Modeling Time-to-Recovery of Adult Diabetic Patients: A Case Study of Jimma University Specialized Hospital

By: Abiyot Negash

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

January, 2014

Jimma, Ethiopia

Modeling Time-to-recovery of adult diabetic patients: A Case Study of Jimma University Specialized Hospital

MSc Thesis

By:

Abiyot Negash

Advisor: Dr. Yehenew Getachew (PhD)

Co-advisor: Mr. Belay Birlie (PhD candidate)

January, 2014

Jimma, Ethiopia

JIMMA UNIVERSITY

SCHOOL OF GRADUATE STUDIES, DEPARTMENT OF STATISTICS

As thesis research advisor, we hereby certify that we have read and evaluated the thesis prepared by **Abiyot Negash Terefe** under our guidance, which is entitled "*Modeling Time-to-recovery of adult diabetic patients: A Comparison of Cox PH and Shared Frailty Model*". We recommend that the thesis be submitted as it fulfills the requirements for the degree of Master of Science (MSc) in Biostatistics.

Dr. Yehenew Getachew		January 15, 2014
Advisor	Signature	Date
Mr. Belay Birlie		
Co-advisor	Signature	Date

As the members of the board of examiners of MSc thesis open defense examination of **Abiyot Negash Terefe**, 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
Main Advisor	Signature
Co-advisor	Signature
Internal Examiner	Signature
External Examiner	Signature

i

STATEMENT OF AUTHOR

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 in the partial fulfillment for the Degree of Master of Science in Biostatistics. The thesis can be deposited 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.

Abiyot Negash Terefe

Date: _____

Signature: _____

Jimma University, Jimma

ACKNOWLEDGMENT

First of all, I would like to express my deepest gratitude to my main advisor Dr. Yehenew Getachew and Co-advisor Mr. Belay Birlie for their excellent guidance, instructions and support. This thesis would not be possible without their advice, suggestions and encouragement.

Thanks to my friends for sharing with me the ups and downs in all these years in post graduate school.

My grateful thanks to my parents, whose love, trust, encouragement, comfort and support always surround.

Also, I really appreciate the precious learning opportunity, environment and budget provided by the Jimma University. I am grateful to all the faculty, department, and friends who ever helped.

Lastly not from least, I would like to thank Jimma University Specialized Hospital (JUSH) Diabetic Patient Clinic for their great help in many ways during data collection.

ABSTRACT

Diabetes is a group of diseases marked by high or low level of glucose resulting from defects in insulin production, insulin action or both. There are two main types of diabetes, namely *type I* and Type II diabetes. The objective of this thesis is to model time-to-first recovery of adult diabetic patients using Cox PH and shared frailty models. A retrospective data was obtained from JUSH diabetic patient clinic. All diabetic patients ≥ 18 years of age and who are under treatments in between September 2010 and August 2013 are included in the study. Time of fasting blood sugar level to reach the first normal range, 70-130 mg/dl, of blood since time of treatment or intervention were the response variable. Due to the impact of residential places and unmeasured shared similarities in a cluster, district (Woreda) is used as a random effect (frailty) term in the survival models. In this thesis, Cox PH and shared gamma frailty models were used. The AIC was used to compare the performance of the different models. First, inseparable diabetic mellitus (DM) was analyzed to identify whether diabetic types significantly influencing recovery time of DM. Second, separate types of DM are analyzed to identify factors influencing recovery time of these types of DM. The median recovery time of type-I and type-II diabetic patients were between 2 and 4 months respectively. The minimum and maximum recovery time of type-I diabetic are 1 and 6 months, respectively, whereas for type-II diabetic mini-max recovery time is found to be 1 and 31 months, respectively. Types of diabetic, bodyweight at baseline, fasting blood sugar at baseline, sex and age of patients are significantly associated with time to first recovery of diabetic patients. These variables are important factors that should be considered during the selection phase a treatment (combination of treatments) for diabetes. Moreover, Cox PH with gamma frailty model have resulted in a minimum AIC as compared to Cox PH model without frailty term in the model. This might be due to the shared environmental and residential factors. Hence, Cox PH model with gamma frailty provide a suitable choice for modeling time to first recovery of DM as compared to Cox PH without frailty term in the model.

Key Words: *frailty, heterogeneity, gamma distribution, AIC, penalized partial likelihood, Cox-Snell, deviance residuals.*

LIST OF ACRONYMS

AIC:	Akaike's Information Criteria
DBP:	Diastolic Blood Pressure
DM:	Diabetic Mellitus
FBS:	Fasting Blood Sugar
HR:	Hazard Ratio
IDDM:	Insulin- Dependent Diabetes Mellitus
JUSH:	Jimma University Specialized Hospital
KM:	Kaplan-Meier
LR:	Likelihood Ratio
NIDDM:	Non-Insulin- Dependent Diabetes Mellitus
PH:	Proportional Hazard
PLH:	Partial Likelihood
PPL:	Penalized Partial Likelihood
SBP:	Systolic Blood Pressure
WHO:	World Health Organization

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1. Background of the Study

Diabetes is a group of diseases marked by high or low level of glucose resulting from defects in insulin production, insulin action or both. It can lead to serious complication and premature death but steps to control the disease and lowers the risk of complications does exist. Insulin replacement is required for survival. The intensive-therapy regimen was designed to achieve blood glucose values as close to the normal range as possible with three or more daily insulin injections or treatment with an insulin pump. Conventional therapy consisted of one or two insulin injections per day (Leong W.Y., 2007).

Diabetes is divided into two main different types which are type I and Type II diabetes. The former was called insulin dependent diabetes mellitus (IDDM) or juvenile-onset diabetes while the latter was called non-insulin- dependent diabetes mellitus (NIDDM) or adult-onset diabetes (National Diabetes Fact Sheet United States, 2005). Diabetes is becoming one of the rapidly increasing non-communicable diseases and an important public health problem all over the world. Connor and Boulton, (1989), notes that the main factors which lead to the cause of diabetes mellitus are hereditary (genetics) and environmental. Type-I diabetes which develops most frequently in children and adolescents can be caused by viruses that have injured the pancreas and destruction of insulin making cells by the body's immune system. Also, a family history of diabetes is a risk factor of type-I diabetes. Type-II is a common and serious global health problem which is associated with rapid cultural and social changes, ageing populations, increasing urbanization, dietary changes, reduced physical activity and other unhealthy, lifestyle and behavioral patterns.

In 2011, 14.7 million adults in the Africa are estimated to have diabetes, with a regional prevalence of 3.8%. The highest prevalence of diabetes in the Africa is in the island of Reunion (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 over half of the total number in Africa (Diabetes atlas 5th edition).

382 million people have diabetes in 2013; by 2035 this will rise to 592 million. The number of people with type 2 diabetes is increasing in every country. 80% of people with diabetes live in low- and middle-income countries. The greatest number of people with diabetes is between 40 and 59 years of age. Ethiopia is one of the developing countries where by the prevalence is increasing time to time. The prevalence of diabetes in Ethiopia for 20-79 age groups in 2013 is 4.89, Uganda 4.81 (IDF, 2013).

In managing diabetic mellitus (DM) proper self-care practice and optimal glucose control is an essential cornerstone in achieving successful health outcomes. DM is a life-long challenge that requires behavioral change and adequate self-care practices for better glycaemic control. In the absence of appropriate self-care practice, the desired therapy targets are difficult, or even impossible to achieve. Glucose control is almost entirely in the hands of the patient who lives with this condition. The patient's motivation to eat, exercise, take medication, test glucose levels and maintain a healthy body weight all play a significant role in the management of DM. If left untreated, it can lead to heart disease, stroke, blindness, and kidney failure (Diabetes in the UK, 2011/12). Diabetes is a common health condition. The chances of developing it may depend on a mix of genes, lifestyle and environmental factors. Environmental factors that contribute to beta cell destruction and genes regulating immune response are involved. Numerous environmental events trigger the autoimmune process in genetically susceptible individuals. There are environmental factors which have a link with DM like, chemical compounds (rodenticides, heavy metals virus, rarely and exposure to bovine milk proteins), and physical factors (penetrative short-wave length rage) etc. People with underlying medical conditions such as diabetes are more vulnerable to the adverse health impacts of climate change. In hotter temperatures, dehydration and heatstroke increases morbidity and mortality in people with diabetes. People with diabetes are predisposed to cardiovascular events during heat waves and higher mortality from heart attack on days of high air pollution (Dereje A., 2005).

A diabetic person has to eat a diet low in fat, high in fiber, and with plenty of starchy foods, fruits and vegetables and should exercise regularly. This reduces the weight, which will help reduce the blood glucose/sugar and risk of having heart attack or stroke. The blood sugar

concentration or blood glucose level is the amount of glucose (sugar) present in the blood of a human or animal. The body naturally tightly regulates blood glucose levels as a part of metabolic homeostasis. The mean normal blood glucose level in humans is about 5.5mM (5.5 mmol/l or 100 mg/dl, i.e. milligrams/deciliter); however, this level fluctuates throughout the day. Glucose levels are usually lowest in the morning, before the first meal of the day (termed "the fasting level"), and rise after meals for an hour or two by a few millimolar. Blood sugar levels outside the normal range may be an indicator of a medical condition. A persistently high level is referred to as *hyperglycemia*; low levels are referred to as *hypoglycemia*. Diabetes mellitus is characterized by persistent hyperglycemia from any of several causes, and is the most prominent disease related to failure of blood sugar regulation. Intake of alcohol causes an initial surge in blood sugar, and later tends to cause levels to fall. Also, certain drugs can increase or decrease glucose levels. Normally, the human body keeps its blood glucose level very stable (between 4mml-7.5mmol/l or 70-130 mg/dl). The body has various systems (regulated by hormones such as insulin and glucagon) to keep the blood glucose level in this range. These systems fail in people with diabetes.

The world health organization (WHO) publishes standards of medical care yearly to promote the importance of achieving optimal glycaemic control. Diabetes was classified according to WHO recommendations. Recommended blood sugar for people with diabetes (according to the WHO) before meals plasma glucose levels within a narrow range 70-130 mg/dl (milligram per deciliter). Blood glucose is balanced between endogenous appearance from the liver (through glycogenolysis and gluconeogenesis) and kidneys, exogenous appearance from the intestines (following a meal), and utilization of glucose by all tissues. Two gross metabolic conditions exist. When fasting, the body relies primarily on glucose stored in the form of glycogen and fatty acids stored in the form of triglycerides to fuel its metabolic needs. After a meal, glucose absorbed from the gut is used to replenish glycogen and fat stores diminished while fasting (Hipszer, B.R, 2001).

The statistical analysis of survival data is an important topic in many areas, including medicine, epidemiology, biology, demography, economics, engineering and other fields. A variety of techniques have been developed to analyse survival data. A common approach to the analysis of survival data is based on the assumption that the study population is homogeneous. That is,

conditional on the covariates, every individual has the same risk of experiencing an event such as death or disease recurrence (Ulviya A., 2013). The event times of individuals in the population, conditional on the observed covariates, are assumed to be independent. However, this cannot be assumed in all applications as many applications require heterogeneous sample, i.e. individuals with different risks and hazards. In practice, there may be an association between the events times of some subgroups of the population since the individuals of these groups share a common trait that cannot be observed. For example, there may be an association in the times to events of cancer or cardiovascular diseases between siblings or married couples, even occurrence of nonlethal diseases within the same individual. Though individuals may look identical in some aspects, they may differ in unmeasured ways. In applications of survival analysis, usually only a few covariates such as age, sex, severity of disease or laboratory data are known. It is known that there are many other factors that can influence survival, including health status, life style, smoking, occupation and genetic risk factors. These factors are unknown and cannot be included in the analysis.

Beard (1959), Vaupel et al. (1979), and Lancaster (1979) suggested a random effects model in order to account for the unobserved heterogeneity due to unobserved covariates. Beard (1959) used the term longevity factor to improve the effect of mortality models in populations. Vaupel et al. (1979) introduced the term frailty in order to account for unobserved heterogeneity, random effects, and association in univariate survival models. He introduced this concept of frailty to biostatistics by applying it on population mortality data. Lancaster (1979) introduced the model to the literature of economics and the model is called the mixed proportional hazards model. The concept, however, goes back to work of Greenwood and Yule on "accident proneness" in 1920. Clayton (1978) discussed the applications of the model to multivariable survival data in his seminar paper on chronic disease incidence in families. Frailty models account for unobserved heterogeneity that occurs because some observations are more prone to failure, and therefore more "frail" than others in a data set. Therefore, the objective is to introduce an additional parameter to the hazard rate that accounts for the random frailties. These frailties can be specific to groups, and are referred to as shared frailty. The overall aim of this thesis is modeling of time-to-first recovery of adult diabetic patients from Jimma University Specialized Hospital (JUSH) using various survival models.

1.2. Statement of Problem

Diabetic mellitus (DM) is a life-long challenge that requires behavioral change and adequate self-care practices for better glycaemic control. For this reason, people living with DM are in need of identifying risk factors and prognostic factor for their survival to control and care themselves.

In the literature, there are many studies on the field of diabetic often researchers examine the effects of covariates on patients using logistic regression (Endalew H. et al (2012)) and chi-square (Charlton, et al (2005), Endalew H. et al (2012)) models. Such data can also be more explored using survival models, such as the classical Cox PH and frailty models. However, to our best knowledge there is a limited work in this line.

Inference for Cox PH model (Cox, 1972) was developed under the assumption that the observations are statistically independent and the population they come from is assumed to be homogeneous with respect to failure. However, this assumption may be violated. Thus, in many epidemiological studies, failure times are clustered into groups such as families or geographical units; some unmeasured characteristics shared by the members of that cluster, such as genetic information or common environmental exposures could influence time to the studied event. In a different context, correlated data may come from recurrent events, i.e. events which occur several times within the same subject during the period of observation. Ignoring the existence of heterogeneity will produce incorrect estimation of parameters and their standard errors in survival analysis. According to Keyfitz and Littman (1979), ignoring heterogeneity overestimates life expectancy based on their study on estimating life expectancy in a heterogeneous population. Lancaster (1990) showed that when heterogeneity is ignored, it caused underestimation of covariate effects in his study of unemployment rates. Henderson and Oman (1999) showed that ignoring frailty leads to regression coefficient estimates biased towards zero by an amount depending on the distribution and the variability of the frailty terms. For such situations, one approach accounting for correlation is to incorporate an additive or multiplicative random effect for each cluster, resulting in a frailty model. Random effect or frailty model attempts to account for the existence of unmeasured attributes (such as genotype, environment and geographical location) that introduce heterogeneity into the study population.

Not taking into account the unobserved frailty will thus under/overestimate the model parameters.

Therefore, in this thesis, we are interested to address the following interesting research questions:

- Which type of DM takes long time to recover to normal blood sugar level;
- Is there heterogeneity among districts with respect to time to recovery of adult diabetic patients;
- What are the covariates influencing the time to recovery for each type of diabetic; and
- Which type of survival models, Cox-PH or shared gamma frailty models, predicts well the recovery time of adult diabetic patients.

1.3. Objective

1.3.1. General Objective

The general objective of the study is modeling of time-to-first recovery of adult diabetic patients from Jimma University Specialized Hospital (JUSH) using Cox-PH and shared gamma frailty Models.

1.3.2. Specific Objective

The specific objectives of the study are;

- to identify the types of diabetic patients that mostly recover blood sugar level into normal range and model them separately;
- to investigate important factors or covariates that are significantly associated with time to first recovery from each type of diabetic; and
- to compare Cox-PH and frailty models and thereby to show the benefit of taking into account the clustering of subjects within districts using shared gamma frailty model.

1.4. Significance of the Study

The results of this study will be very useful in the development of an effective diabetic care and anti-diabetic therapy (ADT) patient monitoring system. Specifically, this study will be helpful to:

- develop implementation plan to deliver the national institute for health and clinical excellence (NICE) quality standards and the national standards framework (NSF) outcomes, for the sake of society, the NHS, and above all for people with diabetes and those at risk of developing diabetes to monitor, care and prevent patients under treatment follow up.
- give essential care standards to reduce complications, costs, diabetic related illness and premature death.

2. LITERATURE REVIEW

2.1. Types of Diabetes

2.1.1. Type-I Diabetes

Type-I diabetes is an autoimmune disease. An autoimmune disease results when the body's system for fighting infection (the immune system) turns against a part of the body. In diabetes, the immune system attacks the insulin-producing beta cells in the pancreas and destroys them. The pancreas then produces little or no insulin. Someone with type-I diabetes needs to take insulin daily to live. At present, scientists do not know exactly what causes the body's immune system to attack the beta cells, but they believe that both genetic factors and environmental factors, possibly viruses, are involved. Type-I diabetes develops most often in children and young adults, but the disorder can appear at any age. Symptoms of type-I diabetes usually develop over a short period, although beta cell destruction can begin years earlier (Lancet, 2010). Undiagnosed or untreated type-I diabetes can make people lose weight and increase blood pressure. In type-I diabetes, the body stops producing the hormone insulin, which is needed to use glucose, the main type of sugar in the blood. Glucose comes from the foods we eat and is the major source of energy needed to fuel the body's functions. In type-I diabetes, the body can't use glucose properly, so flushes the glucose (and the calories) out of the body in urine. As a result, kids who develop type-I diabetes can lose weight and increase blood pressure despite having a normal or increased appetite. Once they're diagnosed and treated, their blood sugar usually returns to normal. Symptoms include increased thirst and urination, constant hunger, weight loss, blurred vision, extreme fatigue and increase blood pressure. If not diagnosed and treated with insulin, a person can lapse into a life-threatening diabetic coma, also known as diabetic ketoacidosis (James N. et al, 2002).

Emmanuel, et al (1991-2005) studied the Incidences, Treatments, Outcomes, and Sex Effect on Survival in Patients with End Stage Renal Disease (ESRD) by Diabetes Status in Australia and New Zealand. The study included 1,284 type-I diabetic (4.5%), 8,560 type-II diabetic (30.0%), and 18,704 non-diabetic (65.5%) patients. The incidence rate of ESRD with type-II diabetes increased markedly over time (+ 10.2% annually, $P< 0.0001$). In patients aged <70 years, rates of renal transplantation in type-I diabetic, type-II diabetic, and non-diabetic patients were 41.8,

6.5 ($P < 0.0001$ vs. other patients), and 40.9% (P=0.56 vs. type-I diabetic patients), respectively. Compared with non-diabetic patients, the adjusted hazard ratio (HR) for death was 1.64 ($P < 0.0001$) in type-I diabetes and 1.13 ($P< 0.0001$) in type-II diabetes. Survival rates per 5-year period improved by 6% in type-I diabetic patients ($P = 0.36$), by 9% in type-II diabetic patients ($P< 0.0001$), and by 5% in non-diabetic patients ($P< 0.001$). In type-II diabetic patients aged ≥ 60 years, the adjusted HR for death in women versus men was 1.19 ($P = 0.0003$).

