Age, family history, having close relative (parent or sibling) with diabetes, having certain racial/ethnic background and having had gestational diabetes or a baby weighing more than nine pounds at birth. Other risk factors, such as overweight or obesity, high blood pressure, low cholesterol, high triglyceride levels, and lack of physical activity are also the risk factors. Obesity is the primary modifiable risk

factors for diabetes (Chondra M.Lockwood et al., 2008).

Viswanathan et al. (1996) in their study found that nearly 75% of the NIDDM patients have first degree family history of diabetes. The prevalence among offspring with one diabetic parent to be 36%, which increased to 54% when there, was a positive family history of diabetes on the non diabetic parental side also. When both parents had diabetes, the prevalence rate increased further 62% typical estimates of the heritability of type II diabetes mellitus and related continuous traits (beta cell function, insulin sensitivity, BMI) lie in the 30 to 60% range (Stumvoll et al., 2005).

Zelalem et al., (2014) used a data obtained from retrospective cohort follow up during 2004-2006 from adult diabetic patients in Jimma University Specialized Hospital, to study the associated factors of FBS of adult diabetic patients and to fit an appropriate statistical model. In this study all diabetic patients whose age 18 and above years were included in the study. A LMM were fitted with quadratic time effects to assess whether good control of fasting blood sugar level changed over time and it was found that fasting blood sugar level of adult diabetic patients decreased over time quadratically. Follow up time (-9.47765, se=1.79), follow up time square (0.2827, se=0.0611), weight (-1.66153, se=0.20217), DBP (0.30078, se=0.1519), and time interaction with weight (0.05221, se=0.023935) was the significant factors for the change in the FBS level. Since the data has a complex profile, fitting using LMM may not be good specification of the model. Therefore this study is motivated to model the data using a semi-parametric mixed model which allows fitting time invariant covariates parametrically and time variant non-parametrically.

2.2 Public health burden of diabetes

Diabetes is the fastest growing chronic disease worldwide (Dawit Worku et al., 2010). It is a progressive, unrelenting and challenging disease with serious complications which can reduce both quality of life and life expectancy. For better surveillance, the full extent of the individual and societal burden of diabetes has become apparent and it is a very common, serious, and costly chronic disease. The worldwide burden of diabetes is likely

to increase dramatically over the next several decades because of an increase in incidence and improvements in detection (Aubert R et al., 1997; Harris MI, et al., 1996; Liebson et al., 1997; Ford ES et al., 1999). Thus, diabetes is increasingly being viewed as both a clinical and a public health challenge (Vinicor F et al., 1994).

Several factors underline the increasing diabetes burden worldwide, particularly a growing incidence of type 2 diabetes, including increase in obesity (Ford ES et al., 1999; Bjorntorp P et al., 1997) and inadequate physical activity, both of which are associated with greater likelihood of developing diabetes (Liebson et al., 1997; www.cdc.gov/Diabetes/statistics); the increasing prevalence of diabetes with age and the steady aging of populations worldwide (Kelly DT et al., 1997).

Planning for health care and public health resources needed to address the significant burden of diabetes (www.cdc.gov/Diabetes/statistics) as an important aspect of population health management, which can be informed by robust prediction tools (Diabetes Atlas, 5th Edition, 2012). This tool can aid policy makers, planners, and physicians by providing reliable estimates of the upcoming diabetes epidemic. In addition, the effectiveness of widespread prevention strategies can be improved by knowing which groups to target and how extensive a strategy is needed to stabilize or reduce the number of new cases (Diabetes Atlas 5th edition, 2012). In Ethiopia the burden of diabetes has becoming increased from year to year and is becoming a public health problem (Ethiopian diabetic association, 2012).

2.3 Statistical models for longitudinal data

2.3.1 Longitudinal Data

In many longitudinal studies, each individual may experience the same event repeatedly at distinct time points during a relatively long follow-up time. These data may occur frequently in a wide variety of setting, including epidemiology, clinical trials, and so on. However, individuals often selectively miss their visits or return at non-scheduled points

in time, as a result, the measurement times are irregular yielding a highly imbalanced data structure. In addition, the frequency and timing of the visits may be informative with respect to the longitudinal outcomes. The data obtained for this study was a type of unbalanced longitudinal measurement and the goal of this study was to develop a general methodology for analyzing data with these features. Due to their unbalanced nature, many longitudinal data sets cannot be analyzed using multivariate regression techniques. A natural alternative arises from observing that subject-specific longitudinal profiles can often be well approximated by linear regression functions (Verbeke and Molenberghs, 2000). There are several different general approaches to the analysis of such longitudinal data.

The standard GLM assumes that the observations are uncorrelated. Extensions have been developed to allow for correlation between observations. The correlation must be accounted for by analysis methods appropriate to the data. Possible consequences of analyzing correlated data as if it were independent are: (i) incorrect inferences concerning regression parameters due to underestimated standard errors; (ii) inefficient estimators, that is more mean square error in regression parameter estimators than necessary. A straightforward application of generalized linear models to longitudinal data is not appropriate, due to the lack of independence among repeated measures obtained on the same individual.

There is an approach for extending generalized linear models to longitudinal data that leads to a class of regression models that are known as marginal models. In marginal models the mean response at each occasion depends only on the covariates of interest, and does not incorporate dependence on random effects or previous responses. This is in contrast to generalized linear mixed models (GLMMs) where the mean response is modeled not only as a function of covariates but is conditional also on random effects.

The most silent feature of marginal model is regression model, with appropriately specified link function, relating the mean response at each occasion to the covariates. For es-

timation of the regression model parameters, marginal models do not necessarily require distributional assumptions for the vector of longitudinal responses. When full distributional assumption for the vector of responses are avoided, the marginal model is said to be semi-parametric, and this leads to a method of estimation know as generalized estimating equation (GEEs) (Liang and Zeger, 1986) allow for the correlation between observations without the use of an explicit probability model for the origin of the correlations, so there is no explicit likelihood. They are suitable when the random effects and their variances are of inherent interest, as they allow for the correlation without explaining its origin. The focus is on estimating the average response over the population ("population-averaged" effects) rather than the regression parameters that would enable prediction of the effect of changing one or more components of covariates on a given individual.

2.3.2 Linear Mixed Model

Linear Mixed-effects Model (LMM) provides a flexible and powerful tool for the analysis of longitudinal data. Mixed-effects model are quite robust to missing data and irregularly spaced measurement occasions and can easily handle both time-invariant and time-varying covariates. It has been a popular method to model the between-subject and within-subject correlations, to handle both balanced and unbalanced scenarios, and allows the inclusion of covariables. In mixed-effect models, response variable are assumed to be a function of fixed effect, non-observable random effect, and error term. When both the fixed and the random effects contribute linearly to the response, the model is called linear mixed-effects model (Davidian, M. et al., 2011). These models are useful in a wide variety of disciplines in the physical, biological and economic sciences. They are particularly useful in settings where repeated measurements are made on the same statistical units, or where measurements are made on clusters of related statistical units (Wondwosen et al., 2010). Linear mixed-effects models rely on assumptions of multivariate normality, and likelihood-based inferences for both the fixed and random effects are relatively straightforward (Davidian, M. et al., 2011).

A major limitation of these methods is that the relationship of the mean of a longitudinal response to covariates is assumed fully parametric. Although such parametric mean models enjoy simplicity, they have suffered from inflexibility in modeling complicated relationships between the response and covariates in various longitudinal studies. Therefore, in order to account this limitation, a semi-parametric mixed model were constructed for longitudinal data, where flexible functional forms can be estimated from the data to capture possibly complicated relationships between longitudinal outcomes and covariates.

2.3.3 Semi-parametric Mixed Model

As we have seen, semi-parametric regression models based on penalized splines can be couched in the mixed model framework, allowing for mixed model estimation and for inferential and computational tools to be used. This synergy is similar in spirit to the mixed model approach to analyzing longitudinal data that commenced (Laird and Ware, 1982). Most of the work that has been undertaken to model longitudinal data has been parametric, in the sense that the effects of continuous covariates have been modeled linearly or by using some parametric nonlinear model (Davidian et al.,2011). An alternative to nonlinear mixed modeling is to incorporate smoothing methods. The mixed model representation of penalized splines allows for a seamless fusion between parametric mixed models and smoothing, which we call semi-parametric mixed models. Extension of smoothing spline to longitudinal data requires explicitly accounting for the within-subject correlation in constructing the penalized likelihood function. Zhang et al., (1998) adopted a penalized likelihood approach based on smoothing splines, which is computationally efficient and can be conveniently implemented using standard software. In this study we apply a semi-parametric mixed effect model to model fasting blood sugar level of adult diabetic patients overtime.

3 METHODOLOGY

3.1 Data description and Study design

This study used the data obtained from retrospective cohort follow up study design of fasting blood sugar level of adult diabetic patients who have been under follow up from 2004 to June 2006 in Jimma University specialized hospital. Jimma University Specialized hospital (JUSH) is located in south west of Ethiopia in Jimma Town. JUSH serves as a teaching and referral hospital for the Jimma area community and, adjacent zones and regions of southwest Ethiopia.

3.2 Study Population and period

All patients aged 18 and above years, under follow up during the period 2004 to June 2006 and have at least two FBS level measurement were eligible to be included in the study leading to a total of 534 study subjects. During follow up subjects come to the center at irregular time (one, two three or more months gap) and during their visit their FBS level is measured and recorded in the individual follow up card along with important characteristics such as, weight of patient, blood pressure (systolic and diastolic), their type of diabetes (type I diabetes mellitus or type II diabetes mellitus), etc.