Hypothesis Surveys in northern Ethiopia have demonstrated that apparent type-I diabetes occurs more frequently than elsewhere in Africa and, indeed, in other parts of the world. They investigated in detail a cohort of diabetic patients from this region to clarify the nature of this type of diabetes. Methods all patients attending the diabetic clinic at Mekelle Hospital in the Tigray region of northern Ethiopia were investigated over a 6 week period. Clinical, demographic and anthropometric data were collected, as well as measurements of HbA1c, fasting lipid profile, fasting serum C-peptide and serum markers of beta cell autoimmunity, i.e. islet antigen-2 and GAD antibodies (GADA). Results of 105 patients seen, 69 (66%) were on insulin treatment and had been from or close to diagnosis. Their median age and diabetes duration were 30 and 5 years, respectively, with a male excess of 2:1. Median BMI was 20.6 kg/m2. Despite these clinical characteristics suggestive of type-I diabetes, only 42 of 69 (61%) patients were C-peptide- negative and 35% GADA-positive. Overall, 38 (36%) of the total group (n=105) had immunological or C-peptide characteristics inconsistent with typical type-I or type-II diabetes. The clinical characteristics, local prevalence of under-nutrition, and GADA and C-peptide heterogeneity suggest a malnutrition-related form of diabetes.

2.1.2. Type-II Diabetes

This form of diabetes usually develops in adults age 40 and older and is most common in adults over age 55. About 80 percent of people with type-II diabetes are overweight and elevates blood pressure. Type-II diabetes is often part of a metabolic syndrome that includes obesity, elevated blood pressure, and high levels of blood lipids. Unfortunately, as more children become overweight and increase their blood pressure, type-II diabetes is becoming more common in young people. When type-II diabetes is diagnosed, the pancreas is usually producing enough insulin, but, for unknown reasons, the body cannot use the insulin effectively, a condition called

insulin resistance. After several years, insulin production decreases. The result is the same as for type-I diabetes glucose builds up in the blood and the body cannot make efficient use of its main source of fuel. The symptoms of type-II diabetes develop gradually. They are not as sudden in onset as in type-II diabetes. Some people have no symptoms. Symptoms may include fatigue or nausea, frequent urination, unusual thirst, weight loss, blurred vision, frequent infections, and slow healing of wounds or sores.

More than 85 per cent of children and young people over the age of 12 have blood glucose levels higher than recommended targets. The percentage of children and young people achieving the HbA1c target of <7.5 per cent varies from 1.6 per cent to 37.2 per cent. 15.5 percent of children and young people have had one episode of Diabetic Ketoacidosis (DKA)* in the last five years, and 10.4 per cent of children and young people have had two or more episodes of DKA in the last five years. DKA is a critical, life-threatening condition caused by prolonged raised blood glucose levels (hyperglycemia) that requires immediate medical attention (State of the nation, 2012 in England).

Diabetes in the UK 2011/2012, most health experts agree that the UK is facing a huge increase in the number of people with diabetes. Since 1996 the number of people diagnosed with diabetes has increased from 1.4 million to 2.9 million. By 2025 it is estimated that five million people will have diabetes. Most of these cases will be Type-II diabetes, because of our ageing population and rapidly rising numbers of overweight and obese people. The figures are alarming and confirm that diabetes is one of the biggest health challenges facing the UK today. If we are to curb this growing health crisis and see a reduction in the number of people dying from diabetes and its complications, we need to increase awareness of the risks, bring about wholesale changes in lifestyle, improve self-management among people with diabetes and improve access to integrated diabetes care services. More men than women have diagnosed diabetes; 56 per cent compared with 44 per cent in those with Type-I diabetes and 55 percent compared with 45 percent in those with Type-II diabetes. This ratio is relatively unchanged from 2001. The estimated diabetes prevalence worldwide for 2011 was 366 million and it is expected to affect 552 million people by 2030.

An estimated 280 Australians develop diabetes every day. The 2005 Australian AusDiab Follow-up Study (Australian Diabetes, Obesity and Lifestyle Study) showed that 1.7 million Australians have diabetes but that up to half of the cases of type-II diabetes remain undiagnosed. By 2031 it is estimated that 3.3 million Australians will have type-II diabetes (Vos et al., 2004). A reduction in the prevalence of type-II diabetes will not only result in cost savings in the health budget, but increased participation and productivity in the workforce and, most importantly, better health outcomes and quality of life for Australians. There is no doubt diabetes is a serious health crisis but it's not all bad news. Up to 60% of cases of type-II can be prevented and we know that good blood glucose control and maintaining a healthy lifestyle can significantly improve the complications associated with diabetes.

Chaudhry, Gannon, Nuttall, (2006) conducted a thesis on stability of body weight in type-II diabetes**.** Data were obtained from the records of 205 adult men who have attended a diabetes clinic for ≥5 years. Their weight and glycohemoglobin at the last visit were compared with the initial visit data. The subjects were categorized according to treatment modalities. The mean follow-up was 9.4 years (range 5–23). For the group as a whole, the mean increase in body weight was 0.23 ± 0.2 kg/year. BMI or initial age had little effect on the rate of weight gain. Treatment regimen used did have an effect on weight change. In subjects treated with insulin, with or without oral agents, body weight increased at a rate of 0.44 ± 0.1 kg/year. In subjects treated with metformin or metformin and a sulfonylurea, there was a mean loss in weight, i.e., 0.24 ± 0.09 kg/year, and with sulfonylureas alone weight increased by 0.42 ± 0.2 kg/year. Concluded that the men treated with insulin alone or insulin combined with oral agents gained weight at a rate comparable with that reported for the general population, i.e., the weight gain was not extraordinary. Metformin treatment resulted in a modest loss of weight.

Gebregziabher, M., et al (2010) used effect of Trajectories of Glycemic Control on Mortality in Type-II Diabetes, Multiple studies have established that poor glycemic control as measured by hemoglobinA1c (HbA1c) level is associated with increased mortality in persons with type-II diabetes. Recently, in the Norfolk, United Kingdom, component of the European Prospective Investigation into Cancer and Nutrition (EPIC-Norfolk), Khaw et al. demonstrated that HbA1c was continuously related to subsequent all-cause, cardiovascular, and ischemic heart disease mortality through the whole population distribution, with the lowest rates being seen among

persons with HbA1c concentrations below 5%. In the EPIC-Norfolk study, an HbA1c level of 5% was used as the reference category, and there was a significant linear relation between HbA1c and risk of death, such that there was an almost 3-fold increased risk of death in men with HbA1c concentrations greater than or equal to 7%.

2.2. Adverse Health Outcomes and Increased Diabetes Risk

Diabetes and climate change are directly and indirectly interconnected. Direct connections refer diabetes and climate change adversely impact upon each other. The indirect connections refer to the common global vectors and pathways that are fuelling both these health and development disasters. People with underlying medical conditions such as diabetes are more vulnerable to the adverse health impacts of climate change. In hotter temperatures, dehydration and heatstroke increases morbidity and mortality in people with diabetes. People with diabetes are predisposed to cardiovascular events during heat waves and higher mortality from heart attack on days of high air pollution (IDF, 2012).

2.3. Cox PH and Frailty Models

Cox PH model keeps the baseline hazard as an arbitrary, unspecified, and nonnegative function of time. It is the most popular and commonly used model by researchers in medical sciences mainly because of its simplicity, and not being based on any assumptions about the survival distribution (Therneau T, & Grambsch P., 2000).

Cox's proportional hazard model (1972) augmented to include time invariant unobserved person specific variables is now widely used in duration analysis (see, e.g., Tuma(1976), Tuma et al. (1979), Lancaster (1979), Flinn and Heckman (1982) and the papers cited in Heckman and Singer (1982)). For single spell duration data, the only estimator of this model that controls for unobserved person specific heterogeneity is the random effect estimator. With this estimator, the analysis estimates the distribution function of unobservable and the parameters of the distribution model conditional on the unobserved variables. Heckman and Singer (1982) and Trussel and Richards (1983) finds that estimates obtained from duration models are very sensitive to arbitrary choices about the functional forms of the distribution of unobservable and the conditional duration distribution.

One alternative model that does allow for dependence between related individuals is the frailty model, which has been studied by a number of authors over the past years, including Clayton (1978), Clayton and Cuzick (1985), Hougaard (1986), Andersen et al. (1993, Chapter-IX), and Hougaard, Myglegaard, and Borch-Johnsen (1994). Another application of the frailty model is to interpret the frailty as modeling the effect of unobserved covariates (e.g., Vaupel, Manton, and Stallard, 1979; Hougaard, 1984). Maximum likelihood estimation in the semiparametric shared frailty model (with gamma- distributed frailties) may be performed using the EM algorithm as suggested by Gill (1985) and further discussed by Nielsen et.al. (1992) and Klein (1992). More recently, interest has focused on a model where the frailties for related individuals need not be shared among them but rather are correlated (Pickles et al., 1994; Yashin, Vaupel, and lachine, 1995). This model has the advantage that separate parameters describe association and unobserved heterogeneity. Also in this model, the EM algorithm may be used for maximum likelihood estimation (Petersen, Andersen, and Gill, 1996).

Andersen, et al. (2013) employed Estimation of variance in Cox's Regression Model with Shared Gamma Frailties. The Cox regression model with a shared frailty factor allows for unobserved heterogeneity or statistical dependence between the observed survival times. Estimation in this model when the frailties are assumed to follow a gamma distribution is reviewed and addressed the problem of obtaining variance estimates for regression coefficients frailty parameter and cumulative baseline hazards using the observed non-parametric information matrix. Comparing the models with and without frailties, concluded that both the estimates and their estimated standard errors are smaller in the models without frailty.

Olive, D., et al (2007) employed a frailty model to study the determinants of recovery time of diabetic patients from three hospitals in Uganda. It was found that Biguanides work better than Insulin, diet and exercise and Sulphonylureas. Disease duration did not have a significant effect on time to remission. It was concluded that duration of the disease does not have any effect on the effectiveness of the interventions. Time to remission was found to decrease with increase in body mass index and age. Males tend to recover faster than the female and the less or non-educated controlled the disease better than the educated ones. It is concluded that Biguanides are better interventions than Insulin, diet and exercise and Sulphonylureas, frailty models are better to model the recovery time of DM.

Semi-parametric inference for frailty models was introduced by Klein et al. (1992) and Nielsen et al. (1992) and as suggested by Gill (1985), they used an EM algorithm applied to the Cox partial likelihood. Hastie and Tibshirani (1993) proposed a general model and suggested estimation through penalized partial likelihood. Therneau and Grambsch (2000) noted a link between the gamma frailty model and a penalized partial likelihood. In the approach of the present paper, we penalize the hazard function (s) while Therneau and Grambsch (2000) penalize the frailties. In Cox and parametric models, hazard function may depend on unknown or non measurable factors which can cause the regression coefficients estimated from such models to be biased. In consequence, in order to overcome the problem and better model survival of patients, the frailty models were introduced. In fact, these models are used to explain the random variation of survival function due to unknown risk factors, such as genetic factors and numerous environmental factors.

Ulviya, A., (2013) employed frailty models for modeling heterogeneity. Suggested as a Semiparametric regression model is an important way to handle heterogeneity. Regression models take lifetime as the dependent variable and explanatory variables as regressors. Sometimes these models may not provide adequate fit to the data. One of the reasons is due to omission of important covariates. Several methods have been developed to model the frailty in survival data during recent years. Used AIC to compare the performance of the models. It is concluded that the generalization of the Cox proportional hazards model (Cox, 1972) is the best and widely applied model that allows for the random effect.

3. DATA AND METHODOLOGY

3.1. Data Source

For this study, longitudinal retrospective cohort follow up (retrospective cohort design) of adult diabetic patients data is collected from Jimma University Specialized Hospital Diabetic Patient Clinic located in southwest of Ethiopia. The data is extracted from the patient's chart which contains epidemiological, laboratory and clinical information of all diabetic patients under insulin treatment follow-up.

3.2. Study Population

A total of 1930 diabetic patients are on active follow up. All diabetic patients greater than or equal to 18 years old and placed under treatments that have followed between September 2010 and August 2013 (three years data) were included. The data for this study consists of 544 individuals. Patents' follow up time was one, two or three months gap according to the order of the doctor and the data was collected from patients' medical follow up card by assigning an identification number per individual by health workers in the chronic follow up clinic, which helps to find the patients profile easily during his/her next visit time. Times of fasting blood sugar level until it reaches the first normal range (70-130mg/dl according to WHO association) were used.

3.3. Variables

3.3.1. Dependent Variable

The outcome variable considered in this study is the time to first recovery of diabetic patients until it reaches normal fasting (before meal) blood sugar level in the follow up period. Time to first recovery means the time until patients comes to normal fasting blood sugar level for first time in the follow up period according to WHO scale (70-130mg/dl). Right censoring is considered when patient is not recovered once between the study time, transferred to other hospital and death before first recovery to normal blood sugar level.

3.3.2. Independent Variables

Covariates are described with their values or codes in Table 1 as follows.

Table 1. Study Covariates

No	Variable	Description	Values/Codes
1	Sex	Sex	0=female, 1=Male
2	Age	Age in years	0=18-29; 1=30-44; 2= 45-59; 3=60-74 and 4= > 74
3	Diabtype	Diabetic types	0=type 1, 1= type2
4	Famhist	Family history	0=no, 1= yes
5	SBP	Systolic blood pressure in Mm/Hg	0= <110 (below), 1=110-130 (normal), 2= >130(high)
6	DBP	Diastolic blood pressure in Mm/Hg	0= <60(below), 1= 60-80(normal), 2= >80(high)
7	BW	Body weight in kg	Continues variable
8	FBS	Fasting blood sugar in mg/dl	Continues variable

3.4. Methods of Data Analysis

3.4.1. Survival Analysis

Survival analysis examines and models the time it takes for events to occur. Although there are well known methods for estimating unconditional survival distributions, most interesting survival modeling examines the relationship between survival and one or more predictors, usually termed covariates in the survival-analysis literature.

Suppose there are n subjects followed over a certain time interval $[0,\tau)$. The i^{th} subject at times $\{t_i, i=1,2,...,n\}$ and a (possibly censored) survival time t_i to a certain endpoint. Let T_i denotes the response for the i^{th} subject (time to event), C_i denote the censoring time for the i^{th} subject, δ_i denote the event indicator

$$\delta_i = \begin{cases} 1 \text{ if the event was observed } (T_i \le C_i) \\ 0 \text{ if the response was censored } (T_i > C_i) \end{cases}$$

The observed response $y_i = \min(T_i, C_i)$. The covariates of interest are denoted by X_i.

3.4.1.1. **Basic Definitions**

Let *T* denote a nonnegative random variable, representing time taken for recovery to occur. Let $f(t)$ and $F(t)$ be the respective density and cumulative distribution functions of *T*. The distribution of survival times is characterized by the survival and the hazard functions.

3.4.1.2. **Survival Function**

The survival function is defined as the probability that the survival time is greater or equal to t.

$S(t) = P(T \geq t), t \geq 0$

3.4.1.3. **Hazard Function**

The hazard function gives the instantaneous failure rate at t given that the individual has survived up to time t, i.e.

$$h(t) = \lim_{\triangle t \to 0} \frac{p(t \leq T < \triangle t / T \geq t)}{\triangle t}, t \geq 0$$

$$h(t) = \frac{f(t)}{S(t)} = \frac{-d \log S(t)}{dt}$$

Or the hazard function is the probability that an individual will experience an event.

Relationship between S(t) and h(t)

$$h(t) = \frac{f(t)}{S(t)} = \frac{-d \log S(t)}{dt}$$

$$S(t) = \exp\left[-\int_0^t h(u)du \right] = \exp(-H(t)), t \geq 0$$

Where $H(t) = \int_0^t h(u)du$ is called the cumulative hazard function, which can be obtained from the survival function since, $H(t) = -\log S(t)$

The survival function is most useful for comparing the survival progress of two or more groups. The hazard function gives a more useful description of the risk of failure at any time point.

3.4.1.2. **Non-parametric Survival Methods**

Nonparametric methods are often very easy and simple to understand as compared to parametric methods. Furthermore, nonparametric analyses are more widely used in situations where there is doubt about the exact form of distribution.

Survival data are conveniently summarized through estimates of the survival function and hazard function. The estimation of the survival distribution provides estimates of descriptive statistics such as the median survival time. These methods are said to be non-parametric methods since they require no assumptions about the distribution of survival time. In order to compare the survival distribution of two or more groups, log-rank tests can be used.

3.4.1.2.1. **The Kaplan-Meier Product Limit Method**

In the nonparametric methods, the most popular and commonly used method is the Kaplan-Meier method. It is used for estimating the survival probabilities from observed survival times both censored and uncensored (Kaplan and Meier, 1958). The method is a modified form of the life table technique, with the condition that each time interval contains exactly one event and event occurs at the beginning of the interval.

Suppose that r individuals have failures in a group of individuals. Let $0 \le t_{(1)} < ... < t_{(r)} < \infty$ be the observed ordered recovery times. Let r_j be the size of the risk set at $t_{(j)}$, where risk set denotes the collection of individuals alive and uncensored just before $t_{(j)}$.

Let d_j be the number of observed recovery at $t_{(j)}$, $j = 1, 2, ..., r$. Then the K-M estimator of $S(t)$ is defined by $S(t) = \prod_{j:t_{(j)}<t} \left(1 - \frac{d_j}{r_j}\right)$, this estimator is a step function that changes values only at the time of each recovery.

Suppose that the distribution is discrete, with atoms h_j at finitely many specified points $0 \le \tau_1 < \tau_2 <, ..., < \tau_j$.

The survival function $S(t)$ may be expressed in terms of the discrete hazard function h_j as

$$S(t) = \prod_{j:t_{(j)}<t} \left(1 - h_j\right)$$

3.4.1.2.2. **Median Survival Time**

Median is the preferred summary measure of the location of the distribution. This is the time beyond which 50% of the individuals in the population under study are expected to survive and is given by that value t(50) which is such that $S\{t(50)\} = 0.5$. When no possible realistic an estimated survival time that makes the survival function exactly equal to 0.5, the estimated median survival time, $\hat{t}(50)$ is defined to be the *smallest* observed survival time for which the value of the estimated survival function is less than 0.5.

Mathematically;

$$\hat{t}(50) = \min\left\{t_i / S(t_j) < 0.5\right\}, \qquad [1]$$

Where, t_i is the observed survival time for the i^{th} individual, $i = 1, 2, ..., n$. t_j is the j^{th} ordered recovery time, $j = 1, 2, 3, ..., r$.

3.4.1.2.3. **Nonparametric Comparison of Survival Distributions**

The K-M survival curves can give us an insight about the difference of survival functions in two or more groups, but whether this observed difference is statistically significant requires a formal statistical test. There are a number of methods that can be used to test equality of the survival functions in different groups. One commonly used non-parametric tests for comparison of two or more survival distributions is the log-rank test.

Let $t_1 < t_2 < ... < t_k$ be the ordered recovery times across two groups.

Suppose that d_j failures occur at t_j and that r_j subjects are at risk just prior to t_j, $j = 1, 2, 3, ..., k$ Let d_{ij} and r_{ij} be the corresponding numbers in group $i (i = 1, 2)$.

The **log-rank test** compares the observed number of recovery with the expected number of recovers for group i. Consider the null hypothesis: $S1_{(t)} = S2_{(t)}$ i.e. there is no difference between survival curves in two groups. Given r_j and d_j the random variable d_{1j} has the hypergeometric distribution

$$\frac{\binom{d_j}{d_{1j}}\binom{r_j - d_j}{r_{1j} - d_{1j}}}{\binom{r_j}{r_{1j}}}$$

Under the null hypothesis, the probability of recovery at $t_{(j)}$ does not depend on the group, i.e., the probability of recovery at $t_{(j)}$ is $\frac{d_j}{r_j}$.

$X^2_{\log rank} = \frac{\left[\sum_{j=1}^{k}(d_{1j} - r_{1j} * d_j / r_j)\right]^2}{\sum_{j=1}^{k}\frac{r_{2j} r_{1j} d_j (r_j - d_j)}{[r_j^2 (r_j - 1)]}}$, this statistic approximate X^2 distribution with 1 df.

3.4.1.3. Cox PH Regression Models

The Cox proportional hazards (PH) regression model (introduced in a seminal paper by Cox, 1972), a broadly applicable and the most widely used method of survival analysis. Survival models are used to quantify the effect of one or more explanatory variables on failure time. This involves specification of a linear-like model for the log hazard. A parametric model based on the exponential distribution may be parameterized as follows:

$$\log h_i(t) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik}$$

Or, equivalently:

$$h_i(t) = \exp(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik}) = \exp(\alpha)\exp(X_i^T \beta)$$

In this case the constant α represents the log-baseline hazard since $\log h_i(t) = \alpha$ when all the x's are zero. The Cox PH model is a semi-parametric model where the baseline hazard $\alpha(t)$ is allowed to vary with time:

$$\log h_i(t) = \alpha(t) + \beta_1 x_{i2} + ... + \beta_k x_{ik}$$

$$h_i(t) = h_0(t)\exp(\beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik})$$

$$h_i(t) = h_0(t)\mu_t = h_0(t)\exp(X_i^T \beta) \qquad [2]$$

Where $h_0(t)$, is the baseline hazard function; X_i is a vector of covariates and β is a vector of parameters for fixed effects.

If all of the x 's are zero the second part of the above equation equals 1 so, $h_i(t) = h_0(t)$. For this reason the term $h_0(t)$ is called the baseline hazard function. With the Cox proportional hazards model the outcome is described in terms of the hazard ratio.

This model is called the proportional hazards model because the hazard of any individual is a fixed proportion of the hazard of any other individual, that is:

$$\frac{h_i(t, x_k + 1)}{h_k(t, x_k)} = \exp(\beta_1(x_1 - x_1) + \beta_2(x_2 - x_2) + ... + \beta_k((x_k + 1) - x_k))$$

$$= \exp(\beta_k), \text{ and is independent of the covariate value } X_i.$$

Where, $\exp(r\beta)$: Hazard ratio of two subjects with a difference of r covariate units. Parameter estimate β refers to the increase in log-hazard with a one unit increase for the continuous covariate.

The **survival function** for Cox-PH model is:

$S(t, X) = [S_0(t)]^{\exp\left(\sum_{i=1}^{p} \beta_i X_i\right)}$, the estimated survival function: $\hat{S}(t, X) = \left[\hat{S}_0(t)\right]^{\exp\left(\sum_{i=1}^{p} \hat{\beta}_i X_i\right)}$

Estimated quantities: $\hat{S}_0(t), \hat{\beta}_i$ are estimated baseline survival and coefficient of covariates respectively.

Assumptions of the Cox proportional hazards model are; (1) the ratio of the hazard function for two individuals with different sets of covariates does not depend on time; (2) time is measured on a continuous scale and (3) censoring occurs randomly.

Interpreting outputs from the Cox model involves examining the coefficients for each explanatory variable. Negative regression coefficient for an explanatory variable indicates that the hazard is lower and thus the prognosis worse. Conversely, positive a regression coefficient implies a better prognosis for patients with higher values of that variable when time to event is recovery and conversely for death.

3.4.1.3.1. **Partial Likelihood Function**

Kalbfleisch and prentice derive a likelihood involving only β and X (not $h_o(t)$) based on the marginal distribution of the ranks of the observed failure times (in the absence of censoring). Cox (1972), derived the same likelihood and generalized it for censoring using the idea of a partial likelihood

Suppose we observe (T_i, δ_i, X_i) for individual i, where T_i is a censored failure time random variable; and δ_i is the failure/censored indicator (1=fail, 0= censor) and X_i represents a set of covariates.

The covariates may be continuous, discrete, or time-varying.

$$L_i(\beta) = \Pr(\textit{individual } j \textit{ fails}/1 \text{ failure from R}(\tau_j))$$
$$= \frac{\Pr(\textit{individual fails / at risk at}\tau_j)}{\Pr(\textit{individual } \text{l fails/at risk at } \tau_j)}$$
$$= \frac{h(\tau_j; X_j)}{\sum_{l \in R(\tau_j)} h(\tau_j; X_l)}$$

Under the PH assumption, $h(t;T) = h_o(t)\exp(\beta X)$, so we get:

$$L^{partial}(\beta) = \Pi_{j=1}^{k} \frac{h_o(\tau_j)\exp(\beta X_j)}{\sum_{l \in R(\tau_j)} h_o(\tau_j)\exp(\beta X_l)}$$
$$= \Pi_{j=1}^{k} \frac{\exp(\beta X_j)}{\sum_{l \in R(\tau_j)} \exp(\beta X_l)}$$

In general, the likelihood contributions for censored data fall into two categories:

- individual is censored at T_i:

$$L_i(\beta) = S(T_i) = \exp[-\int_0^{T_i} h_i(u)du]$$

- Individual fails at T_i:

$$L_i(\beta) = S(T_i)h_i(T_i) = h_j(\mathrm{T_i})\exp[-\int_0^{T_i} h_i(u)du]$$

Thus, everyone contributes $S(T_i)$ to the likelihood and only those who fail contribute $h_j(\mathrm{T_i})$.

This means we get a total likelihood of:

$$L(\beta) = \prod_{i=1}^{n} h_i(T_i)^{\delta_i} \exp[-\int_0^{T_i} h_i(u)du]$$

Now, let's multiply and divide by the term $\left[\sum_{j\epsilon R(T_i)} h_i(T_i)\right]^{\delta_i}$.

$$L(\beta) = \prod_{i=1}^{n}\left[\frac{h_i(T_i)}{\sum_{j\epsilon R(T)} h_i(T_i)}\right]^{\delta_i}\left[\sum_{j\epsilon R(T_i)} h_i(T_i)\right]^{\delta_i}\exp[-\int_0^{T_i} h_i(u)du\Bigg]$$

If we just focus on the first term, then under the Cox PH assumption:

$$L(\beta) = \prod_{i=1}^{n}\left[\frac{h_i(T_i)}{\sum_{j\epsilon R(T_i)} h_i(T_i)}\right]^{\delta_i}$$

$$L(\beta) = \prod_{i=1}^{n}\left[\frac{h_0(T_i)\exp(\beta X_i)}{\sum_{j\epsilon R(T_i)} h_0(T_i)\exp(\beta X_i)}\right]^{\delta_i}$$

$$L(\beta) = \prod_{i=1}^{n}\left[\frac{\exp(\beta X_i)}{\sum_{j\epsilon R(T_i)} \exp(\beta X_i)}\right]^{\delta_i}$$

This is the partial likelihood defined by Cox. Note that, it does not depend on the underlying hazard function $h_0(.)$. Cox recommends treating this is as an ordinary likelihood for making inferences about β in the presence of the nuisance parameter $h_0(.)$.

The log-partial likelihood is:

$$l(\beta) = log\left[\prod_{j=1}^{n}\frac{\exp(\beta X_i)}{\sum_{l\epsilon R(\tau_i)} \exp(\beta X_i)}\right]^{\delta_i}$$

$$= log\left[\prod_{j=1}^{k}\frac{\exp(\beta X_i)}{\sum_{l\epsilon R(\tau_i)}\exp(\beta X_i)}\right]$$

$$= \sum_{j=1}^{k}\left[\beta X_i - \log[\sum_{l\epsilon R(\tau_i)}\exp(\beta X_i)]\right]$$

$$= \sum_{j=1}^{k} l_j(\beta)$$

Where, l_j is the log-partial likelihood contribution at the j^{th} ordered event time. Suppose there is only one covariate (β is one dimensional):

The partial likelihood score equations are:

$$U(\beta) = \frac{d}{d\beta}l(\beta) = \sum_{j=1}^{n}\delta_j\left[X_j - \frac{\sum_{l\epsilon R(\tau_j)}X_l\exp(\beta X_l)}{\sum_{l\epsilon R(\tau_j)}\exp(\beta X_l)}\right]$$

The maximum partial likelihood estimators can be found by solving $U(\beta) = 0$.

Analogous to standard likelihood theory, it can be shown that

$$\frac{(\hat{\beta}-\beta)}{se(\hat{\beta})} \sim N(0,1)$$

The variance of $\hat{\beta}$ can be obtained by inverting the second derivative of the partial likelihood,

$$Var(\beta) \sim \left[-\frac{d^2}{d\beta^2}l(\beta)\right]^{-1}$$

Newton Raphson is used by many of the computer packages to solve the partial likelihood equations.

3.4.1.4. **Frailty Models**

The concept of frailty provides a suitable way to introduce random effects in the model to account for association and unobserved heterogeneity. In its simplest form, a frailty is an unobserved random factor that modifies multiplicatively the hazard function of an individual or a group or cluster of individuals. Vaupel et al. (1979) introduced the term frailty. Clayton (1978) promoted the model by its application to multivariate situation on chronic disease incidence in families.

A random effect model takes into account the effects of unobserved or unobservable heterogeneity, caused by different sources. The random effect, called frailty and denoted here by *Z* is the term that describes the common risk, acting as a factor on the hazard function.

3.4.1.4.1. **Shared Frailty Model**

A natural extension of the univariate frailty model would be a multivariate survival model where individuals are allowed to share the same frailty value. Frailty models are getting more and more popular to account for over-dispersion and/or clustering in survival data. Gets name because they attempt to account for unobserved heterogeneity that occurs because some observations are more failure prone and hence, more "frail" than other observations in a data set. The basic idea is to introduce into the hazard rate, an additional random parameter that accounts for the random frailties. The concept of frailty was introduced by Vaupel et al. (1979) who studied the model with Gamma distributed frailties. In recent decades, a large amount of papers on "frailty models'" have appeared. The assumption of a shared frailty model is that both individuals in a pair share the same frailty *Z*, and this is why the model is called the shared frailty model. It was introduced by Clayton (1978) and extensively studied in Hougaard (2000), Therneau and Grambsch (2000), Duchateau et al. (2002), (2003) and Duchateau and Janssen (2004). These frailties may be individual-specific or group-specific thus giving rise to the nomenclature "individual frailty" or "shared frailty" models. Shared-frailty models are appropriate when you wish to model the frailties as being specific to groups of subjects, such as subjects within families. Here a shared-frailty model may be used to model the degree of correlation within groups; i.e., the subjects within a group are correlated because they share the same common frailty.

In this situation, individuals j in a cluster i are supposed to share the same frailty Z_i, the conditional hazard for individual j in cluster i is:

$$h(t_{ij} / Z_i) = Z_i h(t_{ij}), \qquad [6]$$

Where, $h\left(t_{ij}\right) = h_0(t_{ij})\exp(\beta X_{ij})$ in the Cox regression model. The Z_i are independent identically distributed following a chosen distribution, like in the univariate frailty models. This model is therefore an extension of the described model. The interpretation of this model is that the between-groups variability (the random variation of Z) leads to different risks for the groups, which then show up as dependence within the group.

3.4.1.4.1.1.**Shared Gamma Frailty model**

Suppose there are n clusters and that the i^{th} cluster has k_i individuals and associates with an unobserved frailty, Z_i, $1 \le i \le n$. A vector $X_{ij}, 1 \le i \le n; 1 \le j \le k_i$ is associated with the ij^{th}. Suppose T_{ij} is the survival time of the j^{th} individual in the i^{th} cluster. Conditional on frailties, Z_i, the survival times are assumed to be independent and their hazard functions to be of the form

$$h(t_{ij} / Z_i) = Z_i h_0(t_{ij})\exp(\beta^T X_{ij}), i = 1,2,\ldots n; j = 1,2,\ldots,k_i \qquad [7]$$

Where, $h_0(t)$ are the baseline hazard functions and β is a vector of fixed effect parameters to be estimated. The frailties, Z_i are assumed to be identically and independently distributed random variables with a common density function $f(z,\theta)$ where, θ is the parameter of the frailty distribution. Individuals in cluster i with $Z_i > 1$ tend to fail at a faster rate than that under independence model. A semi-parametric shared frailty model is a frailty model with a nonparametric baseline hazard function $h_0(t)$.

The gamma frailty model assumes a gamma distribution for the frailties. The Gamma distribution is the most widely applied frailty distribution. The shared gamma frailty model was suggested by Clayton (1978) for the analysis of the correlation between clustered survival times in genetic epidemiology. An advantage is that without covariates its mathematical properties are

convenient for estimation (Oakes, 1982, 1986). From an analytical and computational view gamma is a very convenient distribution. Arguably, this is the most popular frailty model due to its mathematical tractability, by Duchateau and Janssen (2008). When observations are clustered into groups such as districts, or when observations are recurrent events times (cancer relapses), the shared gamma frailty model is the most often adapted model (Rondeau et al. 2003). In the model above or the use of the frailty model was justified on the ground that patients given the same intervention may not necessarily be coming from the same environment or were not exposed to the same events.

It is assumed that the Z_i 's are independently and identically distributed from a gamma distribution with mean 1 and unknown variance θ; the probability density function is thus:

$$f_z(z) = \frac{z^{\frac{1}{\theta}-1} \exp(-z/\theta)}{\theta^{\frac{1}{\theta}} \Gamma(\frac{1}{\theta})} \qquad [8]$$

Where, $\Gamma(.)$ is gamma function, it corresponds to a gamma distribution, $Gamma\ (\mu,\theta)$ with $\Gamma(.)$ μ fixed to 1 for identifiability. Its variance is then θ with Laplace transform

$$L(u) = (1+u/\theta)^{-\theta},\ \mathrm{u} \geq 0$$

Large values of θ signify a closer positive relationship between the subjects of the same group and greater heterogeneity among the groups.

The conditional survival function of the gamma frailty distribution is given by: (Gutierrez, 2002)

$$S_\theta(t) = [1-\theta \ln(S_0(t))]^{-1/\theta}$$

and the conditional hazard function is given by:

$$h_\theta(t) = h_0(t)[1-\theta \ln(S_0(t))]^{-1/\theta}$$

Where, $S_0(t)$ and $h_0(t)$ are the survival and the hazard functions of the baseline distributions. In the case of gamma distribution for Z, $EZ=1$ and $varZ=\theta$. So, small value of $1/\theta$ reflect a greater degree of heterogeneity among groups and a stronger association within groups. For the Gamma distribution, the Kendall's Tau (Hougaard 2000), which measures the association

between any two event times from the same cluster in the multivariate case, can be computed by: $\tau = \frac{\theta}{\theta + 2} \epsilon (0,1)$.

The joint survival function for the k_i individuals within the i^{th} cluster is easily written by:

$$S\left\{t_{i1},\ldots,t_{ik_i}\right\} = \Pr\left(T_{i1} > t_{i1},\ldots T_{ik_i} > t_{ik_i}\right) = \int_0^{\infty} \prod_{j=1}^{k_i} \Pr(T_{i1} > t_{i1} / Z_i) g(z_i) dz_i$$

$$= \left[1 + \frac{1}{\theta} \sum_{j=1}^{k_i} H_0(\mathrm{t_{ij}}) \exp(\beta^{\mathrm{T}} \mathrm{X_{ij}})\right]^{-\theta}$$

In this model, the estimates of β, θ, $H_0(t)$ are obtained by using the penalized likelihood maximization (PLM). PLM is more elegant estimation tool.

3.4.1.4.1.1.1. **Penalized Likelihood Method**

Let's introduce a semi-parametric approach to jointly estimate the parameters β, θ and the baseline hazard function $\mathrm{h_0(t)}$, which is assumed to be smooth. A possible means for introducing such an a priori knowledge is to penalize the likelihood by a term which has large values for rough functions. (O'Salivan, 1988; Joly, Commenges and letenneur, 1998). Thus for the vectors of baseline hazard functions, if conditionally on b the censoring is independent and non-informative also of b**,** then the likelihood for model (6) in terms of the parameters ($h_0(\mathrm{t}), \beta, \theta$) is

$$L(h_0(\mathrm{t}), \beta, \theta) = \int \prod_{i=1}^{n} h_i(\mathrm{t}/\mathrm{b})^{\delta_i} \mathrm{S_i(t/b)} \mathrm{p(b; D/\theta))db}$$

$$= \int \prod_{i=1}^{n} (h_0(\mathrm{t}) \exp(X_i^T \beta + Z_i b))^{\delta_i} exp\left[-H_0(\mathrm{t}) \exp(X_i^T \beta + Z_i b)\right] x\, p(b, D(\theta))\, db$$

Where $H_0(\mathrm{t}) = \int_0^t h_0(\mathrm{u})\,\mathrm{du}$ and the unobserved frailties are integrated out.