3.3 Variables

3.3.1 Dependent variables

The response variable of interest in this study is the repeated Fasting blood sugar level measurement in mg/dL at each patient visit.

3.3.2 Independent variables

The independent variable which are expected to be associated with FBS level and considered to be investigated in this study include Age, Time of FBS measurement, Sex, blood pressure (Systolic, diastolic), type of diabetes (type I and type II), family history and weight. Detailed information about this is given in Table 1.

Table 1: Description of study variables with its classification

Variable name	Variable Description	Code
1. Gender	Gender of Patients which is a time-invariant variable	0=female, 1=male
2. Age	Age of patients which is time-invariant variable	
3. DT	Diabetic types	0= type I (IDDM), 1= Type II (NIDDM)
4. Famhist	Family history of patients	0=family has no diabetes, 1= family has diabetes
5. SBP	Systolic blood pressure (time-varying covariates)	MmHg
6. DBP	Diastolic blood pressure (time-varying covariates)	MmHg
7. Bodyweight	Body Weight of patients (time-varying covariates	Kg
8. Time	Observation time (time-varying covariates)	month

3.4 Method of data analysis

3.4.1 Exploratory Data Analysis

In this study we have used various mean structure and random effect structure exploration techniques to choose the appropriate fixed effect structure in the proposed mixed model and to decide which parameters in the model, if any, should have a random-effect component included to account for between-group variation. We will examine OLS residuals and information criteria statistics to select the appropriate covariance structure for the random effect and correlation structure for the random noise.

3.4.2 Linear Mixed-Effects Models

Mixed-effects models include both fixed effects and random effects, where random effects are usually introduced to model correlation within a cluster and/or spatial correlations. They provide flexible tools to model both the mean and the covariance structures simultaneously. A general linear mixed-effects (LME) model assumes that

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{\epsilon} \tag{3.1}$$

where $\mathbf{Y}$ is an n-vector of observations on the response variable, $\mathbf{X}$ and $\mathbf{Z}$ are design matrices for fixed and random effects, respectively, $\boldsymbol{\beta}$ is a q1-vector of unknown fixed effects, $\mathbf{b}$ is a q2-vector of unobservable random effects, and $\boldsymbol{\epsilon}$ is an n-vector of random errors. It is assumed that $\mathbf{b} \sim N(\mathbf{0}, \mathbf{D})$ and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Lambda})$ with $\mathbf{b}$ indepenent of $\boldsymbol{\epsilon}$. The mean structure is modeled by the fixed effects, and the covariance structure is modeled by the random effects and random errors.

3.4.3 Semi-parametric linear mixed-effect models

Many statistical models rely on the assumption that the effects of continuous predictors are linear. However, the linearity assumption may be too simple to represent the effects of some risk factors correctly. More specifically, if the linearity assumption is incorrect for a given risk factor, the parametric estimate may underestimate its effect over some range of values or overestimate the effect over some other range, or both. In the last decade, a number of flexible nonparametric extensions of the conventional linear model have been proposed in the statistical literature (Green and Silverman 1995, Eubank 1999). These nonparametric regression methods eliminate the restrictive linearity assumption and thus allow greater flexibility in modeling the data so that the estimated effects of continuous predictors may follow an arbitrary continuous smooth function. Accordingly the risk of bias is greatly reduced as the estimates depend more on empirical data and less on a priori assumptions. Semi-parametric mixed models (SPMMs; Diggle et al., 2002; Zhang et al., 1998) are useful extensions to linear mixed models (Lard and Ware, 1982; Verbeke and Molenberghs, 2000; Diggle et al., 2002) and provide a flexible framework for analyzing longitudinal data. Many authors have studied semi-parametric models for longitudinal data (e.g., Wang, 1998; Diggle et al., 2002; Ruppert et al., 2003).

A Semi-parametric mixed model uses parametric fixed effects to represent covariate effects and a smooth function to model the time effect, modeling the within-subject correlation using random effects and stochastic processes. Zhang et al., (1998) adopted a penalized likelihood approach based on smoothing splines, which is computationally efficient and can be conveniently implemented using standard software. In this study we have used the Zeger and Diggle (1994) model to a more general class of models termed semi-parametric mixed models. This class of models assumes that the mean of the outcome variable is an arbitrary smooth function of time and parametric functions of the covariates, where the non-parametric function of time is estimated using smoothing spline methods, which allows us to make inference on this nonparametric function as part of the overall inference procedure.

Let $\boldsymbol{Y_{ij}}$ denote the fasting blood sugar level of patients measured at month time $\boldsymbol{t_{ij}}$ on patients i(i=1,2,. . . ,$\boldsymbol{n_j}$). A penalized smoothing spline model, with a subject-specific random intercept $\boldsymbol{U_i}$ based on a truncated line basis function can be written as:

$$\mathbf{Y}_{ij} = \mathbf{f}(\mathbf{t}_{ij}) + \boldsymbol{\epsilon_{ij}} \tag{3.2}$$

with $\boldsymbol{f(t_{ij})}$ a smooth function and error $\boldsymbol{\epsilon_{ij}, i = 1, \ldots,}n_j$ iid N(0,$\boldsymbol{\sigma_\epsilon^2}$). The smooth function $\boldsymbol{f(t_{ij})}$ can be found as result of minimization of the residual sum of squares plus a roughness penalty,

$$\textstyle\sum_{i=1}^{n}(\boldsymbol{y_{ij}-f(t_{ij})})^2+\boldsymbol{\lambda}\int(\boldsymbol{f^{(p)}(t)})^2\boldsymbol{dt} \tag{3.3}$$

where $\boldsymbol{Y_{ij}}$ denote the fasting blood sugar level of patients measured at month time $\boldsymbol{t_{ij}}$ on patients i(i=1,2,. . . ,$\boldsymbol{n_j}$). $\boldsymbol{f^{(p)}}$ is the pth derivative of the function $\boldsymbol{f(t)}$. The result curve fitted to the data is a piecewise polynomial of degree 2p-1. The smoothing parameter governs the trade-off between smoothness and goodness of fit. This parameter is often unknown in practice and needs to be estimated from the data. A classical data-driven approach to selecting the smoothing parameter is cross-validation, which leaves out one subject's entire data at a time. However, this approach is often computationally intensive and the subsequent inference is difficult. There are different types of basis functions to represent the smooth function, among them Linear truncated spline basis function is used to represent $\boldsymbol{f(t)}$.

Let $\boldsymbol{K_1, \ldots, K_{d-p}}$ be a set of distinct numbers inside the range of the $\boldsymbol{t_i}$'s, and let $\boldsymbol{t+ = max(0, t)}$. A random coefficient linear regression spline model for $\boldsymbol{f(t)}$ is

$$\boldsymbol{f(t_{ij})} = \boldsymbol{\beta_0+\beta_1 t+\textstyle\sum_{k=1}^{K} u_k(t-K_k)_+} \tag{3.4}$$

where $\boldsymbol{u} = [\boldsymbol{u_1, \ldots, u_k}]^T$ $\boldsymbol{N(0,\sigma_u^2 I)}$ is independent of $\boldsymbol{\epsilon} = [\boldsymbol{\epsilon_1, \epsilon_2, \ldots, \epsilon_n}]^T$, set of functions such as $\boldsymbol{(t - K_k)_+}$ is called a linear spline basis and the values of $\boldsymbol{K_k}$ are referred as a knots. When $\boldsymbol{\sigma_u^2 = 0}$, $\boldsymbol{f(t)}$ is linear, but for $\boldsymbol{\sigma_u^2 > 0}$, the truncated lines $\boldsymbol{(t - K_k)_+}$ flexibility allow for nonlinearities in $\boldsymbol{f}$. More smoothness could be attained

using $\boldsymbol{p}$ global polynomial terms $\boldsymbol{t_1, \ldots, t}$ and $\boldsymbol{p-d}$ truncated polynomials $\boldsymbol{(t-K_k)_+^p}$, k=1,2,,d-p

$$\mathrm{f}(\mathrm{t}_{ij})=\boldsymbol{\beta_0+\beta_1 t+\beta_2 t^2+\beta_3 t^3+\ldots+\beta_p t^p+\sum_{k=1}^{d-p} u_k (t-K_k)_+^p} \quad \textbf{(3.5)}$$

we can combine equations (3.2) and (3.5) in one model,

$$\mathbf{Y=X\beta+Zu+\epsilon} \quad (3.6)$$

where

$$\mathrm{X}=\begin{bmatrix} \mathbf{1} & \boldsymbol{t_1} & \boldsymbol{t_1^2} & \ldots & \boldsymbol{t_1^p} \\ \mathbf{1} & \boldsymbol{t_2} & \boldsymbol{t_2^2} & \ldots & \boldsymbol{t_2^p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{1} & \boldsymbol{t_n} & \boldsymbol{t_n^2} & \ldots & \boldsymbol{t_n^p} \end{bmatrix}, \boldsymbol{Z} = \begin{bmatrix} (\boldsymbol{t_1-K_1})_+^p & \ldots & (\boldsymbol{t_1-K_{d-p}})_+^p \\ \vdots & \ddots & \vdots \\ (\boldsymbol{t_n-K_1})_+^p & \ldots & (\boldsymbol{t_n-K_{d-p}})_+^p \end{bmatrix} \quad \textbf{(3.7)}$$

where

$$\mathrm{cov}\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{\epsilon} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\sigma_u^2 I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\sigma_u^2 I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\sigma_\epsilon^2 I} \end{bmatrix}$$

Here $\boldsymbol{\sigma_u^2}$ measures the between-subject variation, $\boldsymbol{\sigma_\epsilon^2}$ measures the within-subject variation, and $\boldsymbol{\sigma_u}$ controls the amount smoothing done to estimate $\boldsymbol{f}$. Equation (3.6) is nothing but a normal linear mixed model and, for any given $\boldsymbol{\sigma_u}$and $\boldsymbol{\sigma_\epsilon}$, the estimated best linear unbiased predictor (EBLUP) of y,

$$\boldsymbol{\hat{f} = X\hat{\beta} + Z\hat{u}} \quad \textbf{(3.8)}$$

Unbiased refers here to the property that the average value of the estimate is equal to the average value of the quantity being estimated, that is $\boldsymbol{E(f') = E(f)}$. Equation (3.8) can be rewritten (McCulloh and Searle 2001) as

$$\boldsymbol{\hat{f} = C(C^T C+\lambda^2 D)C^T y} \quad \textbf{(3.9)}$$

Where $\mathbf{C}=\begin{bmatrix} \boldsymbol{X} & \boldsymbol{Z} \end{bmatrix}, \mathbf{D} = \boldsymbol{diag}\begin{pmatrix} \mathbf{0_{p+1}}, & \boldsymbol{I_k} \end{pmatrix} \boldsymbol{and}\ \boldsymbol{\lambda^p} = \boldsymbol{\sigma_\epsilon/\sigma_u}$ for the $\boldsymbol{p^{th}}$ degree of penalized spline(p=2 in the case of the quadratic spline model).