We restrict *b* to follow a multivariate normal distribution, but the derived likelihood approximations can be easily adapted to other frailty distributions as well. The approximate marginal log likelihood:-

$$L\left(h_{\mathrm{o}}(\mathrm{t}),\beta,\ \theta\right)\approx-\frac{1}{2}log\left|D(\theta)\right|-\frac{1}{2}log\left|\sum_{i=1}^{n}H_{o}(\mathrm{t})\exp\left(X_{i}^{T}\beta+Z_{i}\grave{b}\right)Z_{i}\grave{Z}_{i}-D(\theta)^{-1}\right|+$$

$$\sum_{i=1}^{n}\delta_{i}\left[\log\left(h_{0}(t)\right)+X_{i}^{T}\beta+Z_{i}\grave{b}\right]-H_{o}(\mathrm{t})\exp\left(X_{ij}^{T}\beta+Z_{i}\grave{b}\right)-\frac{1}{2}\grave{b}\grave{}\ D(\theta)^{-1}\grave{b} \qquad [9]$$

If both θ were known and *b* were considered a fixed effects parameter, then the second line in (9) would be a penalized log likelihood (Green, 1987), where $-\frac{1}{2}\grave{b}\grave{}\ D(\theta)^{-1}\grave{b}$ is the penalty term penalizing for extreme values of *b*. Since the second line is the full likelihood for a Cox model with *b* as another set of parameters and a penalty term, it turns out that it can be maximized using penalized fixed effects partial likelihood (PPL),

$$\delta_{i}(\left[X_{i}^{T}\beta+Z_{i}b\right]-\log\sum_{\mathrm{j}\epsilon\mathrm{R}(\mathrm{t}_{\mathrm{i}})}\exp(X_{i}^{T}\beta+Z_{i}b))-\frac{1}{2}\grave{b}\,D\left(\theta\right)^{-1}b$$

For given θ, the estimating equations based on the first partial derivatives of the PPL are, for β,

$$\sum_{i=1}^{n}\delta_{i}\left[X_{i}-\frac{X_{i}\exp(X_{i}^{T}\beta+Z_{i}b)}{\sum_{\mathrm{j}\epsilon\mathrm{R}(\mathrm{t}_{\mathrm{i}})}\exp(X_{j}^{T}\beta+Z_{j}b)}\right]=0 \qquad [10]$$

For b

$$\sum_{i=1}^{n}\delta_{i}\left[Z_{i}-\frac{Z_{i}\exp(X_{i}^{T}\beta+Z_{i}b)}{\sum_{\mathrm{j}\epsilon\mathrm{R}(\mathrm{t}_{\mathrm{i}})}\exp(X_{j}^{T}\beta+Z_{j}b)}\right]-D(\theta)^{-1}=0 \qquad [11]$$

and $(\grave{\beta}\left(\theta\right),\ \grave{\mathrm{b}}(\theta))$ can be found by alternating between solving (10) and (11). Note that equation (10) can be solved with standard Cox regression software using estimated values of frailties as the offset term.

3.4.1.5. **Model Selection**

3.4.1.5.1. **Likelihood Ratio Tests (LRT)**

The likelihood ratio test (LRT) statistic is an adequate test as the new model is nested in the previous model. Suppose there are $(p+q)$ explanatory variables measured: $x_i,\ldots,x_p,x_{p+1},\ldots,x_{p+q}$ and proportional hazards are assumed. Consider the following models

Model 1: contains only the first p-covariates $\frac{h_i(t,X)}{h_o(t)} = \exp(\beta_1 x_1 + \ldots + \beta_p x_p)$

Model 2:-contains all $(p+q)$ covariates $\frac{h_i(t,X)}{h_0(t)} = \exp(\beta_1 x_1 + \ldots + \beta_{p+q} x_{p+q})$.

These are nested models. For such nested models, we can construct a likelihood ratio test of $\mathrm{H}_0 : \beta_{p+1} = \ldots = \beta_{p+q} = 0$ as: $X_{LR}^{\ 2} = -2[\log(\hat{L}(1)) - log(\hat{L}(2))]$ under H_0 , this test statistic is approximately distributed as X^2 with q df.

The likelihood-ratio test of $\theta = 0$ is a boundary test and thus requires careful consideration concerning the calculation of its p-value. In particular, the null distribution of the likelihood-ratio test statistic is not the usual $\chi_1^{\ 2}$ but is rather a 50:50 mixture of a $\chi_0^{\ 2}$ (point mass at zero) and a denoted as $\bar{\chi}_{01}^{\ 2}$. See Gutierrez, Carter, and Drukker (2001) for more details.

3.4.1.5.2. **Akaike's Information Criterion (AIC)**

Some models are nested within gamma model. For comparing models that are not nested, the Akaike information criterion (AIC) can be used instead, which is defined as:- $AIC = -2(log\ likelihood) + 2(k)$, where, k is the number of covariates in the model. The addition of $2(\mathrm{k})$ can be thought of as a penalty if non predictive parameters are added to the model. Although the best-fitting model is the one with the largest log likelihood, the preferred model is the one with the smallest AIC value.

3.4.1.6. **Model Checking and Diagnosis**

The use of diagnostic procedures for model checking is an essential part of the modeling process. There are different commonly used model checking to evaluate whether the appropriate functional form for a covariate is used in the model to assess the fitted model.

3.4.1.6.1. **Cox-Snell residuals**

The residual that is most widely used in the analysis of survival data is the Cox-Snell residual, so called because it is a particular example of the general definition of residuals given by Cox and Snell (1968). The Cox-Snell residual for the i[th] individual, $i = 1, 2, \ldots, n$, is given by properties and features of residuals, when survival outcome are modeled, have been extensively studied in the literature. The Cox -Snell residuals are commonly used for a direct assessment of excess events (i.e., to reveal subjects that are poorly fit by the model), and for evaluating whether the appropriate functional form for a covariate is used in the model.

$S(t:X) = \left[S_o(t)\right]^{\exp(\beta X)}$ or, in terms of hazards: $h(t;X) = h_o(t)\exp(\beta X)$ So, for each person with covariates x_i, $\mathrm{S(t:x}_i) = \left[S_o(t)\right]^{\exp(\beta x_i)}$ then we can calculate $\hat{h}_i = -\log[\hat{S}(T_i; x_i)]$

Or first predict survival probability at the actual survival time for individual, then log-transform it. The residuals in right censored data constitute a censored sample of the unit exponential distribution $r_{Ci} = \hat{H}_i\left(t_i^*\right) = -\log \hat{S}_i\left(t_i^*\right)$, [13]

Where $\hat{H}_i\left(t_i^*\right)$ and $\hat{S}_i\left(t_i^*\right)$ are the estimated cumulative hazard and survivor functions, respectively, for the i[th] individual at the censored survival time.

Then the modified Cox-Snell residual is given by

$r'_{Ci} = 1 - \delta_i + r_{Ci}$ [14]

- Plotting $-\log \hat{S}(T_i)$ vs t should yield a straight line
- Plotting $\log[-\log(\hat{S}(T_i))]$ vs $\log(\mathrm{t})$ should yield a straight line through the origin with slope=1.

Note: - the Cox-Snell residuals will not be symmetrically distributed about zero and cannot be negative.

3.4.1.6.2. **Deviance Residuals**

Although martingale residuals share many of the properties possessed by residuals encountered in other situations such as in linear regression analysis, they are not symmetrically distributed about zero, even when the fitted model is correct. This skewness makes plots based on the residuals difficult to interpret. The deviance residuals, which were introduced by Therneau et al.(1990) are much more symmetrically distributed about zero. They are defined by

$$r_{Di} = sgn(r_{Mi})[-2\{r_{Mi} + \delta_i \log(\delta_i - r_{Mi})\}]^{1/2} \qquad [16]$$

Where, r_{Mi} is the martingale residual for the ith individual and the function $\text{sgn}(.)$ is the sign function. This is the function that takes the value +1 if its argument is positive and -1 if negative.

3.4.1.6.3. **Influential Observations**

Observations that have an undue effect on model-based inference are said to be influential. The most direct measure of influence is $\hat{\beta}_j - \hat{\beta}_{j(i)}$ where, $\hat{\beta}_j$ is the jth parameter, j =1, 2... p in a fitted Cox PH model and is obtained by fitting the model after omitting observation i. To check the influence of observations on a parameter estimate, Cain and Lange showed that an approximation to $\hat{\beta}_j - \hat{\beta}_{j(i)}$ is the j component of the vector $r'_{si}\,\text{var}(\hat{\beta})$ where r_{si} is the $p \times 1$ vector of score residuals for the ith observation, which are modifications of Schoenfeld residuals and are defined for all the observations, and Var(β) is the variance-covariance matrix of the vector of parameter estimates in the fitted Cox PH model. The jth element of this vector is called delta-beta statistic for the jth explanatory variable $\Delta\hat{\beta}_j = \hat{\beta}_j - \hat{\beta}_{j(i)}$, which tells us how much each coefficient will change by removal of a single observation.

3.4.1.7.**Testing the Assumption of Proportional Hazards**

It is always a good practice to check the assumption of proportional hazards, before proceeding further with other inferential activities. Schoenfeld residuals can be used for this purpose.

3.4.1.7.1. **Using Schoenfeld residuals**

The expected value of the i[th] scaled Schoenfeld residuals for the j[th] explanatory variable is given by $E[r^*_{p_{ji}}] \approx \beta_j(t_i) - \hat{\beta}_j$, where,

- $\beta_j(t_i)$, the value of time varying coefficient of x_j at the i[th] death time.
- $\hat{\beta}_j$, is the estimated value of β_j in the fitted Cox model.

Plot $r^*_{pji} + \hat{\beta}_j$, against the recovery time. A horizontal line would suggest that the coefficient of x_j is constant and the proportional hazards assumption is satisfied.

3.5. ETHICAL CONSIDERATION

The data for the analysis is obtained from Jimma University Specialized Hospital (JUSH), and an ethical clearance for the study was provided by research ethics review board of Jimma University and the Department of Statistics was written an official support letter to JUSH. Qualified data collectors were carefully recruited and trained before the start of the data collection phase.

4. STATISTICAL ANALYSIS AND RESULTS

4.1. Descriptive and non-parametric survival analysis of time to first recovery of diabetic patients

Baseline categorical covariates are illustrated in Table 2. The data consists of 544 patients aged equal to or above 18 years old and placed under treatments that have followed between 1st September 2010 and 30th August 2013 (a three years data) at JUSH were included to find out their time to first recovery in to normal blood sugar level.

The outcome response is time to first recovery. From the total of 544, 404 (74.26%) were recovered to normal blood sugar level and the rest 140 (25.74%) loss to follow-up from the study. The median recovery time is 3 month but it varies depending on the covariates included in the study.

The majority of the cases, 423(77.76%) were type-II diabetic, 351(64.52%) were males, 423 (77.76%) had no family history of the disease. The results further shows that the majority of the patients were first detected at the age of 45-59 years (30.51%) followed by age group 30-44 (28.49%) and a few cases at age above 74 years (4.96%). It can be seen that most of the cases had normal upper (systolic) (62.68 %) and lower (diastolic) (77.76 %) blood pressure.

In type-I diabetic from the total of 121, 103 (85.12%) experienced the event and the rest 18(14.88%) loss to follow-up and in type-II diabetic from the total of 423, 301 (71.16%) experienced the event and the rest 122 (28.84%) loss to follow-up from the study. The median recovery time for type-I and type-II diabetic were between 2 and 4 months respectively. The minimum and the maximum recovery time of type-I diabetic were 1 and 6 months and for type-II diabetic 1 and 31 months respectively. In type-I diabetic from the total of 103 recovered patients, 71(84.52%) were males and median recovery time was 2 months whereas, 32(86.49%) were females and median recovery time 1 month and in type-II from the total of 423 recovered patients, 202(75.66%) were males and median recovery time 3 months whereas, 99(63.46%) were females and median recovery time 8 months. Majorities of females with type-I DM are recover to normal blood sugar level as compared to males and majorities of males with type-II DM were recover to normal blood sugar level as compared to females.

Table 2: Diabetes mellitus patient baseline covariates of median recovery, percentage and frequencies

			Type-I Diabetic Mellitus (DM)			Type-II Diabetic Mellitus (DM)		
Covariates		**Total No (%)**	**Total (%) (for type-I)**	**Recovery**	**Median**	**Total (%) (for type-II)**	**Recover**	**Median**
Sex	Male	351 (64.52)	84(23.93)	71 (84.52)	2	267(76.07)	202(75.66)	3
	Female	193 (35.48)	37(19.17)	32(86.49)	1	156(80.83)	99(63.46)	8
Family history	Yes	121(22.24)	121(22.24)	103(85.12)	2			
	No	423(77.76)				423(77.76)	301(71.16)	4
Age at	<30	109(20.04)	39(35.78)	38(97.44)	1	70(64.22)	64(91.43)	1
baseline (yrs)	30-44	155(28.49)	53(34.19)	44(83.02)	2	102(65.81)	82(80.39)	3
	45-59	166(30.51)	22(13.25)	19(86.36)	2	144(86.75)	101(70.19)	5
	60-74	87(15.99)	4(4.6)	2(50)	4	83(95.4)	44(53.01)	10
	>74	27(4.96)	3(11.11)	0	NA	24(88.89)	10(41.67)	16
SBP(mm/Hg)	<110	94(17.28)	19(20.21)	18(94.74)	1	75(79.79)	57(76)	5
	110-130	341(62.68)	84(24.63)	71(84.52)	2	257(75.37)	186(72.37)	4
	>130	109(20.04)	18(16.51)	14(77.78)	2	81(74.31)	58(71.60)	5
DBP(mm/Hg)	<60	8(1.47)	2(25)	2(100)	1	6(75)	4(66.67)	5
	60-80	423(77.76)	101(23.88)	86(85.15)	2	322(76.12)	227(70.50)	4
	>80	113(20.77)	18(15.93)	15(83.33)	1.5	95(84.07)	70(73.68)	4
Overall		544	121(22.24)	103(85.12)	2	423(77.76)	301(71.16)	4
Overall DM		544	404(74.265)			3(Median)		

Area: Jimma University Specialized Hospital (JUSH); study time: between September 2010 and August 2013 (a three year data); Median: Median recovery time; *DM: Diabetic Mellitus.*

All patients with type-I diabetic had family history of the disease whereas type-II diabetic had no family history of the disease. The results further shows that the majority of the type-I and II diabetic patients were first detected at the age of 30-44 (43.8%) and 45-59 (34.04%) years respectively. Type-I DM detected at young age whereas type-II DM at adult age. Patient's age group 18-29 years was fast to recover to normal blood sugar level both for type-I and II DM.

Table 3: Baseline characteristics of DM patients and types of DM for continues variables

Covariates	DM		Type-I DM		Type-II DM	
	Mean ± SD	Median	Mean ± SD	Median	Mean ± SD	Median
Bodyweight	61.63± 16.16	60	55.52± 14.89	56	63.38± 16.10	61
Age	44.66 ± 16.11	45	35.93± 13.12	35	47.16± 16.02	48
Upper (systolic) blood pressure	120.55± 16.17	120	119 ± 13.56	120	120.99± 16.83	120
Lower(Diastolic) blood pressure	77.83± 10.12	80	77.77± 9.7	80	77.85± 10.25	80
Fasting Blood sugar	215.38± 91.37	196	209.34± 89.1	188	217.999± 90.793	199

Table 3 describes the baseline characteristics of diabetic mellitus patients and types of diabetes for continue variables. The mean age at the start of follow up was 35.93 yrs for type –I DM with a standard deviation of 13.12 and median 35 yrs, the mean age for type-II DM was 47.16yrs with a standard deviation of 16.02 and median age was 48 yrs, the mean bodyweight at the start of follow up for type-I DM was 55.52kgs with a standard deviation of 14.89 and median 56kgs, the mean bodyweight for type-II DM was 63.38kgs with standard deviation of 16.10 and median 61kgs, the mean upper (systolic) blood pressure for type-I DM was 119 mm/Hg with a standard deviation and median 120mm/Hg, the mean upper (systolic) blood pressure for type-II DM was 120.99mm/Hg with a standard deviation of 16.83 and median 120mm/Hg and the mean lower(diastolic) blood pressure for type-I DM was 77.77mm/Hg with a standard deviation of 9.7 and median 80mm/Hg and the mean lower(diastolic) blood pressure for type-II DM was 77.85mm/Hg with a standard deviation of 10.25 and median 80mm/Hg. Type-II diabetic patients were the oldest whereas type-I diabetic patients were the youngest. Bodyweight and FBS at baseline for type-II (61kg and 199mg/dl) patients were large as compared to type-I (56kg and 188mg/dl) diabetic patients.

Kaplan-Meier Estimates and Logrank Tests

The logrank test (Section 3.4.3.3) and a plot of Kaplan-Meier (Section 3.4.3.1) estimates for only two selected categorical covariates; diabetic type and sex are displayed below. The K-M curves for each study sub group (category) provide an initial insight difference for each subgroup.

Table 4: The log-rank test for Diabetic Type

Diabetic Type	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
DT=type-I	123	104	62.3	27.91	42.9
DT=type-II	421	300	341.7	5.09	42.9
Chisq= **42.9** on 1 degrees of freedom, **p= 5.7e-11**					

Table 5: The log rank test for Sex

Sex	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
Sex=female	193	131	185	15.8	38.2
Sex=male	351	273	219	13.3	38.2
Chisq= 38.2 on 1 degrees of freedom, p= 6.49e-10					

As per to the log-rank presented in table 4 and 5 there is a significant difference in the cumulative incidence of recovery time for diabetic type (p= 5.7e-11) and sex (p= 6.49e-10). Figure 1 shows patients with type-I diabetic and male patients have a higher probability to recover than type-II and female DM patients respectively. The results are consistent with the log rank test.

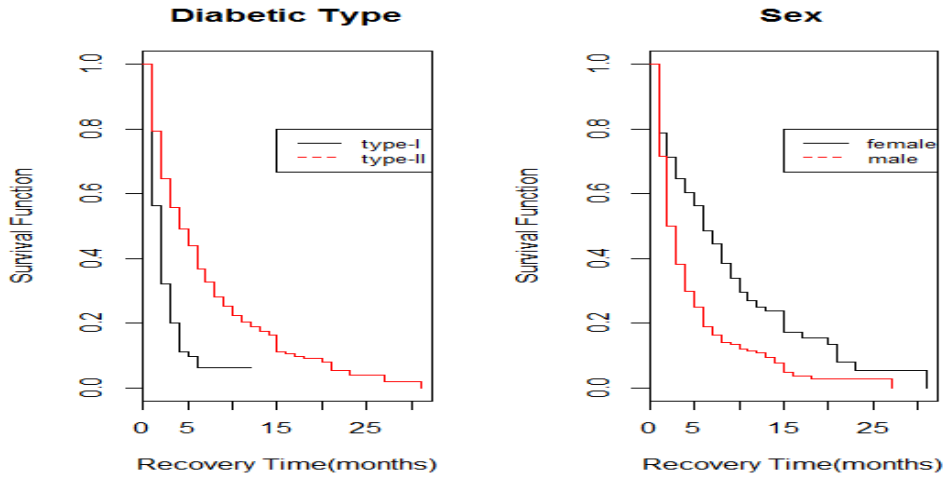


Figure 1. Kaplan-Meier estimate of the Diabetic type and Sex survivor function

4.2. Result: from Cox-PH Models

4.2.1. **Modeling Recovery Time for DM Inseparably**

In order to select variables in the model, first univariate analysis is used to check all the covariates associated with recovery time. Accordingly, the univariate Cox proportional hazards regression models are fitted for every covariate shown (Appendix-I A). In this study, the predictors in the multivariate model is considered, if the test for the univariate model has a p-value less than or equal to 0.1 in the univariate analysis. Then the full multivariate Cox PH model is fitted including all the potential covariates which are significant at 10% at the univariate level and from multivariate model variables non-significant at 10% were eliminated using backward selection method. Accordingly variables with minimum AIC are; bodyweight at baseline, age group, sex, fasting blood sugar (FBS) at baseline and diabetic types are significant covariates selected for the model (Table 6).

Table 6: Multivariate Cox-PH model for the diabetic types inseparably

Covariates	Coef$\left(\hat{\beta}\right)$	se(Coef$\left(\hat{\beta}\right)$)	P-value	Hazard Ratio(HR)	95%C.I for HR
Age					
30-44 years	-0.387	0.135	0.004125*	0.6788	(0.5209, 0.8845)
45-59 years	-0.695	0.139	5.84e-07*	0.4992	(0.3801, 0.6556)
60-74 years	-1.378	0.1858	1.22e-13*	0.2521	(0.17514, 0.3629)
>74 years	-1.909	0.3465	3.63e-08*	0.1483	(0.07517, 0.2924)
Bodyweight(kg)	-0.0216	0.0032	1.68e-11*	0.9786	(0.97248, 0.9848)
Diabetic Type (Type-II)	-0.482	0.1240	0.000101*	0.6173	(0.48409, 0.7872)
FBS	-0.0032	0.0006	9.91e-08*	0.9968	(0.9956, 0.998)
Sex (Male)	0.582	0.1127	2.39e-07*	1.79	(1.43525, 2.2324)
Likelihood ratio test= 251.2.1, p=0.0001, Wald test = 222.2 p=0.0001, Score (logrank) test = 243.3, p=0.0001, AIC= 4138.79.					

Area: *Jimma University Specialized Hospital (JUSH);* ***study time:*** *between September 2010 and August 2013 (a three year data),* ***Inseparable:*** *when two types of diabetic is in one data set,* ***Coef:*** *coefficient for covariate,* ***HR:*** *hazard ratio,* ***p-value:*** *probability value,* ***95%C.I HR:*** *95% confidence interval for HR, FBS: Fasting Blood Sugar,* ******* *Significant at 0.05 level.*

In the univariate (Appendix-I A) Cox PH models, the model with a covariate, age group 30-44 (P-value=0.00148), 45-59(P-value=5.62e-09), 60-74 (P-value=< 2e-16), >74(P-value=1.89e-11) (when age group <30 as a reference), bodyweight (kg) (P-value=2.22e-15), sex (P-value=1.71e-09 when female as a reference) , diabetic type (P-value= 1.07e-14 when type-I as a reference) and FBS(mg/dl)(p-value=3.44e-08) at base line shows statistically significant association with

time-to-recovery at 10% level of significance. But, upper (Systolic) (110-130, P-value= 0.617& >130, P-value =0.359) (<110 as a reference) and lower (Diastolic) (60-80, P-value=0.843 &>80, P-value =0.837) (<60 as a reference) blood pressure at baseline are not significant at 5% let leave alone10% revealing that this variable will not be included in the model.

The multivariate results of a Cox PH model fitted to this dataset were obtained on table 6. It is now observed that effects of age group, bodyweight (kg) (p-value= 1.68e-11) at baseline, sex (p-value=2.39e-07), FBS (mg/dl) (p-value=9.91e-08) at baseline and diabetic type (p-value= 0.00010 1) are significantly associated. Again also, the likelihood ratio test (p-value=0.0001), Wald test (p-value=0.0001) and score (logrank) test (p-value=0.0001) are highly significant.

Estimating the hazard and survival functions

The semi-parametric Cox PHs model is the most commonly used model in hazard regression. In this model, the conditional hazard function, given the covariate value x, is assumed to be of the form

$h(t \setminus x) = h_0(t)\exp\{\beta^T x\}$, where, $\beta = (\beta_1,...,\beta_p)^T$ is the vector of regression coefficients, and $h_0(t)$ denotes the baseline hazard function. No particular shape is assumed for the baseline hazard; it is estimated non-parametrically. The contributions of covariates to the hazard are multiplicative.

The baseline survival function is estimated as $S_0(t) = \exp\{-h_0(t)\}$, by $\hat{S}_0(t) = \exp\{-\hat{h}_0(t)\}$

In the Cox PHs model, the survival function $S(t \setminus x)$ of an individual with covariate values x is given by $S(t \setminus x) = S_0(t)^{\exp(\beta^T x)}$

The final multivariate Cox PH model (2) (table6)is then written as:-

$$h_i(t) = h_o(t)\exp(-0.387 * Age_{30-44i} - 0.695 * Age_{45-59i} - 1.378 * Age_{60-74i} - 1.909 * Age_{>74i} - 0.0216 * bodyweight_i - 0.482 * DT_i - 0.0032 * FBS_i + 0.582 * Sex_i)$$

The estimated survival model:

$$\hat{S}(t,X)=\left[\hat{S}_0(t)\right]^{\left(\begin{array}{c}\exp(-0.387*Age_{30-44i}-0.695*Age_{45-59i}-1.378*Age_{60-74i}-1.909*Age_{>74i}\\ -0.0216*bodyweight_i-0.482*DT_i-0.0032*FBS_i+0.582*Sex_i)\end{array}\right)}$$

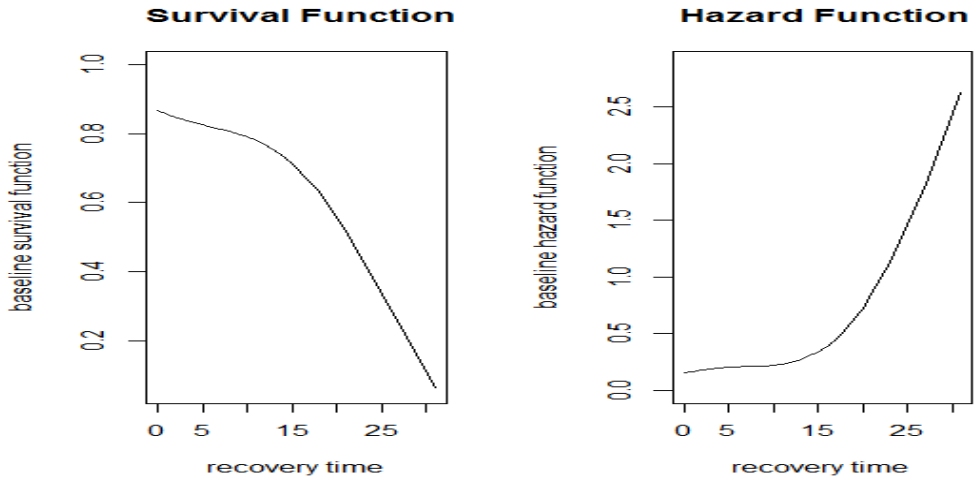


Figure 2: Estimated of the baseline survival and hazard in the PHs models respectively for inseparable DM patients of Cox PH model

The baseline survival and hazard are decreasing and increasing respectively as time goes (Figure 2). The Cox PHs model is a semi-parametric model where the baseline hazard $h_0(t)$ and survival $S_0(t)$ vary with time. A baseline hazard function is left unspecified but must be positive. The patients' chance of recovery time up to 15 months slowly decreases and increases for baseline survival and hazard function respectively.

The PH assumption checking with graphical method based on the Schoenfeld residuals have been described (Figure 3) and included in the model. Systematic departures from a horizontal line are indicative of non-proportional hazards, since PH assumes that estimates $\beta_1, \beta_2, ..., \beta_p$ do not vary much over time. Also, the graphs for some of the categorical variables displayed (Figure 1) using Kaplain Meir appeared were parallel, implying that the proportional-hazards assumption among categorical and continues variables has not been violated.

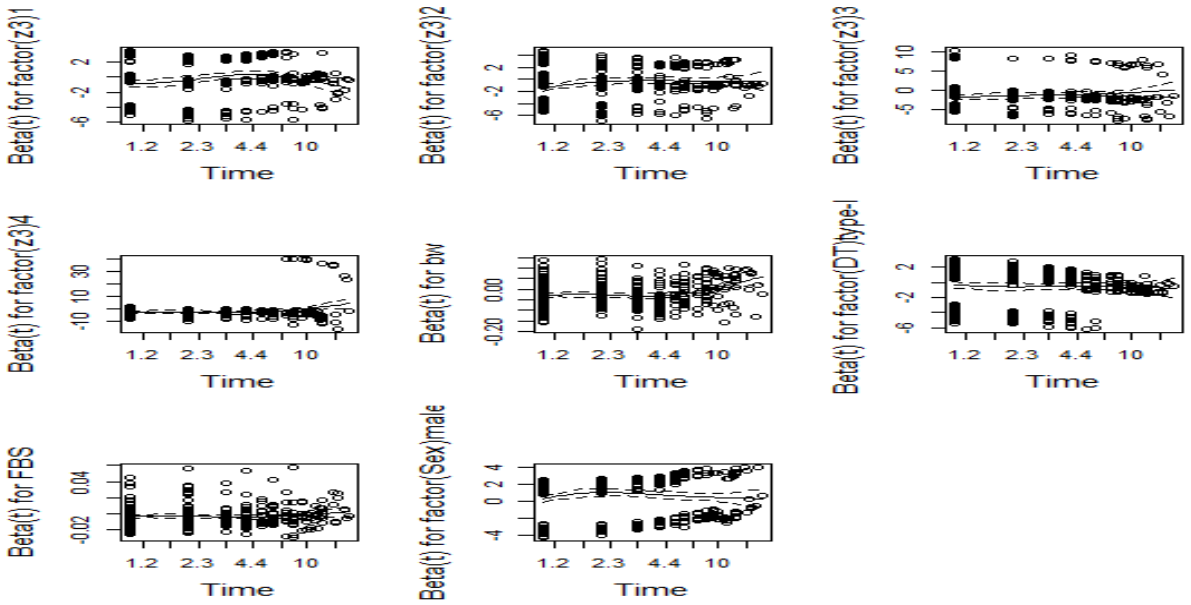


Figure 3: Plots of Scaled Schoenfeld Residuals for each Covariate in the model for inseparable DM data

4.2.1.1. Model Diagnosis for Diabetic Types Inseparably

A plot of the Cox-Snell residuals against the cumulative hazard is presented (Figure 4). The hazard function follows the 45 degree line very closely except for very large values of time. It is very common for models with censored data to have some wiggling at large values of time and it is not something which should cause much concern. Overall, the final model fits the data very well.

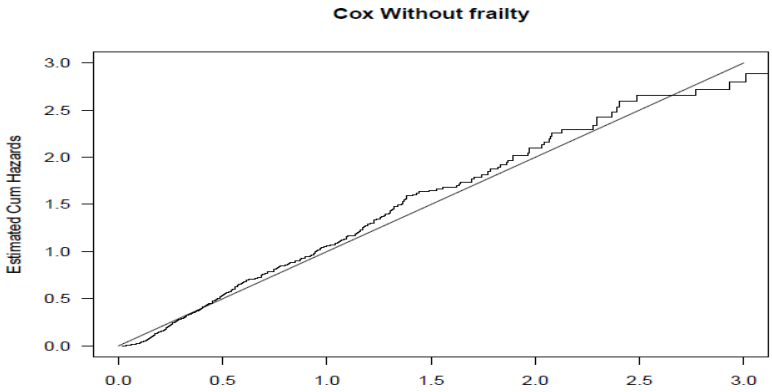


Figure 4: Cox-Snell residuals obtained from fitting Cox PH model to the DM data.

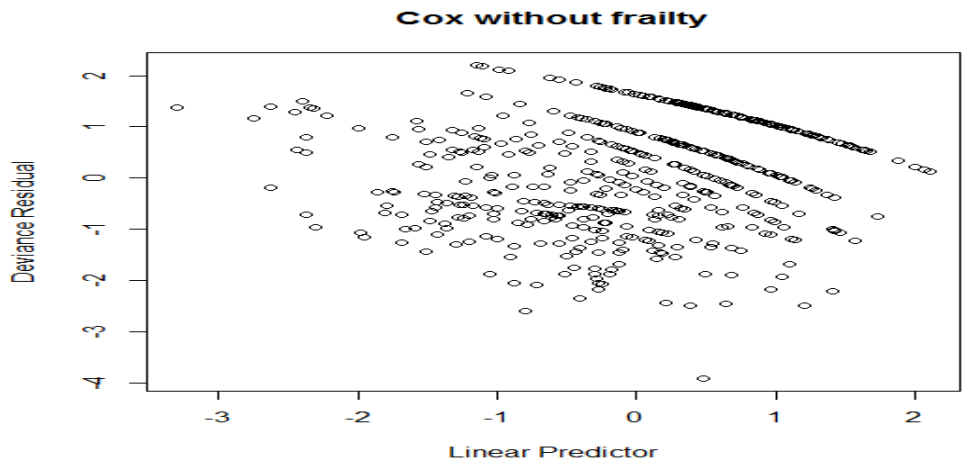
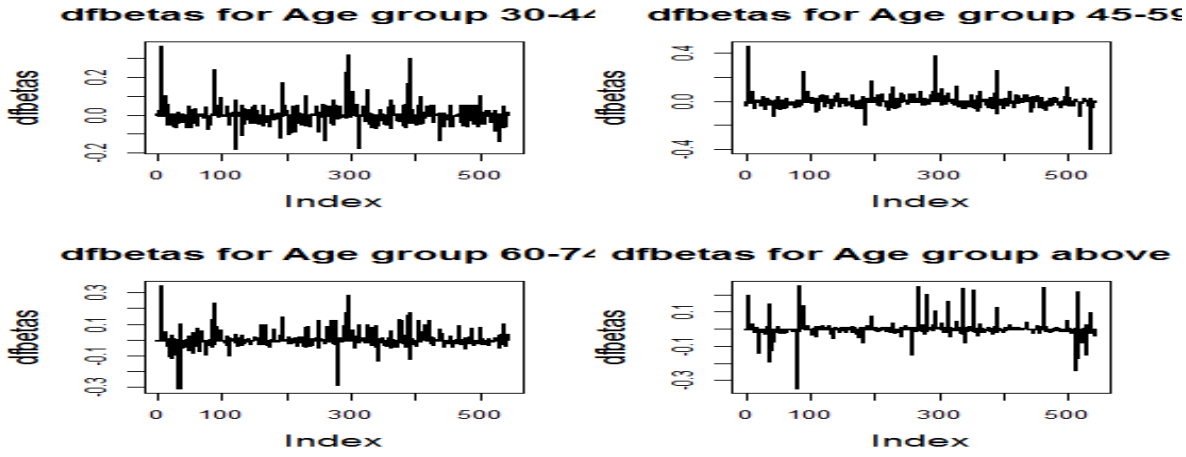


Figure 5: Deviance Residuals for Cox PH model to the DM data

The plot of deviance residual (Figure 5) shows that the deviance residuals seem to be approximately symmetrically distributed about zero and there exists not as such clearly outlying observation. There is a pattern in the residuals that do not necessarily indicate any problems with the model. Therefore, there is almost some concern about the adequacy of the fitted Cox PH model.



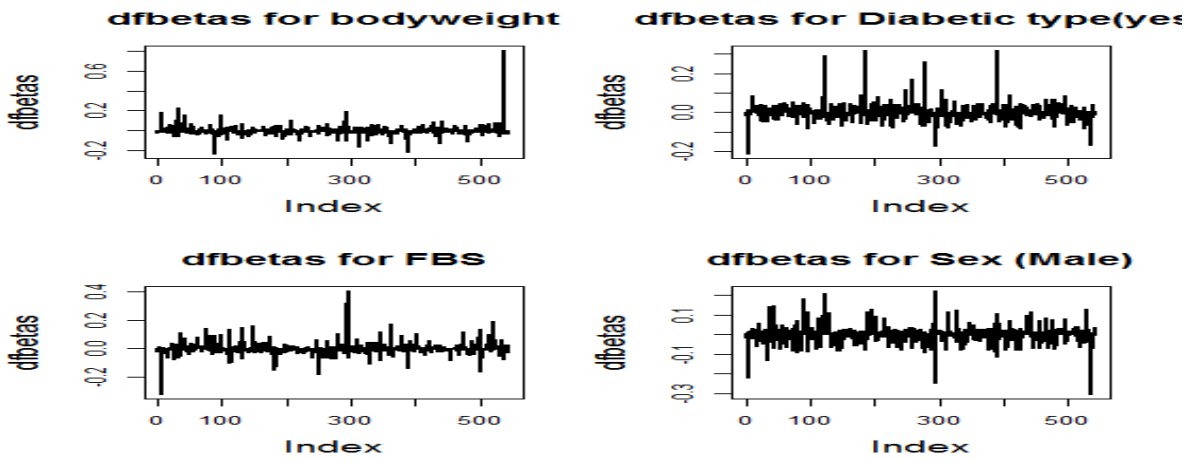


Figure 6: Influential observations for Cox PH model for DM data set

The index plots produced in Figure 6, comparing the magnitudes of the largest dfbeta values to the regression coefficients suggests that none of the observations is terribly influential individually for inseparable DM.