A generalized semi-parametric mixed model can be written as

$$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{Zu}+\mathbf{Zb}+\boldsymbol{\epsilon} \tag{3.10}$$

Where $\boldsymbol{\beta}$ is a vector of regression coefficients that models the effects of covariates $\boldsymbol{X}$, the random effect $\boldsymbol{b_i}$ follows a normal distribution. Two sets of random effects $\boldsymbol{b_i}$ and $\boldsymbol{U_i}$ are assumed to be independent. The subject-specific random intercept accounts for the correlated nature of the observations. The truncated line basis is simple in formulation and performs adequately in many circumstances and is therefore a sensible choice. The matrix $\boldsymbol{Z}$ contains the elements of the truncated line basis , as well as columns of ones for random subjects. Further we can combine $\boldsymbol{b_i}$ and $\boldsymbol{U_i}$ in to as $\boldsymbol{b} = (\boldsymbol{b_1}, \boldsymbol{b_2}, \ldots, \boldsymbol{b_K}, \boldsymbol{U_1}, \boldsymbol{U_2}, \ldots, \boldsymbol{U_n})'$. Their covariance function is given below.

$$\text{cov}\begin{bmatrix} \mathbf{u} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\sigma_u^2 I} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\sigma_b^2 I} \end{bmatrix} \tag{3.11}$$

The estimators of the regression coefficients and the truncated s-spline estimator of the non-parametric function are obtained using maximum penalized likelihood. The random effects are estimated using the conditional mean given the data. Fitting penalized splines by the linear mixed model approach has some appealing advantages, such as the automatic determination of the smoothing parameter, a unified framework for inference and ease to extend the models. In what follows, we proceed along these lines by formulating a series of hypothetical semi-parametric mixed models.

3.4.4 Estimation and Inference

For simplicity, the estimation and inference procedures for the semi-parametric linear mixed model (3.9) is presented here. We assume that the covariance matrices D and $\boldsymbol{\lambda}$ depend on an unknown vector of covariance parameter $\boldsymbol{\tau}$. The marginal distribution of $\mathbf{y}$ is

$\mathbf{y} \sim \boldsymbol{N(\eta(\beta, f), \sigma^2 W^{-1}), where}$ $\mathrm{W}^{-1} = \boldsymbol{ZDZ^T + \Lambda}$ and $\boldsymbol{\eta(\beta, f) = X\beta + \gamma(f)}$ for fixed $\boldsymbol{\tau}$,$\boldsymbol{\beta}$ and $\boldsymbol{f}$ are estimated by minimizing penalized weighted least squares

$\text{(y-}\boldsymbol{\eta(\beta, f)})^T \boldsymbol{W(y - \eta(\beta, f))} + \boldsymbol{\lambda} \sum_{k=1}^{K} \boldsymbol{\theta_k^{-1}} ||\boldsymbol{f_k}||^2$.

Inference for the parametric components in $\boldsymbol{\beta}$ can be based on the result of $\boldsymbol{cov}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X^T V^{-1} X})^{-1}$, where V is replaced by $\hat{\boldsymbol{V}}$,in which the variance components in$\boldsymbol{V}$ are replaced by their REML estimates. Under $\boldsymbol{H_o}$ the asymptotic distributional assumption theory for penalized spline mixed model is reasonable. The hypothesis tests about the overall effect and linearity of nonparametric component can be achieved by testing.

$\boldsymbol{H_o : \sigma_u^2 = 0}$ versus $\boldsymbol{H_1 : \sigma_u^2 > 0}$ and for that overall effect of density reduces to

$\boldsymbol{H_o : \beta = \sigma_u^2 = 0}$ versus $\boldsymbol{H_1 : \beta \neq 0}$ or $\boldsymbol{\sigma_u^2 > 0}$

3.4.5 First order derivative test and estimation via penalized splines

For first order derivative estimation via penalized splines, it is recommended that higher degree polynomial basis functions to be used to insure that the resulting derivative estimates are smooth. We will start by describing first order derivative estimation, for which a quadratic spline are the simplest basis leading to continuous fits.

Let $\hat{\boldsymbol{f}}$ be a quadratic penalized spline fit:

$\hat{\boldsymbol{f}}(\boldsymbol{x}) = \hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_1 \boldsymbol{x} + \hat{\boldsymbol{\beta}}_2 \boldsymbol{x^2} + \sum_{k=1}^{K} \hat{\boldsymbol{u}}_k (\boldsymbol{x - K_k})^2$

This is a piecewise quadratic function that can be differentiated over each piece to obtain the piecewise linear estimates of $\hat{\boldsymbol{f}}$:

$\hat{\boldsymbol{f'}}(\boldsymbol{x}) = \hat{\boldsymbol{\beta}}_1 + 2\hat{\boldsymbol{\beta}}_2 \boldsymbol{x} + \sum_{k=1}^{K} 2\hat{\boldsymbol{u}}_k (\boldsymbol{x - K_k})$

Operationally, a derivative estimate at location $\boldsymbol{x}$ can be obtained from the quadratic fit coefficients

$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_0 & \hat{\boldsymbol{\beta}}_1 & \hat{\boldsymbol{\beta}}_2 \end{bmatrix}^{\boldsymbol{T}}$ and $\hat{\boldsymbol{u}} = \begin{bmatrix} \hat{\boldsymbol{u_1}} & \dots & \hat{\boldsymbol{u_k}} \end{bmatrix}^{\boldsymbol{T}}$ by setting

$\boldsymbol{X'_x} = \begin{bmatrix} \mathbf{0} & \mathbf{1} & \mathbf{2x} \end{bmatrix}$ and$\boldsymbol{Z'_x} = \begin{bmatrix} \mathbf{2(x - K_k),} & \boldsymbol{for} \mathbf{1} \leq \boldsymbol{k} \leq \boldsymbol{K} \end{bmatrix}$

then $\hat{\boldsymbol{f}}'(\boldsymbol{x})$=$\boldsymbol{X'_x}\hat{\boldsymbol{\beta}} + \boldsymbol{Z'_x}\hat{\boldsymbol{u}}$

Also $\text{var}(\hat{\boldsymbol{f}}'(\boldsymbol{x}) - \boldsymbol{f}'(\boldsymbol{x})) \cong \boldsymbol{C'_x Cov} \begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\boldsymbol{u}} - \mathbf{u} \end{bmatrix} \boldsymbol{C'^{T}_x} = \boldsymbol{\sigma^2_\epsilon C'_x}\left(\boldsymbol{C^T_x C} + \frac{\boldsymbol{\sigma^2_\epsilon}}{\boldsymbol{\sigma^2_u}}\boldsymbol{D}\right)^{-1}\boldsymbol{C'^{T}_x}$

Where $\mathbf{C'_x}$=$\begin{bmatrix} \mathbf{X'_x} & \mathbf{Z'_x} \end{bmatrix}$ $\boldsymbol{and}$ $\mathbf{D}$=diag$\left(\begin{array}{cccccc} \mathbf{0,} & \mathbf{0,} & \mathbf{0,} & \mathbf{1,} & \cdots & \mathbf{,1} \end{array}\right)$

3.5 Model diagnosis

For model diagnosis, we assessed normality (the QQ-plot and the histogram of residuals), Homogeneity (the residual versus predictor plot, and residuals versus fitted values plot, also called the linear predictor plot for the Gaussian distribution with identity link), and model fit(fitted values versus observed values plot).

3.6 Software

Data were analyzed by using R version 3.0.3, 3.1.0, and 3.1.1 Statistical Software. The GAMM package was used. The Full r-code used for this analysis can be found from ANNEX II.

3.7 Ethical consideration

The data for the analysis obtained was from Jimma university specialized hospital, and an ethical clearance for the study was provided by Research Ethics Review Board of Jimma University and the department of statistics wrote an official co-operation letter to Jimma university specialized hospital. Careful recruitment and training for data collectors from each patient's card was undertake.

4 RESULT AND DISCUSSION

4.1 Explorattory data analysis

4.1.1 Patients Baseline Characteristics

A total of 558 patients who start follow up from September 2004 till end of June 2006, with a minimum of two and maximum of thirty five measurements per individual which results a total of 4405 measurements were included for this study. Among these patients 192(35.96%) were females and 342 (64.04%) of them were males (Table 4.1). The minimum and the maximum age of patients were 18 and 93 years. Their fasting blood sugar level with other important variables was routinely recorded from patient's identity card. The average fasting blood sugar level of the patients is 164.8899 mg/dL.