4.3. Separate Analysis of Type of Diabetic Mellitus

First, inseparable diabetic mellitus is modeled using Cox- PH above to identify whether diabetic types significantly influencing the recovery time of diabetic patients. Second, since, diabetic types are significantly influencing recovery time of patients, then separate (when a type of diabetic is in different data set) models for type-I and II were undertake to identify factors affecting the recovery time of patients.

4.3.1. **Modeling Recovery Time for Type-I Diabetic**

The univariate analysis (Appendix-I B) for type-I diabetic with a covariate, age group ((30-44, P-value=0.0036), (45-59, P-value=0.0577), (>60, P-value=0.0015)), bodyweight (P-value=0.0236), lower (Diastolic) (60-80, p-value=0.0572, >80, p-value=0.0934) blood pressure and fasting blood sugar (P-value=0.1) at baseline are significant with time-to-recovery at significance level of 10%. While, sex (p-value=0.467) and upper (Systolic) (110-130, p-value=0.253, >130, p=0.462) blood pressure are non-significant at 5% let leave alone10% revealing that this variable will not be included in the model and from multivariate model variables non-significant at 10%

were eliminated using backward selection method. Accordingly variables with minimum AIC are age group, bodyweight, and FBS at baseline.

Multivariate results of a Cox PH model (2) (Appendix-I C) shows that effects of age groups, bodyweight and FBS at baseline had a statistically significant impact on time-to-recovery while diastolic blood pressure is non-significant at 10% level of significance. Therefore, these variables reduced from the model. In the appendix-I C, an individual with lower DBP recovers faster than a higher DBP.

It is now observed in the table 7 that effects of age groups, bodyweight and FBS at baseline had a statistically significant impact on time-to-recovery. Therefore, age group, bodyweight and FBS at baseline are selected for the final model reducing the non-significant covariates. Again also, the likelihood ratio test (P-value=1.615e-05), Wald test (P-value=0.0002363) and score (logrank) test (P-value=5.684e-05) are highly significant.

Table 7: Multivariate Cox-PH model for type-I DM

Covariates	Coef$\left(\hat{\beta}\right)$	se(Coef$\left(\hat{\beta}\right)$)	P-value	Hazard Ratio(HR)	95%C.I for HR
Age					
30-44 years	-0.5827	0.2254	0.00972*	0.5584	(0.35899, 0.8685)
45-59 years	-0.61796	0.283	0.02919*	0.539	(0.30934, 0.9393)
> 59 years	-2.079	0.7398	0.00495 *	0.125	(0.02933, 0.5331)
bodyweight(kgs)	-0.018	0.00835	0.03358*	0.9824	(0.96648, 0.9986)
FBS(mg/dl)	-0.0028	0.0012	0.01737*	0.9972	(0.99492, 0.9995)

Likelihood ratio test= 29.8, p-value=1.615e-05, Wald test = 23.81, p=0.0002363, Score (logrank) test = 27.01, p-value=5.684e-05, AIC= 803.804.

*Coef: coefficient for covariate, HR: hazard ratio; p-value: probability value, 95%C.I HR: 95% confidence interval for HR, * Significant at 0.05 level.*

The multivariate model in table 7, the hazard of bodyweight is 0.9824, implies that for a unit increase of bodyweight in type-I DM who has not yet recovered by a certain time has 0.9824 times the chance of being recovered at the next point in time after controlling other factors in type-I DM or for a unit increase of bodyweight, the recovery time of diabetic patient delayed by 0.0176. The hazard of fasting blood sugar (FBS) is 0.9972, implies that for a unit increase of FBS at baseline in type-I DM who has not yet recovered by a certain time has 0.9972 times the

chance of being recovered at the next point in time after controlling other factors in type-I DM or for a unit increase of FBS at baseline, the recovery time of diabetic patient prolonged by 0.0028. Thus, the lower the bodyweight, age and FBS at baseline the faster the rate of the blood sugar level returning to normal range in this type of DM.

The final Cox PH model (2) for type-I DM is then given by:-

$$h_i(t) = h_0(t)\exp(-0.5827 * Age_{30-44i} - 0.61796 * Age_{45-59i} - 2.079 * Age_{60-74i} - 0.018 * bw_i - 0.0028 * FBS_i)$$

Estimated Survival model for type-I DM for is:-

$$\hat{S}(t, X) = \left[\hat{S}_0(t)\right]^{\left(\exp(-0.5827 * Age_{30-44i} - 0.61796 * Age_{45-59i} - 2.079 * Age_{60-74i} - 0.018 * bw_i - 0.0028 * FBS_i)\right)}$$

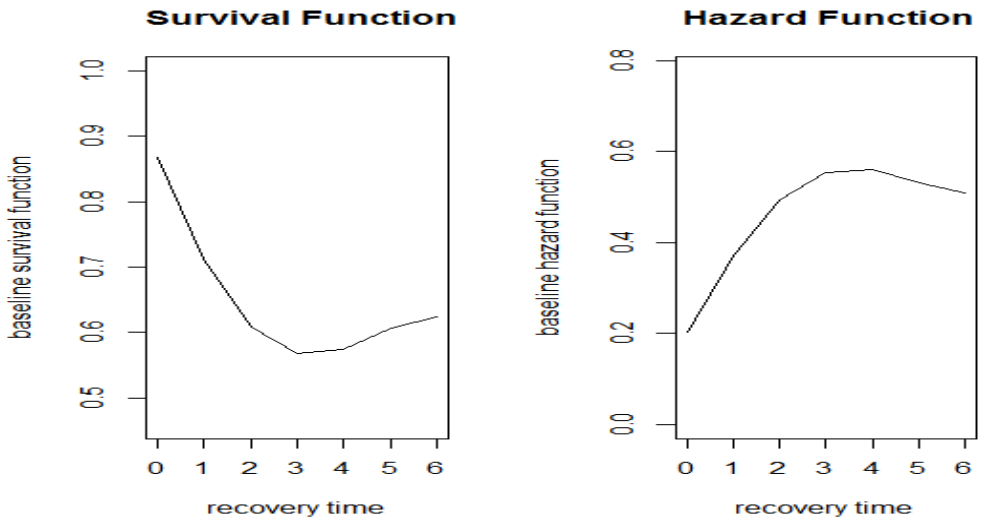


Figure 7: Estimated baseline survival and hazard in the PHs models respectively for type-I DM patients of Cox PH model

From figure 7 the baseline survival and hazards are varying with time. The patients' chance of recovery time up to 3 month decreases and increases and then increases in survival and decreases in hazard at baseline survival and hazard function respectively.

The PH assumption checking with test statistics based on the Schoenfeld residuals have been described (table 8) to test whether the correlation between Schoenfeld residual for these

covariates is zero. The PH assumption for all variables is checked and included in the model. Hence, PH assumption among age group, bodyweight and FBS variables are not violated and the global test is also not quite statistically significant.

Table 8: Cox PH assumption checking test statistics based on Schoenfeld residuals for type-I DM

Age group	rho	chisq	P
factor(z3)1	0.1175	1.3607	0.243
factor(z3)2	0.0889	0.7758	0.378
factor(z3)3	-0.0106	0.0125	0.911
bodyweight	-0.0352	0.1629	0.686
FBS	-0.0635	0.3634	0.547
GLOBAL		2.0860	0.837

4.3.1.1. **Model Diagnosis for Type-I DM Patients**

A plot of the Cox-Snell residuals against the cumulative hazard is presented for the final model in figure 8. This plot reveals that there is little evidence of a systematic deviation from the straight line, which gives us only some concern about the adequacy of the fitted models. It is very common for models with censored data to have some wiggling at large values of time and it is not something which should cause much concern.

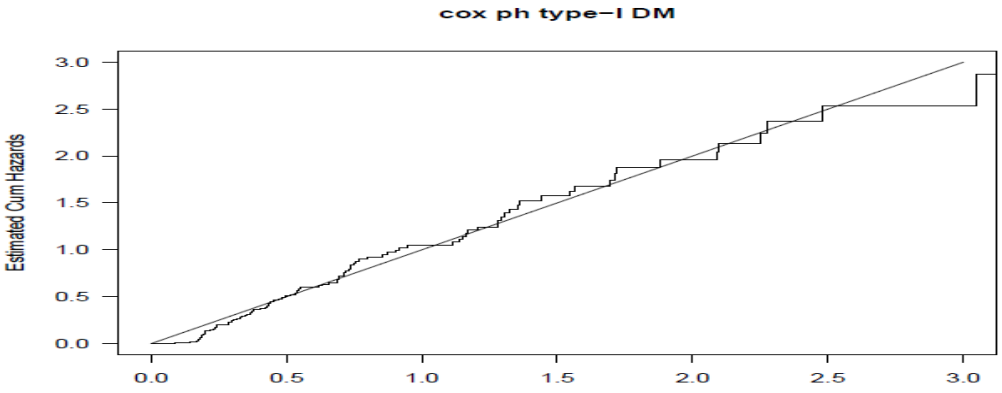


Figure 8: Cox-Snell residuals obtained from fitting Cox PH model to the Type-I Diabetic data.

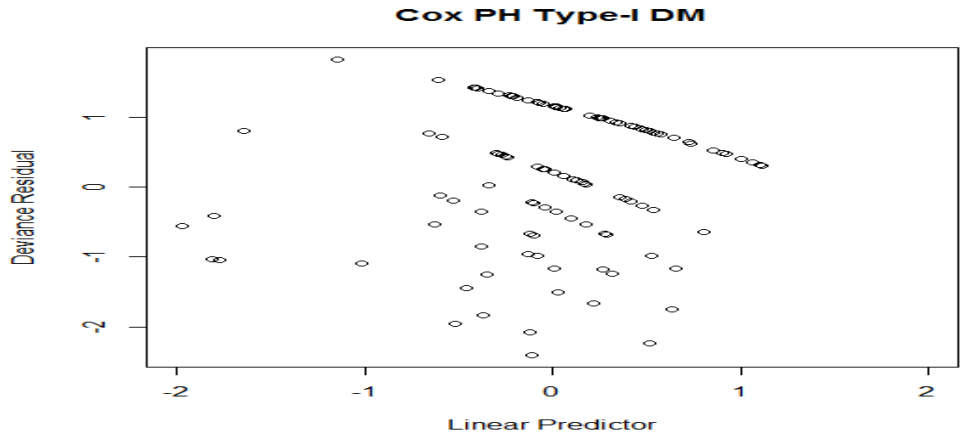


Figure 9: Deviance Residuals for Cox PH model for type-I DM

The plot of deviance residual (figure 9) shows that the deviance residuals seem to be approximately symmetrically distributed about zero and there exists some outlying observation. There is a pattern in the residuals that do not necessarily indicate any problems with the model. Therefore, we have almost some concern about the adequacy of the fitted Cox PH model.

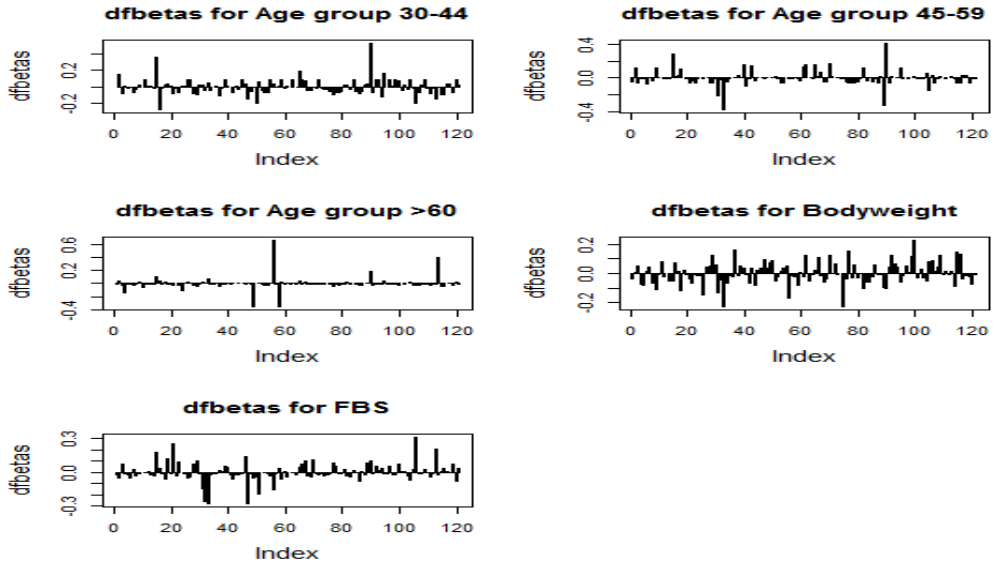


Figure 10: Influential observations for Cox PH model for Type-I DM data set

The index plots produced in figure 10, comparing the magnitudes of the largest dfbeta values to the regression coefficients suggests that none of the observations is terribly influential individually for type-I diabetic.

4.3.2. **Modeling Recovery Time for Type-II Diabetic**

The univariate analysis (Appendix-I D) for type-II with a covariate, age group ((30-44, P-value =0.05), (45-59, P-value =6.13e-05), (60-74, P-value =2.77e-11)and(>74,P-value=4.65e-09) (18-29 is as a reference)), bodyweight (P-value =1.37e-10), sex(P-value =1.83e-10) and FBS(P-value=3.92e-07) at baseline are significant with recovery time at significance level of 5% let leave alone 10% significance level.

Hence, age group, bodyweight, FBS and sex are the significant covariates associated with the recovery time of type-II diabetic mellitus whereas upper (Systolic) (110-130, P-value=0.701, >130, P-value =0.696) and lower (Diastolic) (60-80, P-value=0.907, >80, P-value=0.796) BP at baseline are not significantly associated with recovery time at 10% level of significance, hence these variables not included in multivariate analysis.

Table 9: Multivariate Cox-PH model for type-II DM

Covariates	Coef$\left(\hat{\beta}\right)$	se(Coef$\left(\hat{\beta}\right)$)	P-value	Hazard Ratio(HR)	95%C.I for HR
Age					
30-44 years	-0.3694	0.16896	0.0288 *	0.6911	(0.49629, 0.9625)
45-59 years	-0.6988	0.16351	1.92e-05*	0.4972	(0.36087, 0.6850)
60-74 years	-1.3351	0.20118	3.21e-11*	0.2631	(0.17739, 0.3903)
>74 years	-1.8210	0.35671	3.31e-07*	0.1619	(0.08045, 0.3257)
Bodyweight(kg)	-0.0212	0.00354	2.23e-09*	0.9791	(0.9723, 0.9859)
Sex (Male)	0.7588	0.12992	5.19e-09*	2.1357	(1.65563, 2.7551)
FBS(mg/dl)	-0.0031	0.00072	1.79e-05*	0.9969	(0.99553, 0.9983)

Likelihood ratio test= 175.1 , p=0.0001, Wald test =156.5, p=0.0001, Score (logrank) test = 166 , p=0.0001

*Coef: coefficient for covariate, HR: hazard ratio; p-value: probability value, 95%C.I HR: 95% confidence interval for HR, * Significant at 0.05 level.*

The results of a Cox PH model (2) fitted to this dataset were obtained on table 9. It is observed that effects of age groups (30-44, P-value= 0.0288, 45-59, P-value=1.92e-05, 60-74, P-value=3.21e-11, >74, P-value=3.31e-07), sex (P-value=5.19e-09), bodyweight (kg) (P-value= 2.23e-09) and FBS (P-value=1.79e-05) at baseline had a statistically significant impact on time-

to-recovery for type-II diabetic mellitus. A male patient recovers faster in type-II diabetic. A hazard ratio of 2.1357 corresponds to 68.11% chance of the male patient's recovered first.

The final Cox PH model (2) for type-II DM is then written by:-

$$h_i(t) = h_0(t)\exp(-0.3694 * Age_{30-44i} - 0.6988 * Age_{45-59i} - 1.3351 * Age_{60-74i} - 1.821 * Age_{>74i} - 0.0212 * bodyweight_i + 0.7588 * Sex_i - 0.0031 * FBS_i)$$

Survival model for type-II DM is:-

$$\hat{S}(t, X) = \left[\hat{S}_0(t)\right]^{\left(\begin{array}{c}\exp(-0.3694*Age_{30-44i} - 0.6988*Age_{45-59i} - 1.3351*Age_{60-74i} - 1.821*Age_{>74i} \\ -0.0212*bodyweight_i + 0.7588*Sex_i - 0.0031*FBS_i)\end{array}\right)}$$

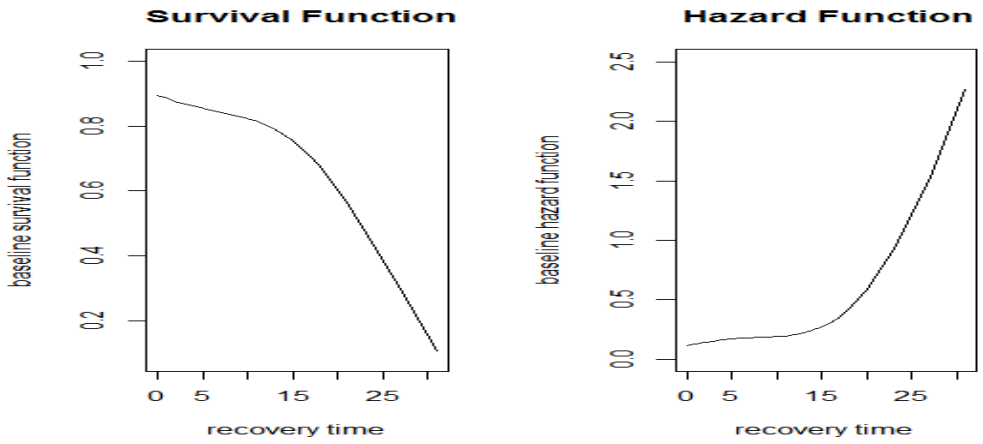


Figure 11: Estimated baseline survival and hazard in the PHs models for type-II DM patients of Cox PH model

From figure 11 the baseline survival and hazard are decreasing and increasing respectively. The patients' chance of recovery up to 15 months slowly decreases and increases at baseline survival and hazard function respectively.