Table 2: Distribution of diabetes status of adult diabetic patients at baseline, Jimma University Specialized Hospital, from 2004 to June 2006

Characteristics		Type I diabetes n(%)	Type II diabetes n(%)	p-value
Gender	Male	87(16.29%)	255(47.75%)	**0.9935**
	Female	48(8.99%)	144(26.97%	
Family History	Family has no diabetes	37 (6.93%)	380(71.16%)	$< 2.2e^{-16}$
	Family has diabetes	98(18.35%)	19(3.56%)	
Age		34.5355 (11.896.9)*	48.64033 (13.77569)*	2.2e-16
Weight		58.91393(11.17389)*	63.91077(13.73605)*	2.2e-16
FBS		171.7(102.4946)*	162.8(80.73385)*	0.01137
SBP		117.6(115.3, 120)**	122.7(121, 124.5)**	$7.66e^{-04}$
DBP		77.2 (75.5, 78.95)**	78.71 (77.7,79.73)**	$\mathbf{< 2.2e^{-16}}$

4.1.2 Exploring the mean structure using loess smoothing

Figure 4.1 shows that FBS level trajectories by patients using loess smoothing and it can be seen from that in general, the FBS level declines over time, especially in the early months, but with tremendous heterogeneity. From the plot we can guess that there is a linear or quadratic time effects on FBS level of the diabetic patients.

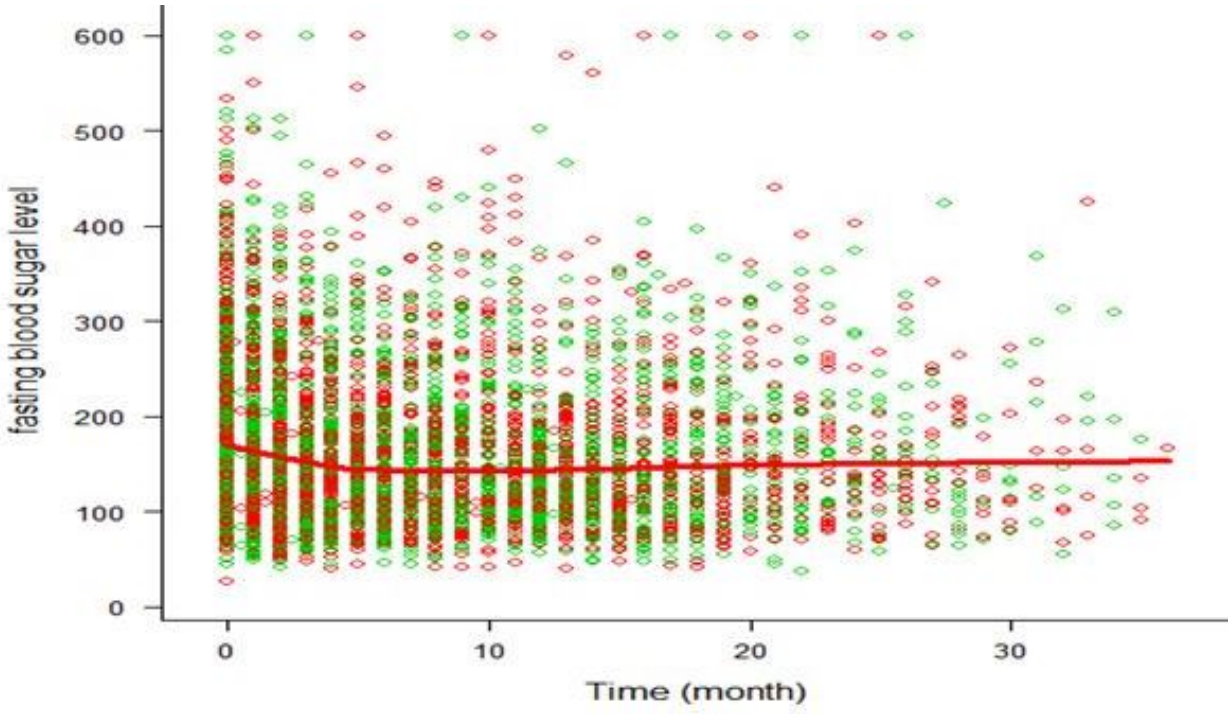

Figure 1: Overall profile plots of FBS level of diabetic patients over time (month) using loess smoothing

4.2 Mixed model building

4.2.1 Parametric Linear mixed models of changes in FBS over time

First a linear mixed effect model was fitted with a quadratic time effects, since the mean evolution declines quadratically with time. In a simple repeated measure linear mixed effects model, time in month was significantly associated with FBS. FBS level declined by 4.82 percentage points per month (P-value$<$ 0.001). From table 6 in ANNEX I LMM of FBS with quadratic time effect, the fixed effects with small p-values (P-value$<$0.05) show that the fixed effects are significant. When all the covariates with time interaction was added to the model, except time

with weight interaction, the other interactions were not significant, but the predictive ability of the model improved significantly (likelihood ratio test statistics= 100.4122, df=24, p<0.001) suggesting that the only significant covariates term should be retained in the model (Table 7, ANNEX I). Of the covariates that were tested in this mode, only weight, time and weight with time interaction was statistically significant. Finally, after removing the non-significant covariates, a model with weight with time interaction were fitted. The final fitted linear mixed effect model framework was given below.

$$\text{FBS}_{ij} = \boldsymbol{\beta_{0i}} + \boldsymbol{\beta_{1i}weight} + \boldsymbol{\beta_{2i}time_{ij}} + \boldsymbol{\beta_{3i}time_{ij}^2} + \boldsymbol{\beta_{4i}weight_{ij} * time_{ij}} + \boldsymbol{b_{0i}} + \boldsymbol{b_{1i}time_{ij}} + \boldsymbol{b_{2i}time_{ij}^2} + \boldsymbol{\epsilon_{ij}} \quad \textbf{(4.1)}$$

In this linear model, weight and time had a negative coefficients (the negative implies that FBS is declining with time), but only square time effect and weight with time interaction has a positive coefficient. Thus, FBS tended to fall with time, but there was significant heterogeneity among the individual in the rate of change in FBS. The summary statistics for the selected model is displayed in table 3.

Table 3: Parameter estimates of linear mixed model with quadratic time effect

Fixed Effects	Value	Std.Error	Deg.fred	t-value	p-value
Intercept	303.00504	13.090723	3866	23.146548	0.0000
weight	-1.96306	0.208243	3866	-9.426738	0.0000
time	-6.09087	1.126493	3866	-5.406930	0.0000
time^2	0.16130	0.026024	3866	6.197941	0.0000
weight:time	0.02871	0.016142	3866	1.778418	0.0754
Random Effects					
Variance Covariance					
Intercept	57.7367563				
time	6.7166296				
time2	0.2407067				
residual	68.6562477				
AIC	BIC	loglik			
50748.23	50824.9	-25362.12			

4.2.2 Semi-parametric mixed model of changes in FBS level over time

Linear mixed model assumes that the relationship between FBS and other covariates was linear. However, this assumption is not without question. To explore this possible non-linearity of the relationship between FBS and time and other covariate, we applied a linear truncated spline method. A semi-parametric linear mixed model is an extension of a linear mixed model by the use of smooth function terms instead of linear terms for covariates. Here, smooth function terms were used for time (month). Building on the results of the LME analyses, we used these variables significant in LME modeling as predictors. For model diagnostics, we assessed i: normality (the Q-Q plot and the histogram of residuals), ii: Homogeneity (residuals vs predictors plot, residuals vs fitted value plot also called the linear predictor plot for the Gaussian distribution with identity link), iii: A model fit (fitted values vs observed values plots). In order to show the significance of adding random component to model these correlated data, a linear regression and linear mixed model has been fitted and compared based on their AIC values and it was found that adding a random term with quadratic time effect is appropriate.

4.2.2.1. GAMM Model 1: (Null Model; Predictors: Smoothing time)

The nonlinear model (GAMM) produced results qualitatively similar to the linear model. The first GAMM model included only the smoothing spline terms. We did not want to assume that the relationship between FBS and time was linear, and therefore we modeled this relationship using a linear truncated s-spline model with 20 knots. The SPMM framework is given below:

$$\text{FBS}_{ij} = \boldsymbol{\beta_{0i}}+\boldsymbol{\beta_{1i}time_{ij}}+\boldsymbol{\beta_{2i}time^2}+\sum_{k=1}^{K} \boldsymbol{u_k}(\boldsymbol{time_{ij}} - \boldsymbol{K_k})_+^2 + \boldsymbol{b_{0i}} + \boldsymbol{b_{1i}time_{ij}} + \boldsymbol{b_{2i}time_{ij}^2} + \boldsymbol{\epsilon_{ij}}$$

This model includes quadratic time effects in month and suggested that the relationship between FBS and time was non-parametric. From ANNEX I Table 8, the effective degrees of freedom (edf) of the smoothed term for the time in month was 8.385, where edf of 1.0 would denote linearity, and values above that non-linearity; the larger the edf, the more non-linearity. The p-value for the smoothed term was less than 0.0001 (F=20.88), indicating that the FBS declined over time. The R-squared of the model was 0.0302. A plot of the relationship between FBS

and smoothed time is shown in Figure 2 (a). In this graph, the y-axis is the contribution of the smoother to the fitted values for FBS. Visually one can appreciate that FBS declines with time: the graph with 95% confidence bands shows that the fitted values for FBS are greater than zero in months 1, 2 and 3 and less than zero for months 5 to 35. The plot for the predicted value of fitted model is shown in figure 2 (b), which shows that the predicted value also decreases with time.