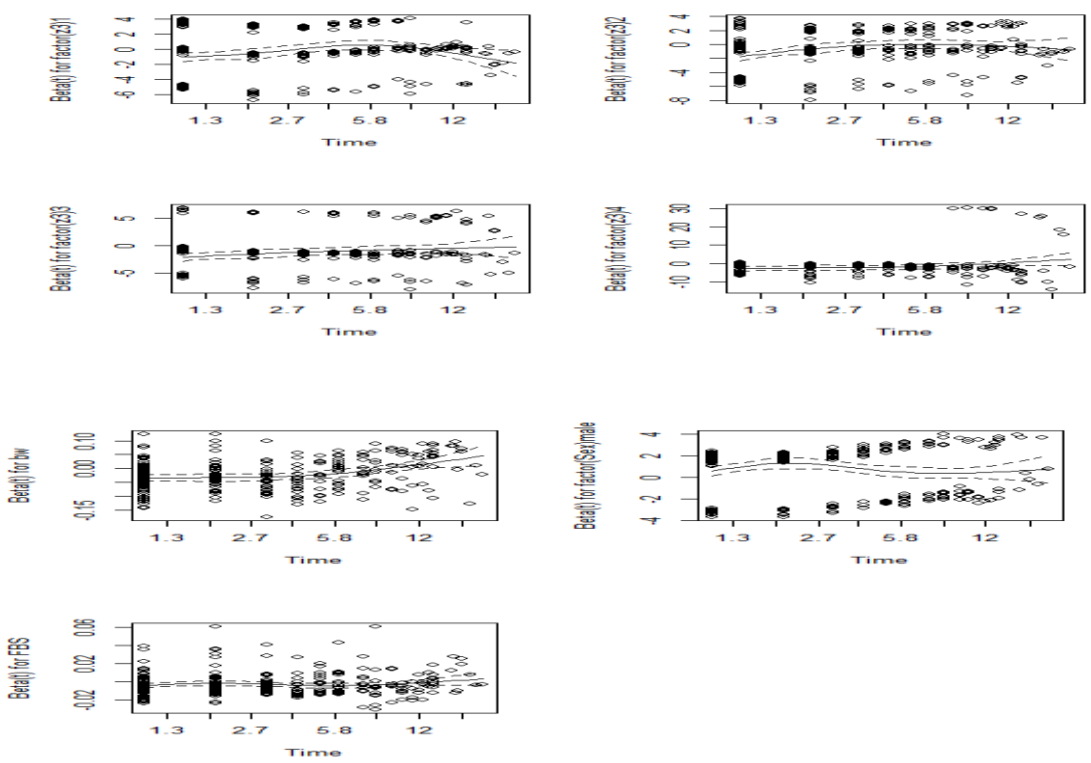


Figure 12: Plots of Scaled Schoenfeld Residuals for each Covariate in the model for type-II DM

The PH assumption checking with graphical method based on the Schoenfeld residuals have been described in figure 12. The PH assumption for all variables is checked in appendix and included in the model, implying that the proportional-hazards assumption among categorical and continues variables has not been violated.

4.3.2.1. **Model Diagnosis for Type-II DM Patients**

A plot of the Cox-Snell residuals against the cumulative hazard is presented in figure 13. This plot reveals that there is little evidence of a systematic deviation from the straight line which gives us only some concern about the adequacy of the fitted models. The hazard function follows the 45 degree line very closely except for very large values of time. It is very common for models with censored data to have some wiggling at large values of time and it is not something which should cause much concern. Overall, the final model fits the data very well.

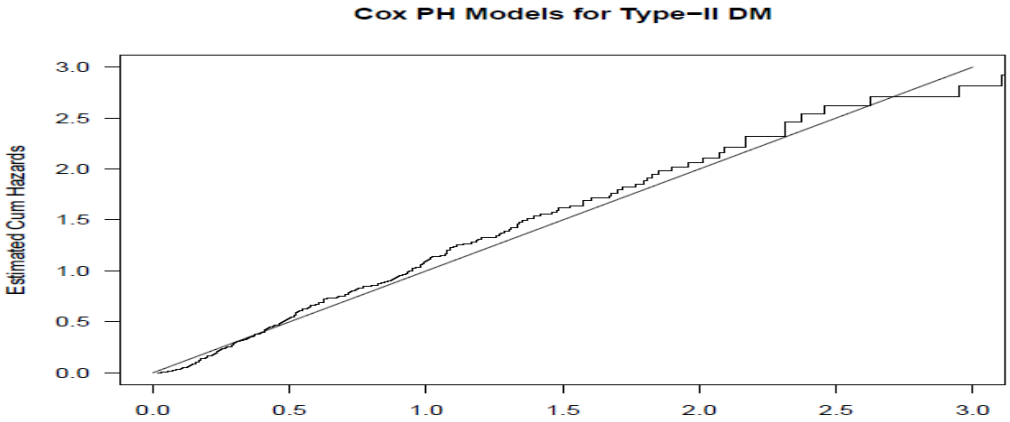


Figure 13: Cox-Snell residuals obtained from fitting Cox PH model to the type-II DM

The plot of deviance residual (figure 14) shows that the deviance residuals seem to be approximately symmetrically distributed about zero and there exists not as such clearly outlying observation. There is a pattern in the residuals that do not necessarily indicate any problems with the model. Therefore, we have almost some concern about the adequacy of the fitted Cox PH model.

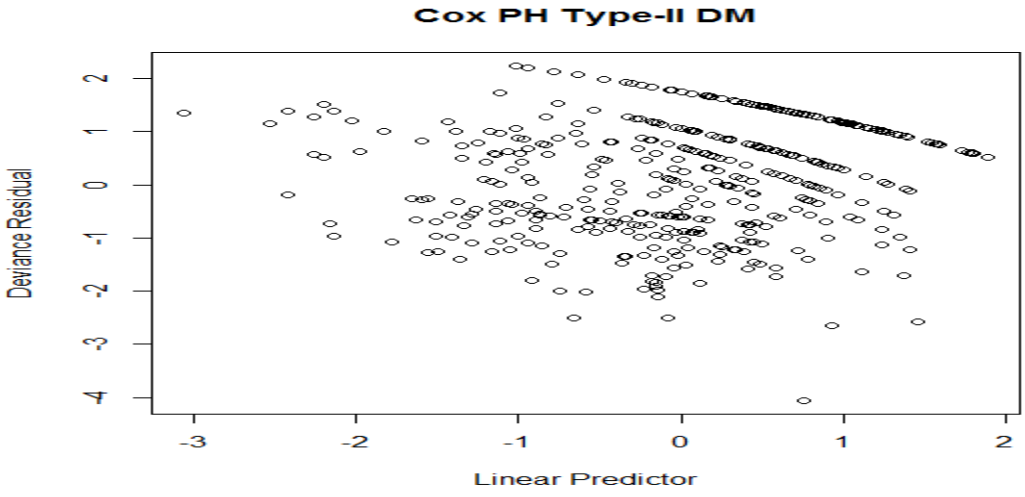


Figure 14: Deviance Residuals for Cox PH Model to the type-II DM

The index plots produced in figure 15, comparing the magnitudes of the largest dfbeta values to the regression coefficients suggests that none of the observations is terribly influential individually for type-II DM.

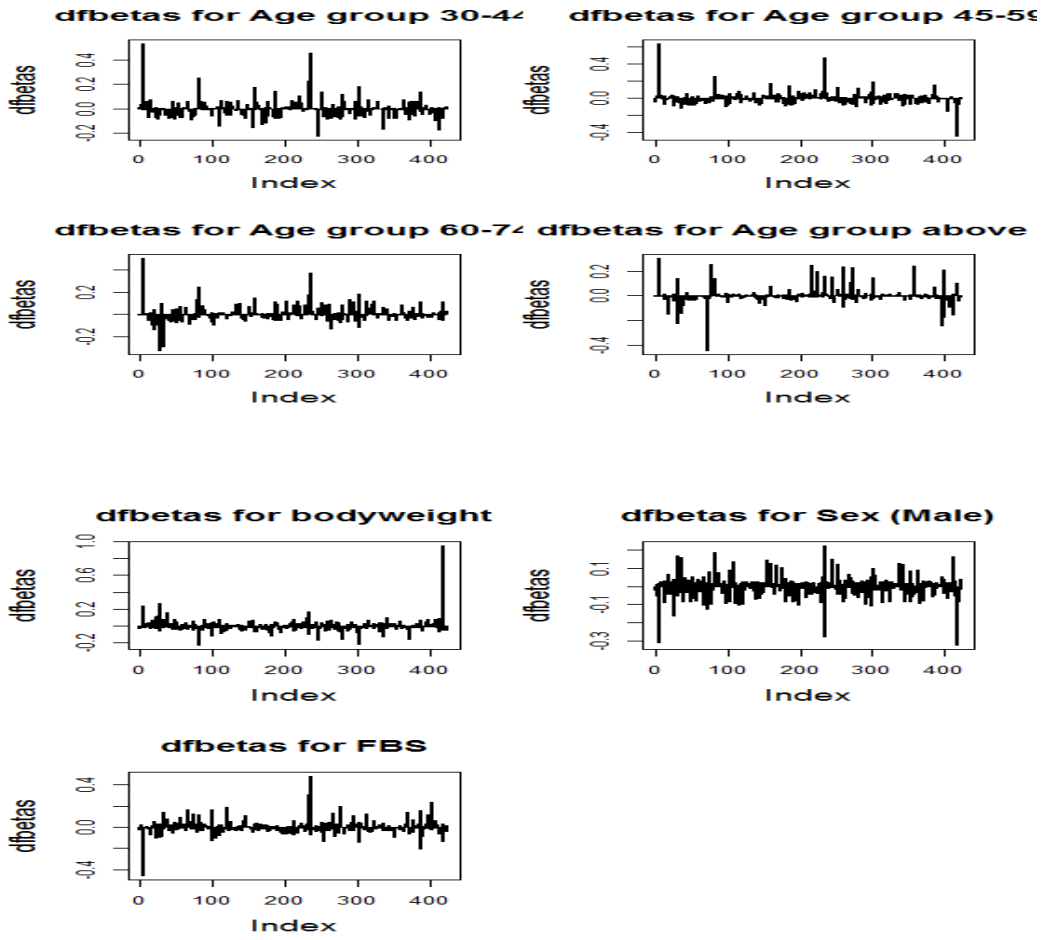


Figure 15: Influential observations for Cox PH model for type-II DM data set

4.4. **Cox-PH with Shared Gamma Frailty Model**

4.4.1. **Modeling Recovery Time of DM Inseparably**

The dependent variable used was time from start of treatment to return to normal blood sugar levels and the status variable was whether the blood sugar was in normal range or not. The districts are used to capture the random effect in the model.

In Cox-PH models, hazard function may depend on unknown or non-measurable factors which can cause the regression coefficients estimated from such models to be biased. In consequence, in order to overcome the problem and better model survival of patients, the frailty models were introduced. In fact, these models are used to explain the random variation of survival function due to unknown risk factors, such as genetic factors and numerous environmental factors.

From figure 16 below, the median first recovery time for different districts are significantly different, because the median recovery time for all districts are not the same, describe us presence of heterogeneity.

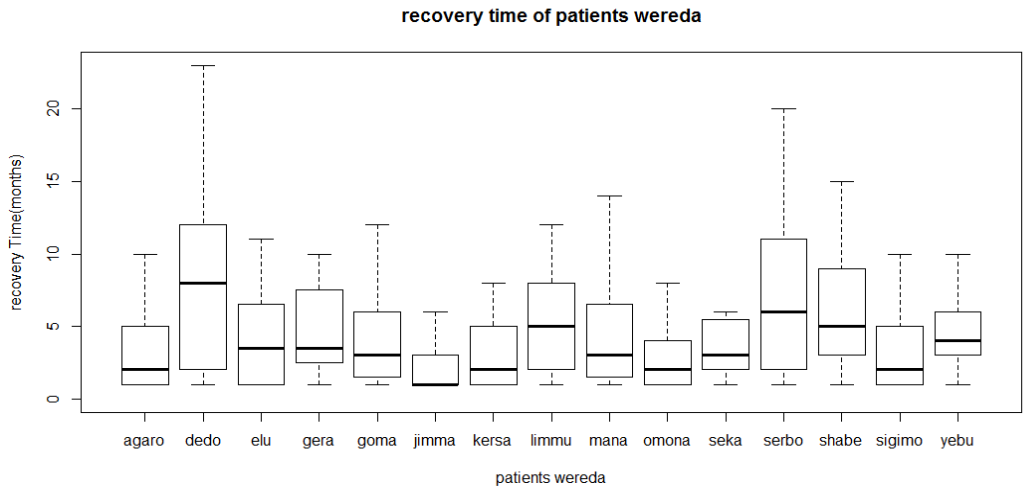


Figure 16. Box plot for recovery time (months) of patients District

Since, patients in dedo and serbo districts takes long time to recover, therefore, the median recovery times are delayed whereas patients in Jimma district, they are fast to recover into normal blood sugar level.

4.4.2. **Between-Cluster Variance Estimate**

The results in table 10 show the estimated shared gamma frailty model with random effects.

Table 10: Cluster Variance Estimate

Between-Cluster Variance Estimate		
Cluster	*Theta*	*Standard error*
District	*0.18*	*0.081*

A large value of cluster variance (theta=0.18) indicates a greater degree of heterogeneity among districts and strong association within districts.

In Cox PH frailty models (7) same to Cox PH (2) done above, first univariable (Appendix-II A) analysis were done for all variables to select variables at 10% level of significance, then variables significant at 10% were considered to fit in multivariable analysis to identify the significant variables associated with the disease accounting frailty in the model. The results in the univariable analysis (Appendix-II A) shows that age group, bodyweight at baseline, sex, fasting blood sugar at baseline and diabetic types are significantly associated with time to first recovery after controlling other prognostic factors and accounting frailty.

Table 11: Multivariable Cox-PH with Shared Gamma Frailty model for Diabetic Types Inseparably

Covariates	Coef$\left(\hat{\beta}\right)$	se(Coef$\left(\hat{\beta}\right)$)	P-value	Hazard Ratio(HR)	95%C.I for HR
Age					
30-44 years	-0.387	0.137	4.8e-03*	0.679	(0.519, 0.888)
45-59 years	-0.645	0.1418	5.4e-06*	0.525	(0.3973, 0.693)
60-74 years	-1.299	0.189	6.3e-12*	0.273	(0.1883, 0.395)
>74 years	-1.839	0.3514	1.7e-07*	0.159	(0.0799, 0.317)
Bodyweight(kg)	-0.022	0.00332	6.4e-11*	0.979	(0.9722, 0.985)
Diabetic Type (Type-II)	-0.541	0.1251	1.5e-05*	0.582	(0.4556, 0.744)
FBS	-0.0033	0.00063	2.6e-07*	0.997	(0.995, 0.998)
Sex (Male)	0.547	0.1143	1.7e-06*	1.728	(1.3809, 2.161)
Theta (θ) =0.0526, $\bar{\chi}_{01}^{2}$ = 16.79, p-value= 2.088e-05, 50 Newton-Raphson, I-likelihood = -2053, Likelihood ratio test= 287, p=0.0001, Wald test= 202 , p=0.0001, AIC= 4117.636					

*Note: standard errors of hazard ratios are conditional on theta, * Significant at 0.05 level.*

But, the upper (systolic) and lower (diastolic) blood pressure at baseline have no significant effect at 5% let leave alone 10% significance level for recovery time of diabetic patients after accounting frailty in the model the same to Cox without frailty. Let us now drop the upper

(Systolic) and lower (Diastolic) blood pressure at baseline with the aim of introducing unobserved heterogeneity. The *likelihood-ratio test* and *random effects* are significant at 5% level.

The multivariate results of a Cox PH with gamma frailty model (7) fitted to this dataset was obtained on table 11. In this table, all covariates are statistically significant at 5% level of significance. The improvement in *log-likelihood* (-2043.381) relative to the no-frailty (-2061.395) model is largest for the shared gamma frailty model.

The LR of a Cox-PH model without frailty (*log likelihood*=-2061.395) and with gamma frailty model in frailty (*I-likelihood*=-2053) is 2(2061.395-2053) =16.79, its' p-value=2.088e-05, there is a significant frailty effect, implies correlation within district cannot be ignored. In inseparable DM, gamma frailty models indicating that frailty variable (districts) is very highly significantly related to the time to first recovery of DM. Thus, there is much evidence pointing towards a population that is indicating heterogeneity.

After controlling for other prognostic factors and accounting for frailty in table 11, patients with age group 30-44, 45-59, 60-74 or >74 are recovered 0.679, 0.525, 0.273 and 0.159 times age group of 18-29. Being young is associated with better recovery. The results reveal that after accounting for heterogeneity and other confounders in the data, time to recovery takes longer time with a unit increase of FBS and bodyweight at baseline in diabetic patients. This implies that the lower the age, FBS and bodyweight at baseline the faster the rate of recovery (blood sugar level reaching the normal range) of diabetic patients. Type of diabetes has a significant effect on the life of diabetic patients. A hazard ratio of 0.582 indicates, 36.79% chance of the type-II diabetic patients recovered first as compared to type-I diabetic patients after accounting and controlling other factors in the model. Thus, an individual suffering from type-II diabetes delayed to recover as compared to type-I diabetic. And also, sex is seen to be significantly associated with the recovery time of the diabetic patient. After accounting for heterogeneity and other factors, the male diabetic recovers 1.728 times the female. That is, males recover faster than females (HR=1.728, 63.34% chance of the male diabetic patients recovered first) inseparable DM data set.

The final model then given by:-

$$h(t_{ij} / Z_i) = h_0(t_{ij}) Z_i \exp(-0.387 * Age_{30-44ij} - 0.645 * Age_{45-59ij} - 1.299 * Age_{60-74ij} - 1.839 * Age_{\succ 74ij} - 0.022 * bodyweight_{ij} - 0.541 * DT_{ij} - 0.0033 * FBS_{ij} + 0.547 * Sex_{ij})$$

and $f_z(z) = \dfrac{z^{\frac{1}{0.0526}-1} \exp(-z/0.0526)}{0.0526^{\frac{1}{0.0526}} \Gamma(\frac{1}{0.0526})}$

Data reveals moderate dependence ($\theta \sim 0.0526, kendall's\ \tau = 0.02563$). Since, Kendall's τ is 0.02563 for Cox-PH with gamma frailty, thus there is a positive correlation of 0.0263 between the recovery times of diabetic patients within district.