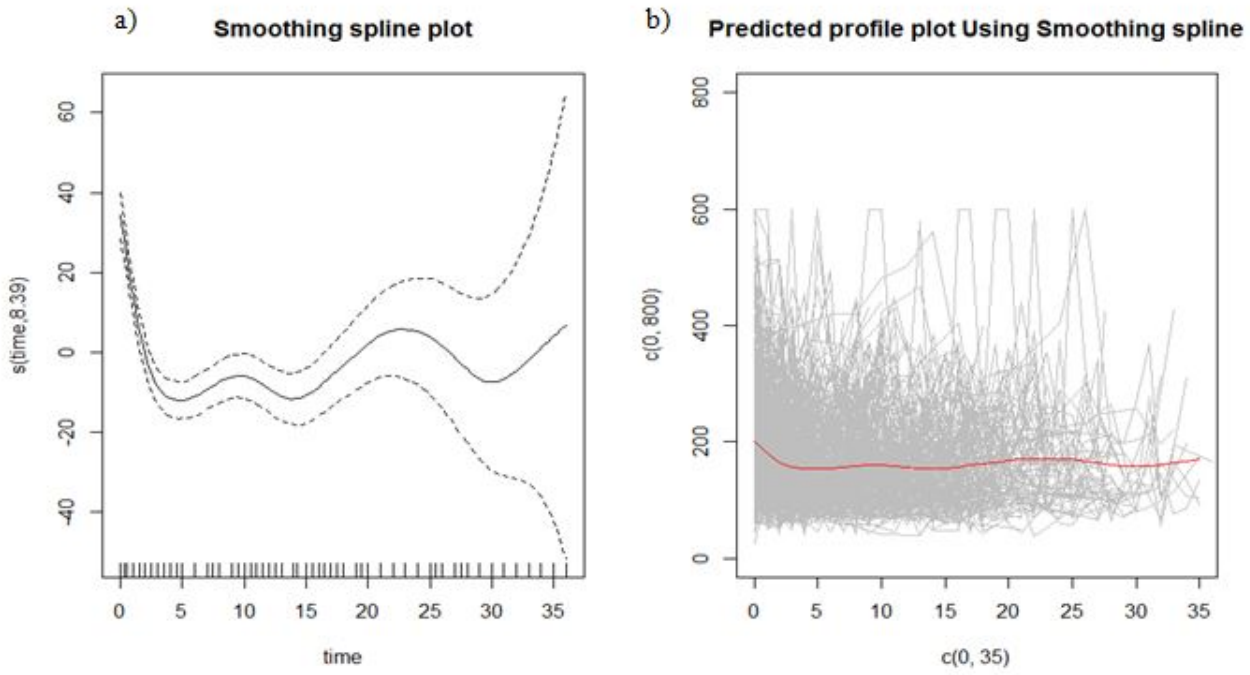

Figure 2: Smoothing and predicted plot for null model.

4.2.2.2. GAMM model 2: (Adding time-invariant covariates on Null Model. Predictors: Age, sex, DT, Famhist, Smoothing time)

For the second GAMM model, we study the effect of time invariant-covariates with smoothing time on change of FBS. The semi-parametric mixed model framework is given below:

$$\boldsymbol{FBS_{ij}}=\boldsymbol{\beta_{0i}} + \boldsymbol{\beta_{1i}time_{ij}} + \boldsymbol{\beta_{2i}time^2} + \sum_{k=1}^{K} \boldsymbol{u_k(time_{ij} - K_k)_+^2} + \boldsymbol{\beta_{4i}age_i} + \boldsymbol{\beta_{5i}DT} + \boldsymbol{\beta_{6i}famhist}+\boldsymbol{\beta_{7i}sex}+\boldsymbol{b_{0i}}+\boldsymbol{b_{1i}time_{ij}}+\boldsymbol{b_{2i}time_{ij}^2}+\boldsymbol{\epsilon_{ij}} \quad (4.3)$$

where FBSt is the fasting blood sugar level of patient at time t (month),$\boldsymbol{\beta_0 t}$ is the intercept, t is the observation time in month, DT is the Diabetic type , famhist is the Family history of diabetic status, Sex is the sex of patient and $\boldsymbol{\epsilon_t}$ is the independently normally distributed noise. The amount of smoothing for smoother can be determined with the AIC or with an automatic selection tool like the cross-validation. From ANNEX 1 Table 9, SPMM of FBS with truncated S-spline with 20 knots, the variable DT, sex, famhist and Age are not significant (p-value 0.05).

4.2.2.3. GAMM model 3 :(Adding time-variant covariates on Model 2 predictors: Weight, smooth time)

A stepwise variable selection method was applied here to select the significant time-dependent covariates. From ANNEX I, Table 10, and Table 11, SBP and DBP with weight interactions were not significant (p-value 0.05). After removing all insignificant covariates and interaction term step by step, a SPMM model with quadratic time effect and smoothing time effect was fitted and compared with parametric linear mixed model. The semi-parametric mixed model framework is given below:

$$\text{FBS}_{ij}=\boldsymbol{\beta_{1i}weight_{ij}}+\boldsymbol{\beta_{2i}weight_{ij}*time_{ij}}+\boldsymbol{\beta_{3i}time_{ij}\beta_{4i}time_{ij}^2}+\sum_{k=1}^{K} \boldsymbol{u_k(time_{ij} - K_k)_+^2} + \text{b}_{0i}+\boldsymbol{b_{1i}time_{ij}}+\boldsymbol{b_{2i}time_{ij}^2}+\boldsymbol{\epsilon_{ij}} \quad \textbf{(4.4)}$$

where FBSt is the fasting blood sugar level of patient at time t (month), t is the observation time in month, and $\boldsymbol{\epsilon_{ij}}$ is the independently normally distributed noise. After removing all the insignificant variables and interaction terms step by step, finally model 3 was fitted and the summary statistics for the parametric estimate and non-parametric estimate is given in table 4. From the Table 4. we can see that weight with time interaction and square of time in months

are not significant.

Table 4: Summary table of comparing semi-parametric and parametric estimate of the linear mixed effects model with quadratic time effects

Semi-Parametric estimate				Parametric esti		
Fixed effects	Value	std.error	p-value	Fixed effects	value	std
Xweight	-1.8406	0.1990	0.0000	Intercept	298.90953	13.3
Xtime	27.4824	6.1481	0.0000	time	-5.75162	1.1
xtime^2	0.4171	0.4050	0.3031	time^2	0.16307	0.0
Xweight:time	0.0175	0.0147	0.2327	weight	-1.89121	0.2
Xs(time)Fx1	-3065.1153	600.2950	0.0000	time:weight	0.02231	0.0
variance covariance				variance covariance		
Random effects	Std.Dev			Random effects	Std.Dev	
Intercept	41.524695			Intercept	57.8614734	
time	0.0016846			time	6.6557840	
time2	8.5882e-05			time2	0.2340871	
residual	72.502262			residual	68.7270782	
	Variance of smooth terms					
s(time)	14.03372					
hline	Approximate significance of smooth terms					
	edf	F	p-value			
s(time)	6.985	91.65	$< 2e^{-16}$ ***			
AIC	BIC	loglik				
50634.55						
50724.0						
-25303.27						

4.2.3 First order derivative test of the smooth function

As mentioned earlier, the primary objective of this study is to estimate the average change of FBS level over time, to compare their profile over time as well as their rate of change as a function of time. The proposed linear truncated S-spline model (equation 4.4) was fitted with a roughness penalty on the second order derivative m=2 to obtain a smooth first order derivative. Figure 3 shows the smoothing spline for FBS level where solid line is a group specific mean smoothing spline and the dotted points show the fitted value vs time profiles. The FBS level decreases markedly in the first 3 to 4 months and becomes more or less stable (does not shows

a moderate change in FBS level). A 95% confidence bands is also plotted around the average line. Figure 4, shows the plot of Predictive value for selected model and first order derivative of the fitted model. The interesting features of this plot are to look at its slope and how it changes with time, in our mind, we note where the slope is positive, where it is negative and where it is essentially zero. It appears that $\boldsymbol{f'}$ is small and negative for early months, meaning that patients were able to decrease or control their fasting blood sugar level up to month 5. But eventually $\boldsymbol{f'}$ reaches a plateau, suggesting that patients do not experience much (if any) rise in FBS level. There is some suggestion in the data that $\boldsymbol{f'}$ has an alternative sign of slope after month 5, meaning that patients do actually may not follow their follow-up time correctly or most patients would not come for treatment after some month due to some unknown reason.

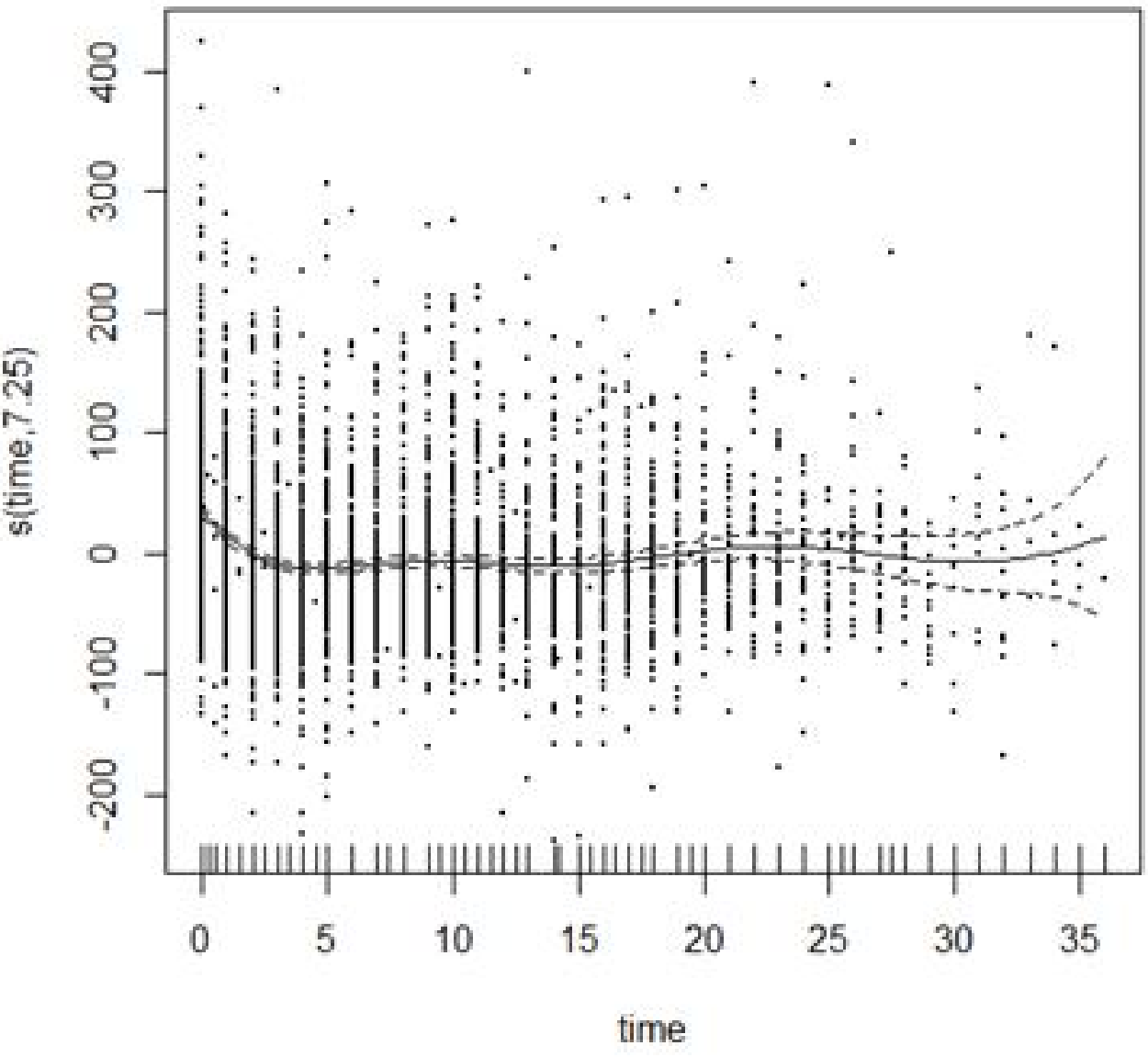

Figure 3: Plot of fitted values vs Observation time in month with a 95% confidence bands

4.3 Model Diagnostics

Graphical diagnostics for the model are shown in Figure 5. The QQ-plot and the histogram of residuals show some non-normality. Although a transformation on FBS could be applied to alleviate this problem, the disadvantage of a transformation is that it changes the type of relationship between response and explanatory variable. Residuals versus predictor plot and residuals versus fitted values plot show that variance is approximately constant. Model fit (fitted values versus observed values plot) shows that some positive linear relationships.

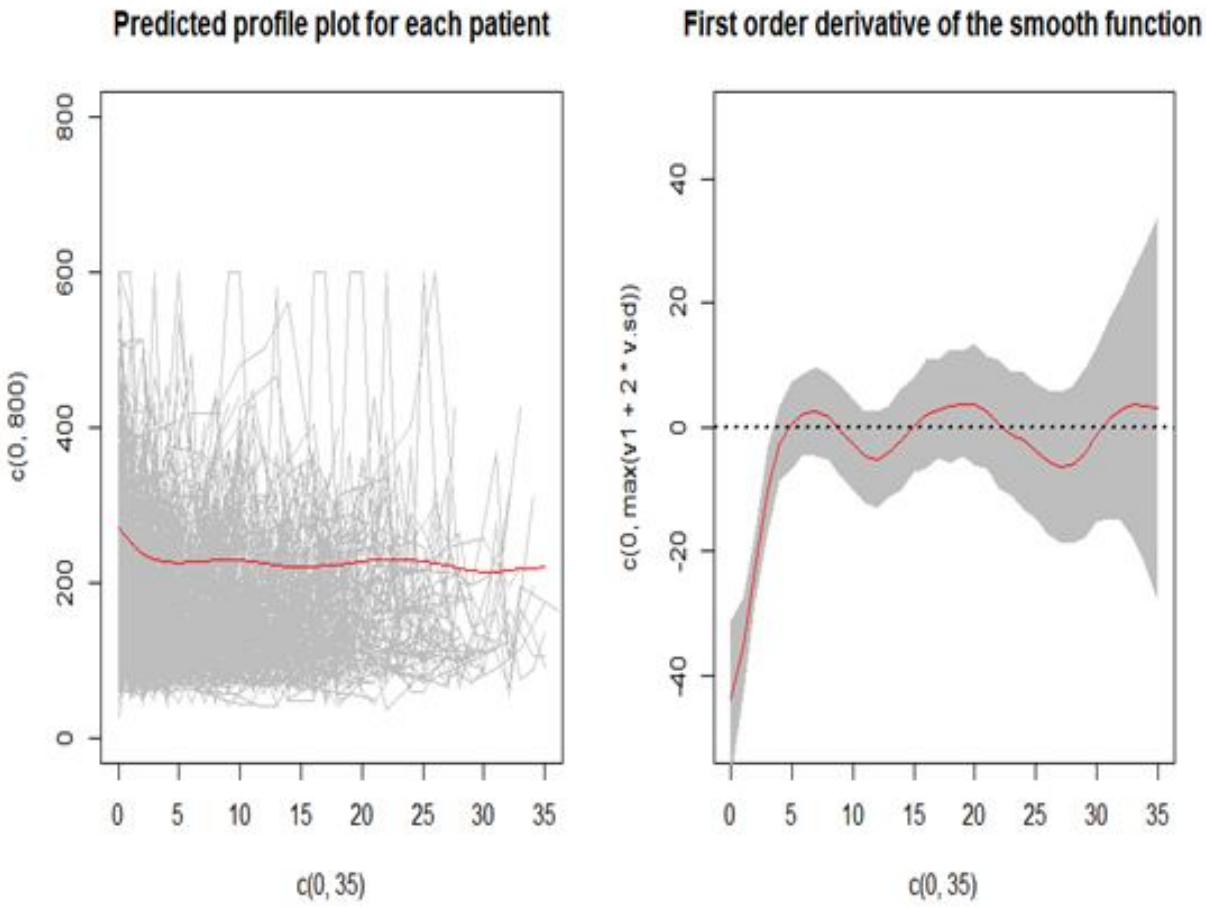

Figure 4: a) Predicted values for FBS from GAMM by time. b) First order derivative plot of the smooth function with confidence bands.

4.4 Comparing Semi-parametric mixed model with parametric mixed model

For the GAMM model 3 (equation 4.4), and LME model (equation 4.1), Table 5 compares the model fitted with linear mixed effect method and model fitted using semi-parametric mixed effect method by AIC, BIC, log(likelihood), and adjusted R-square. The model fitted using a semi-parametric mixed effect approach is the best optimal model by AIC and BIC.

Table 5: Model comparison by AIC, BIC and log likelihood method

Model	AIC	BIC	log likelihood
LME	50757.67	50834.433	-25366.83
SPMM	50634.55	50724.01	-25303.27

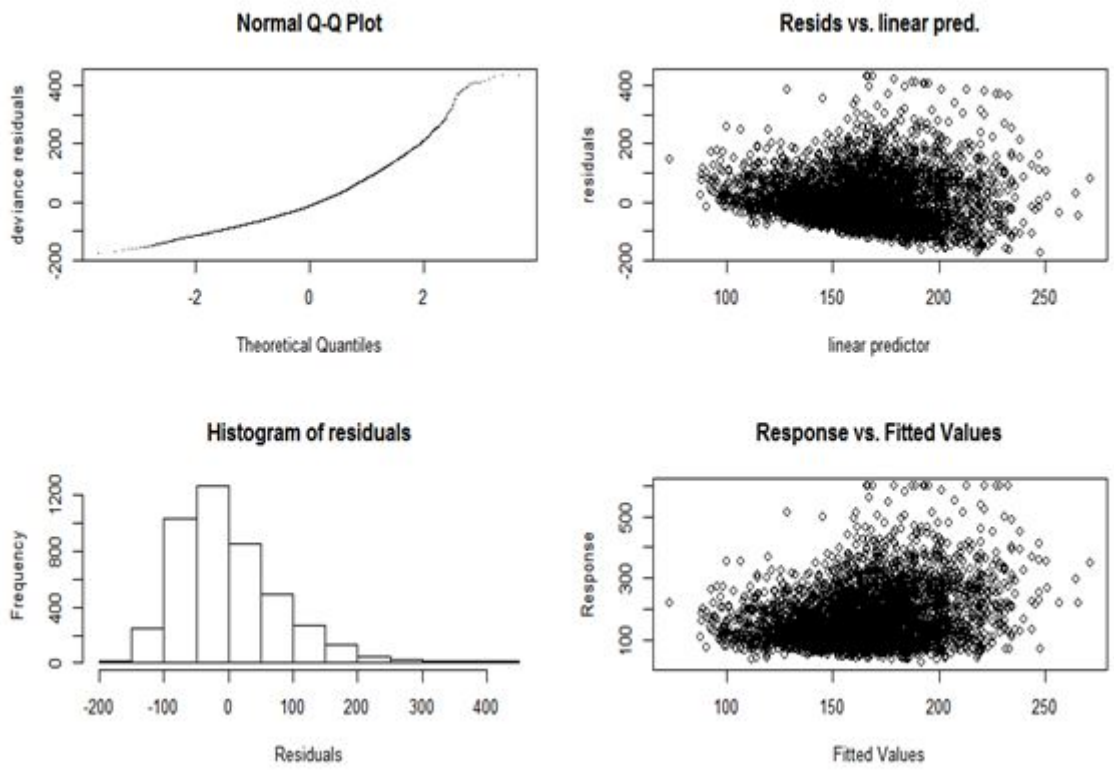

Figure 5: : Model validation graphs for the GAMM model 3

4.5 Discussion

This study was focused on modeling fasting blood sugar level of adult diabetic patients using longitudinally measured data to examine the effect of different covariates on fasting blood sugar level. Modeling FBS is important because it would tell us about the disease progress of the adult diabetic patients. Therefore, in this study we have applied a semi-parametric linear mixed effect model to model the change in the expected value of FBS level over time, because the data are continuous and correlated.

A SPMMs (Diggle, et al., 2002) are a useful extension to linear mixed model (Liard and Ware,1982) and provide a flexible framework for analyzing such a continuous and correlated longitudinally collected data. In SPMM, it is of practical interest to model some covariates whose effects are well understood parametrically while, other covariates effects non-parametrically. In this study we have fitted a SPMM, because it uses parametric fixed effects to represent a covariate effects and a smooth function to model the time effects. We mainly discusses the smoothing spline method to estimate the non-parametric function of time. a key advantage of

using smoothing spline method for a Gaussian data is that smoothing spline have a close connection with linear mixed. A challenge in estimating such semi-parametric mixed effect model is that some parameters are finite dimensional while other parameters are infinite dimensional. Due to this reason, statistical inference in such semi-parametric mixed model was approached with caution.

Since the FBS level data are correlated, continuous and unbalanced in nature, different types of exploratory analysis was used to visualize the evolution of fasting blood sugar level over time. From the individual profile plot Fasting blood sugar level have a complex profile over time, more specifically, on average the Fasting blood sugar level of adult diabetic patients decreased over time and then gradually it becomes rebound and stable with tremendous heterogeneity between subjects and within subjects. Patients were tried to control their fasting blood sugar level for the first few months and then the FBS level becomes inside the 95% confidence bands.

From the semi-parametric mixed model, weight and time are the only significant factors for the change in the fasting blood sugar level. Weight has a negative significant effect on the fasting blood sugar level.

In this study the other covariates (age, sex, SBP, DBP, family history of diabetes) did not have significant effect on the change of fasting blood sugar level of adult diabetic patients.

There were several limitations in this study. In order to compare and contrast the finding of this study we could not be able to find a similar research conducted. And also we could not adjust for potentially important patient level covariates including Depression, income, drug type used for treatment, residence, Dietary habit, physical exercise and other important covariates , because these variables were not recorded on patients identity card.

5 CONCLUSION AND RECOMMENDATION

5.1 Conclusion

This study focused on longitudinal data analysis of fasting blood sugar level of adult diabetic patients in Jimma University specialized hospital using an application of semi-parametric mixed model method. From the individual profile plot, the variability at baseline was higher during the follow up time and Fasting blood sugar level of adult diabetic patients has a complex profile plot.

Semi-parametric mixed model with quadratic time random effect with Autoregressive process of order 1 correlation structure fits well to the data. There was statistically significance difference among adult patients FBS level with respect to time and weight but age, DT, famhist, and sex had no significant effect on FBS level. From the first order derivative plot, patients were tried to lower or manage their fasting blood sugar level up to month 5, after that they were not able to manage the FBS level.

The parameters estimates of the LME and SPMM are different with different the standard errors. This shows that fitting the data using semi-parametric mixed model approach was the right choice.

5.2 Recommendation

Since baseline weight and follow up time are the significant factors that cause for the changes in fasting blood sugar level of adult diabetic patients, a patient must do an exercise in order to control their body weight change and follow their treatment according to the schedule ordered by the doctor. Further Health workers must give awareness for the patients in order to lower their fasting blood sugar level.

In this study, some of the important variables like drug type, level of education, dietary habits, the type of exercise that patients practice and their residence were not included in to the study, further studies has to be conducted in order to know the factors which are highly significant factors for the change in their fasting blood sugar level.

At national level further study has to be conducted by governmental or non-governmental organization in order to lower the prevalence of diabetic mellitus in the country.

In this study we have applied a semi-parametric regression method for the analysis of longitudinal data However, theoretical properties of smoothing splines are only available for Gaussian data. Even for Gaussian data, theoretical properties of the REML estimator of the smoothing parameter and the joint maximum penalized log-likelihood estimators of regression coefficients and the non-parametric function in semi-parametric models are not well understood. Therefore, studying the theoretical side for non-parametric and semi-parametric regression method is desired, including convergence of an iterative algorithm, and consistency, efficiency and robustness for model parameter and spline estimator.

Reference

Alexander Pedan: SUGI proceedings: Statistics and Data Analysis, Smoothing with SAS PROC MIXED; 28:268-28, 2003.

Aubert R: Trends in the global burden of diabetes,1995-2025(Abstract).Diabetes 46(Suppl.1): 139A, 1997.

Bjorntorp P: Obesity. Lancet 350: 423-426, 1997.

Chondra M. Lockwood,. The burden of diabetes in Oregon surveillance report. Oregon Diabetes Program, march 2008;971: 673-0984.

Colditz GA, Willet WC, Rotnizky A Manson J. Weight gain as a risk factor for clinical diabetes mellitus in women. Annals of Internal Medicine 1995;122:481-6.

Colditz, G.A., Willet, W.C., Stampfer, M.J. et al., (1990) Weight as a risk factor for clinical diabetes in women. American Journal of Epidemiology ,132,501-13.

David K McCulloch, MD :Patient information: Diabetes mellitus type 2: Overview (beyond the basics) http://www.uptodate.com/contents/diabetes-mellitus-type-2-overview-beyond-the-basics accessed on March 7 2014

Diabetes Report Card (2012).National Center for Chronic Disease Prevention and Health Promotion: www.cdc.gov/Diabetes/statistics A report of the Surgeon General. Atlanta, GA, Centers for Disease Control and Prevention, 1996.

Diggle, P.J., Heagert, P.J., Liang, K.Y., and Zeger,S.L.(2002), Analysis of longitudinal data. Oxford: Oxford University Press.

Diggle, P.J., Liang, K.Y., and Zeger, S.L. (1994), Analysis of Longitudinal data, Oxford, U.K.: Oxford University Press.

Eubank, R.L. (1999), Nonparametric Regression and Spline Smoothing, New York: Marcel Dekker.

Federal Interagency Forum on Aging- Related Statistics: Data Base News in Aging. Washington, DC, Bureau of census, 1996.

Fitzmaurice, G.; Davidian, M.; Verbeke, G.; Molenberghs, G. Longitudinal Data Analysis: AHandbook of Modern Statistical Methods. Boca Raton, Florida: Chapman Hall/CRC; 2008.

Ford ES,Williamson DF,LiuS: Weight change and diabetes incidence: findings from a national cohort of US adults. Am J Epidemiol 146:214-222,1999

Gebre-Yohannes A and Rahlenbeck SI. Glycaemic control and its determinants in diabetic patients in Ethiopia. Diabetes Res Clin Pract,1997,35,129-134.

Green, P.J. and Silverman B.W. (1995), Nonparametric Regression and Generalized Linear Models, London: Chapman and Hall.

Harris MI, Eastman RC: Early detection of undiagnosed non-insulin dependent diabetes mellitus. JAMA 276:1261-1262, 1996

International Diabetes Federation. Diabetes atlas, $\mathbf{5}^{th}$ edition. via http://www.idf.org/diabetesatlas/5e/th
global burden: Accessed 2012 November 20.

Kahn SE, Hull RL, Utzschneider KM: Mechanisms linking obesity to insulin resistance and

type 2 diabetes.Nature444:840 - 846, 2006.

Kelly DT: Our future society: a global challenge. Circulation 95:2459-2464, 1997.

Kim M, Berger D, Matte T. Diabetes in New York City: Public Health Burden and Disparities. New York: New York City Department of Health and Mental Hygiene, 2006.

Laird, N.H., and ware, J.H. (1982),"random-effect models for longitudinal data," Biometrics, 38; 963-974.

Leibson CL,O'Brien PC,AtkinsonE, Palumbo PJ, Melton LJIII: Relative contributions of incidence and survival to increasing prevalence of adult onset diabetes mellitus: a population based study. Am J Epidemiol 146:12-22,1997

Liang, K.Y. and Zeger, S.L.(1986). Longitudinal data analysis using generalized linear models, Biometrika 73,13-22.

McCulloch, C.E. and Searle, S.R. (2001), Generalized, Linear, and Mixed Models, New York: John Wiley Sons.

McCullagh, P. and Nelder, J.(1989). Generalized Linear Models. 2nd edition. London: Chapman and Hall. Pp.150-300

Moyeed, R.A. Diggle, P.J. (1994). rates of convergence in semiparametric modeling of longitudinal data. Australian Journal of Statistics, 36, 75-93.

Olive D. Buhule, Jonathan O. Odwee and Leonard K. Atuhaire: Determinants of Recovery Time of Diabetic Patients in Uganda: Union for African population studies fifth African Population conference Arusha, Tanzania: 10-14 December 2007

Perry IJ, Wannamethee SG, Walker MK, Thompson AG, Whincup PH, Shaper AG. Prospective study of risk factors for development of non-insulin dependent diabetes in middle aged British men. British Medical Journal 1995;310:560-4.

Peter, J. Diggle, Patrick, H., Kung ,Yee Liang., Scott, L. Zeger (2000).Analysis of longitudinal data, Oxford University Press.

Rosella LC. A population based approach to diabetes mellitus risk prediction: Methodological advances and practical applications, Toronto, ON, Canada: University of Toronto, 2009.

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003) Semiparametric Regression, New York, Cambridge University Press.

Shaw JE, Sicree RA, Zimmet PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Res Clin Pract 2010; 87: 414

Shen, J., 2011. Additive Mixed Modeling of HIV Patient Outcomes Across Multiple Studies. University of California.

Stumvoll, M., Goldstein, B.J. and Haeften, T.W. (2005). Type 2 Diabetes: Principles of Pathogenesis and Therapy. Lancet, Vol 365: P 1333-1346.

Vanderpump MPJ, Tunbridge WMG, French JM, Appleton D, Bates D, Clark F, Grimley Evans J, Rodgers H, Tunbridge FT Young ET (1996). The incidence of diabetes mellitus in an English community - a twenty-year follow-up to the Whickham survey. Diabetic Medicine 13:741-

Vinicor F: Is diabetes a public-health disorder? Diabetes Care 17(Suppl. 1): 22-27, 1994.

Viswanathan M, McCarthy MI, Snehalatha C, et al. Familial Aggregation of type 2 (non-insulin-dependent) diabetes mellitus in South India; Absence of excess maternal transmission. Diabetic Medicine 1996;13:232-7.

Verbeke, G. and G. Molenberghs (2000). Linear Mixed Models for Longitudinal Data. Springer

Wang, Y. (1998), Smoothing Spline Models With Correlated Random Errors, Journal of the American Statistical Association, 93, 341-348 .

Wild S, Roglic G, Green A, Sicree R, King H (2004) Global prevalence of diabetes: Estimates for the year 2000 and projections for 2030. Diabetes Care 27: 1047-1053.

Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB Sr.. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. Arch Intern Med 2007;167:1068-74

Wondwosen, K. (2009-2010). Longitudinal Data Analysis in R

Worku, D., L.Hamza and K.Woldemichael, 2010. Patterns of diabetic complications at jimma university specialized hospital, southwest Ethiopia. Ethiop.J.Health Sci., 20:33-39.

World Health Organization. Definition, Diagnosis and Classifi cation of Diabetes Mellitus and its Complications. Report of a WHO Consultation. Part 1: Diagnosis and classification of diabetes mellitus. Geneva: WHO Department of Noncommunicable Disease Surveillance, 1999: 1-59. http://www.who.int

World Health Organization. Facts and figures about diabetes available at http ://www.who.int/diabetes/fa accessed on Dec 5, 2013

Zelalem et al., 2014. Longitudinal data analysis of fasting blood sugar level of adult diabetic patients in Jimma University Specialized hospital,LAP Lambert Academic Publishing 2014.

Zeger, S.L., and Diggle, P.J.(1994),"' Semi-parametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters,"' Biometrics, 50, 689-699.

Zhang, D., Lin, X., Raz, J. and Sowers, M. F. (1998), Semiparametric stochastic mixed models for longitudinal data, Journal of the American Statistical Association 93, 710719.

Appendix

Annex I: Tables

Table 6: LMM of FBS with quadratic time effect

Fixed Effects	Value	Std.Error	DF	t-value	p-value
Intercept	183.26731	3.375740	3866	54.28952	0.0000
time	-4.82118	0.607602	3866	-7.93477	0.0000
time^2	0.17895	0.025994	3866	6.88417	0.0000
Random Effects					
Variance Covariance					
Intercept	62.1577722				
time	7.0286500				
time^2	0.2428374				
residual	69.1773930				
AIC	BIC	loglik			
50848.65	50912.54	-25414.32			

Table 7: LMM of FBS with quadratic time effect when all covariates are included

Fixed Effects	Value	Std.Error	DF	t-value	p-value
(Intercept)	303.48072	20.975054	3857	14.468650	0.0000
age	0.39557	0.222586	3857	1.777140	0.0756
time	-4.84430	1.964228	3857	-2.466264	0.0137
sex	-3.24162	5.984220	530	-0.541694	0.5883
DT	-11.19119	10.401324	530	-1.075939	0.2824
famhist	-12.21103	10.174176	530	-1.200199	0.2306
weight	-1.88200	0.222675	3857	-8.451752	0.0000
SBP	-0.34123	0.139545	3857	-2.445304	0.0145
DBP	0.40295	0.215407	3857	1.870640	0.0615
time2	0.16168	0.025959	3857	6.228409	0.0000
age:time	-0.02222	0.017678	3857	-1.256820	0.2089
time:sex	-0.29027	0.473005	3857	-0.613672	0.5395
time:DT	0.03475	1.202816	3857	0.028893	0.9770
time:famhist	-0.04611	1.197413	3857	-0.038509	0.9693
time:weight	0.02655	0.017893	3857	1.483723	0.1380
time:SBP	0.01960	0.012221	3857	1.603580	0.1089
time:DBP	-0.03034	0.020148	3857	-1.506032	0.1321
Random Effects					
Variance Covariance					
Intercept	57.997822				
time	6.703831				
time2	0.234987				
residual	68.719672				
AIC	BIC	loglik			
50776.24	50929.51	-25364.12			

Table 8: SPMM of FBS with quadratic time effect when smoothing time in month is included

Fixed Effects	Value	Std.Error	DF	t-value	p-value
X(Intercept)	165.45085	2.34957	3869	70.41751	0.0000
X(time)	-11.36977	41.25850	3869	-0.275557	0.7829
Random Effects					
Variance Covariance					
Intercept	56.84406				
time	2.581283				
time2	0.000247899				
residual	69.84482				
Approximate significance of smooth terms					
	edf	F	p-value		
s(time)	8.386	20.98	$\mathbf{< 2e-16}$ ***		
AIC	BIC	loglik			
50794.64	50858.54	-25387.32			

Table 9: SPMM of FBS with quadratic time effect when time-independent covariate and smoothing time in month is included

Fixed Effects	Value	Std.Error	DF	t-value	p-value
Xage	-0.081	0.1789	3867	-0.453555	0.6502
XDT	-17.283	9.0194	531	-1.916198	0.0559
Xsex	-8.961	4.9010	531	-1.828412	0.0680
Xfamhist	15.317	-8.8340	531	-1.733840	0.0835
Xtime	18.352	-1.8274	3867	-10.042721	0.0000
Xtime2	2.841	0.1513	3867	18.774764	0.0000
Xs(time2)Fx1	-7240.235	358.4224	3867	-20.200285	0.0000
Random Effects					
Variance Covariance					
Intercept	56.81558				
time	2.591051				
time2	0.0000091034				
residual	69.57985				
Approximate significance of smooth terms					
	edf	F	p-value		
s(time)	9.127	51.52	$< 2e-16$ ***		
AIC	BIC	loglik			
50796.77	50892.6	-253883.38			

Table 10: SPMM of FBS with quadratic time effect when time-invariant covariate with time interaction and smoothing time in month was included

Fixed Effects	Value	Std.Error	DF	t-value	p-value
Xage	0.030	0.2304	3863	0.129133	0.8973
XDT	-14.201	10.8056	531	-1.314247	0.1893
Xsex	-7.619	6.2985	531	-1.209617	0.2270
Xfamhist	-12.562	10.5707	531	-1.188343	0.2352
Xtime	-16.883	2.2242	3863	-7.590662	0.0000
Xtime:age	-0.013	0.0175	3863	-0.748560	0.4542
Xtime:DT	-0.828	1.2075	3863	-0.685839	0.4929
Xtime:sex	-0.156	0.4840	3863	-0.321981	0.7475
Xtime:famhist	-0.763	1.2004	3863	-0.635925	0.5249
Xtime2	2.769	3863	17.028223	0.0000	
	0.1626				
Xs(time2)Fx1	-7015.518	402.9918	3863	-17.408586	0.0000
Random Effects					
Variance Covariance					
Intercept	57.01418				
time	2.648456				
time2	0.00009545323				
residual	69.57126				
Approximate significance of smooth terms					
	edf	F	p-value		
s(time)	9.128	35.83	$< \mathbf{2e^{-16}}$ ***		
AIC	BIC	loglik			
50806.27	50927.64	-25384.13			

Table 11: SPMM of FBS with quadratic time effect when time-varying covariate with time interaction and smoothing time in month was included.

Fixed Effects	Value	Std.Error	DF	t-value	p-value
XSBP	-0.111	0.0802	3865	-1.380823	0.1674
Xtime	-19.188	2.0497	3865	-9.361306	0.0000
Xweight	-1.839	0.2155	3865	-8.534296	0.0000
Xtime2	3.635	0.1739	3865	20.902568	0.0000
Xtime:weight	0.028	0.0164	3865	1.718602	0.0858
Xs(time2)Fx1	-9756.019	455.2741	3865	-21.428889	0.0000

Random Effects			
Variance Covariance			
Intercept	53.298396334		
time	2.360792105		
time2	0.001688503		
residual	69.182483305		
Approximate significance of smooth terms			
	edf	F	p-value
s(time)	9.143	53.05	$< \mathbf{2e^{-16}}$ ***
AIC	BIC	loglik	
50727.57	50817.02	-25349.79	