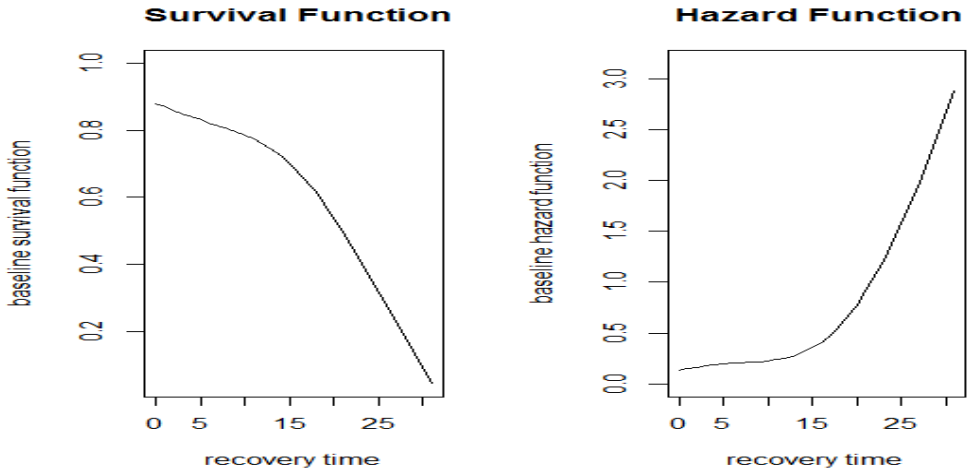


Figure 17: Estimated baseline survival and hazard in the PHs models for inseparable DM patients of Cox PH with gamma frailty model

From figure 17 above, the patients' chance of recovery time up to 15 months in districts slowly decreases and increases at baseline survival and hazard function respectively.

4.4.3. **Model Diagnosis for Diabetic Types Inseparably**

The goodness of fit by residual plots (Section 3.4.8.1) assessed. A plot of the Cox-Snell residuals against the cumulative hazard is presented in figure 18. The hazard function follows the 45 degree line very closely except for very large values of time.

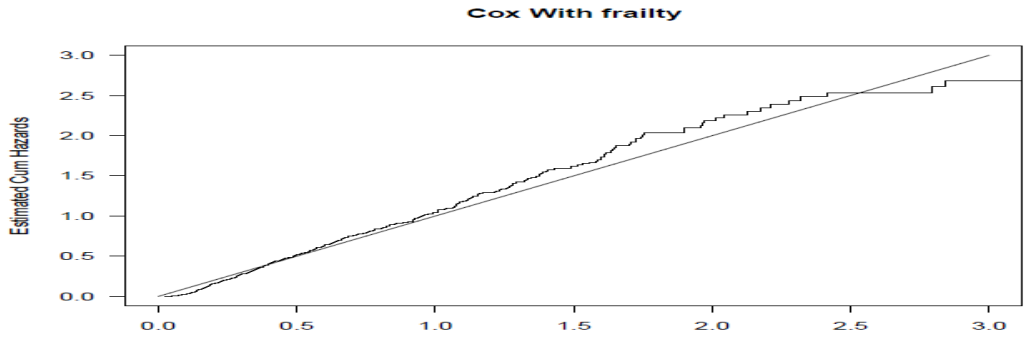


Figure 18: Cox-Snell residuals obtained from fitting Cox PH frailty model to the DM data.

It is very common for models with censored data to have some wiggling at large values of time and it is not something which should cause much concern. Overall, the final model fits the data very well.

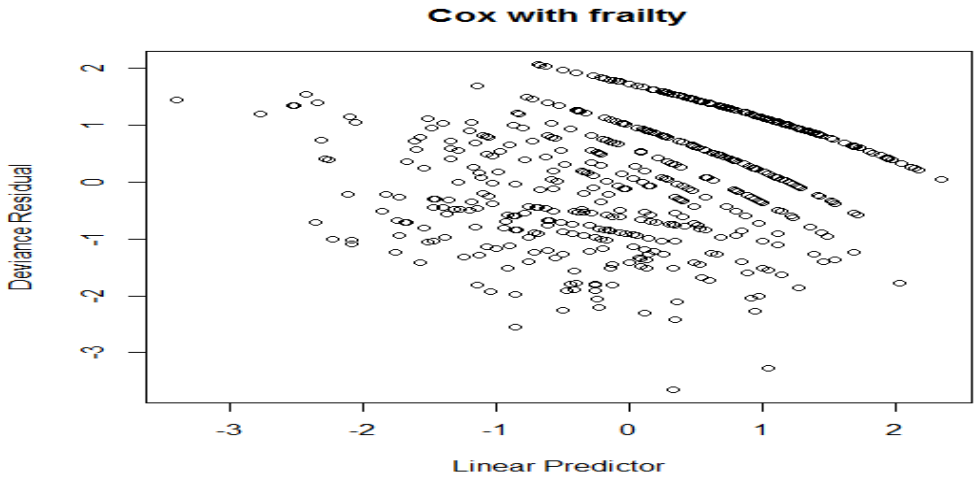


Figure 19: Deviance Residuals for Cox PH Frailty model to the DM data

The plot of deviance residual (Figure 19) shows that the deviance residuals seem to be approximately symmetrically distributed about zero and there exists not as such clearly outlying observation. There is a pattern in the residuals that do not necessarily indicate any problems with the model. Therefore, we have almost some concern about the adequacy of the fitted Cox PH with gamma frailty model of diabetic inseparably.

The index plots produced in figure 20, comparing the magnitudes of the largest dfbeta values to the regression coefficients suggests that none of the observations is terribly influential individually for inseparable DM.

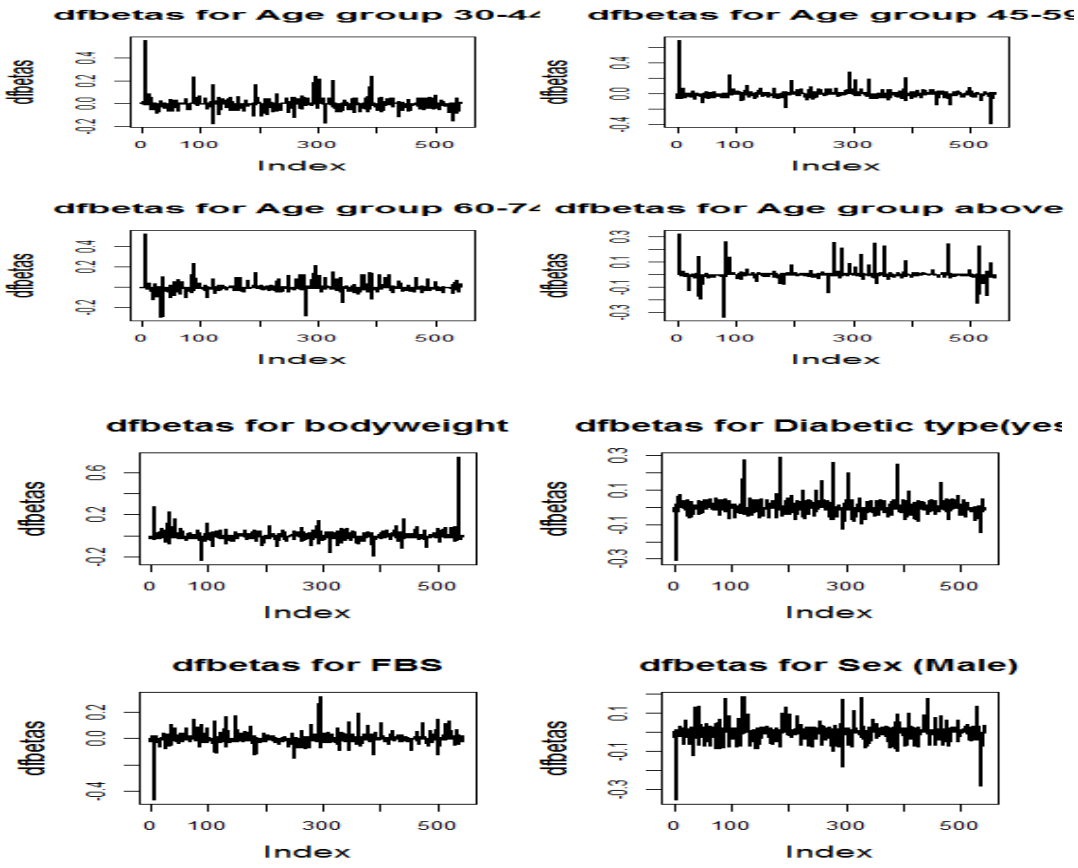


Figure 20: Influential observations for Cox PH with shared gamma frailty model for inseparable DM data set

4.5. Separately Modeling Types of Diabetic Mellitus (DM)

4.5.1. Modeling recovery time for type-I Diabetic Mellitus (DM)

In type-I DM, univariate (Appendix-II B) and multivariate analysis (Appendix-II C and Table 12) used to identify the significant variables associated with the disease accounting frailty in the model to reduce non-significant covariates. The results in the univariate; age group, bodyweight, lower (Diastolic) BP and fasting blood sugar (FBS) at baseline are significant at 10% level of significance, hence this factors associated with blood sugar level. But, sex and the upper

(systolic) blood pressure at baseline are non significant effect on the time to recovery for type-I DM patients after accounting frailty in the model at 10% level of significance the same to Cox-PH without frailty.

Let us now drop the non-significant covariates with the aim of introducing unobserved heterogeneity. Lower (Diastolic) BP is significant in univariate but, become non-significant in multivariate analysis (Appendix-II C). Hence, these covariates are reduced from the model and age group, bodyweight and FBS have a minimum AIC value.

Table 12: Multivariable Cox-PH with shared Gamma frailty model for type-I DM

Covariates	Coef$\left(\hat{\beta}\right)$	se(Coef$\left(\hat{\beta}\right)$)	P-value	Hazard Ratio(HR)	95%C.I for HR
Age					
30-44 years	-0.5673	0.23814	0.017*	0.567	(0.3556, 0.904)
45-59 years	-0.6791	0.30318	0.025*	0.507	(0.2799, 0.919)
>59 years	-2.1373	0.76105	0.005*	0.118	(0.0265, 0.524)
Bodyweight	-0.0165	0.00852	0.043*	0.984	(0.9674, 0.9985)
FBS	-0.0024	0.00127	0.048*	0.998	(0.9951, 0.9995)

Theta (θ) = 0.195, $\bar{\chi}_{01}^{2}$ =4.4, p-value= 0.01797, 28 Newton-Raphson, I-likelihood=-394.7, Likelihood ratio test= 52.5, p=4.34e-07, Wald test = 12.4, p=0.0388, AIC=794.73.

*Note: standard errors of hazard ratios are conditional on theta, * Significant at 0.05 level.*

After controlling for other prognostic factors and accounting for frailty (Table 12), patients with age group 30-44, 45-59 and >59 (years) who has not yet recovered by a certain time has 0.567, 0.507 and 0.118 times the chance of being recovered at the next point in time compared to someone in the age group 18-29 (years) in type-I DM. For a unit increase of bodyweight and FBS, patient recovery time who has not yet recovered by a certain time has 0.984 and 0.998 respectively, after accounting and controlling the effects of the other variables in the model. Thus, the lower the ages, bodyweight and FBS at baseline the faster the rate of the blood sugar level returning to normal range in type-I DM.

The LR of a Cox-PH model without frailty and with gamma frailty model is 2(396.90-394.7) = 4.4, its P-value=0.01797, there is a significant frailty effect, implies that the correlation within district cannot be ignored. In type-I DM, gamma frailty models indicating that frailty variable (districts) is very highly significantly related to time to first recovery. Thus, there is much evidence pointing towards a population that is indicating heterogeneity.

The results of a Cox PH with gamma frailty model fitted to this dataset were obtained on table 12. The hazard of the final model is then given by:-

$$h(t_{ij} / Z_i) = h_0(t_{ij}) Z_i \exp(-0.5673 * Age_{30-44ij} - 0.6791 * Age_{45-59ij} - 2.1373 * Age_{>59ij} - 0.0165 * bw_{ij} - 0.0024 * FBS_{ij})$$

and $f_z(z) = \dfrac{z^{\frac{1}{0.195}-1} \exp(-z/0.195)}{0.195^{\frac{1}{0.195}} \Gamma(\frac{1}{0.195})}$

Data reveals moderate dependence $(\theta \sim 0.195, kendall's\ \tau = 0.089)$. Since, Kendall's τ is 0.089 for Cox-PH with gamma frailty, thus there is on average a positive correlation of 0.089 between the recovery times of type-I diabetic patients within district.

From figure 21 below, the patients' chance of recovery time in districts monotonically decreases and increases at baseline survival and hazard function respectively.

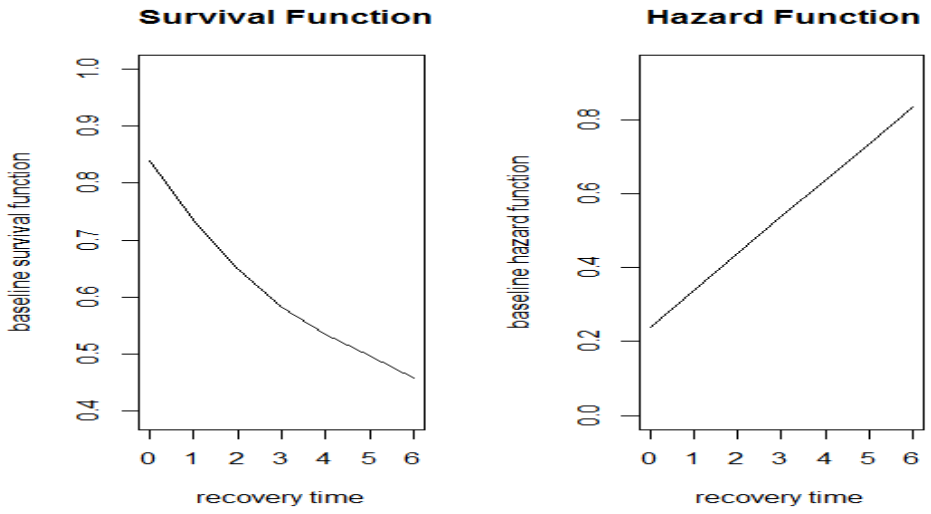


Figure 21: Estimated baseline survival and hazard in the PHs models for type-I DM patients of Cox PH with gamma frailty model

4.5.1.1. **Model Diagnosis for Type-I DM Patients**

The goodness of fit by residual plots (Section 3.4.8.1) assessed. A plot of the Cox-Snell residuals against the cumulative hazard is presented above in figure 22. The hazard function follows the 45 degree line very closely except for very large values of time. It is very common for models with censored data to have some wiggling at large values of time and it is not something which should cause much concern.

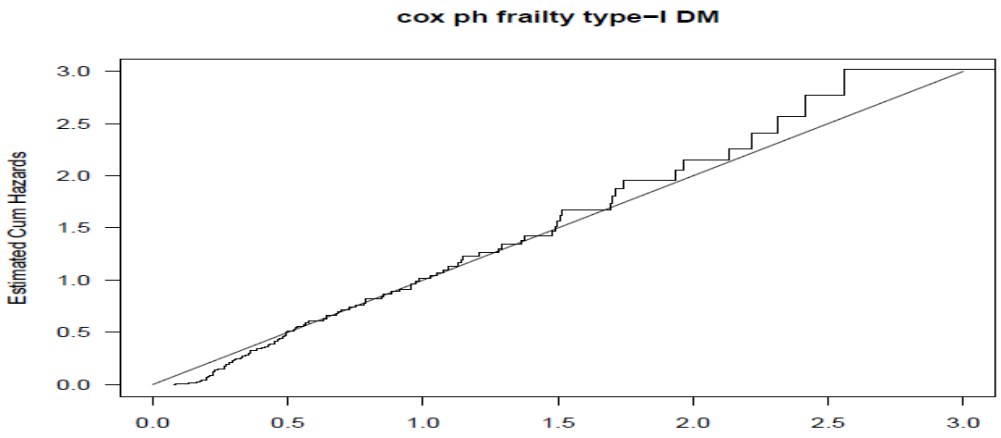


Figure 22: Cox-Snell residuals obtained from fitting Cox PH with Gamma frailty model to the type-I DM.

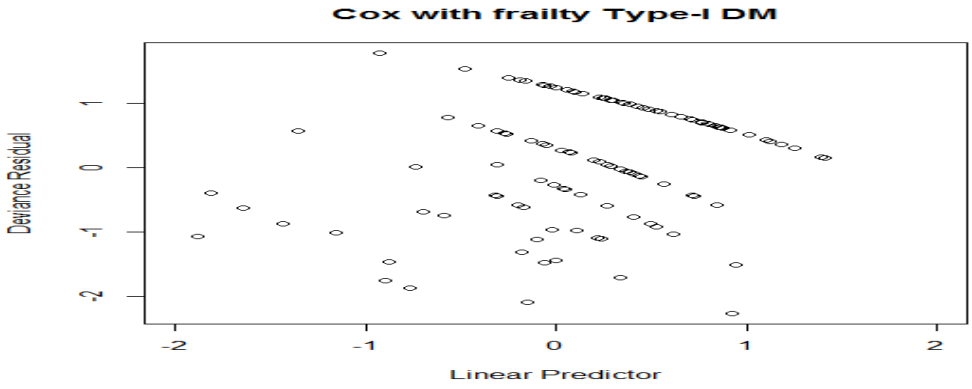


Figure 23: Deviance Residuals for Cox PH with Shared Gamma Frailty model to the type-I DM

This plot reveals that there is little evidence of a systematic deviation from the straight line which gives us only some concern about the adequacy of the fitted models. Overall, the final model fits the data very well. Again, the plot shows that the line related to the Cox-Snell

residuals of the Cox PH with frailty model were nearest to the line through the origin as compared to Cox PH without frailty (figure 8) indicating that this model describes the type-I DM dataset well.

The plot of deviance residual (figure 23) shows that the deviance residuals seem to be approximately symmetrically distributed about zero and there exists not as such clearly outlying observation. There is a pattern in the residuals that do not necessarily indicate any problems with the model. Therefore, we have almost some concern about the adequacy of the fitted Cox PH with gamma frailty model.

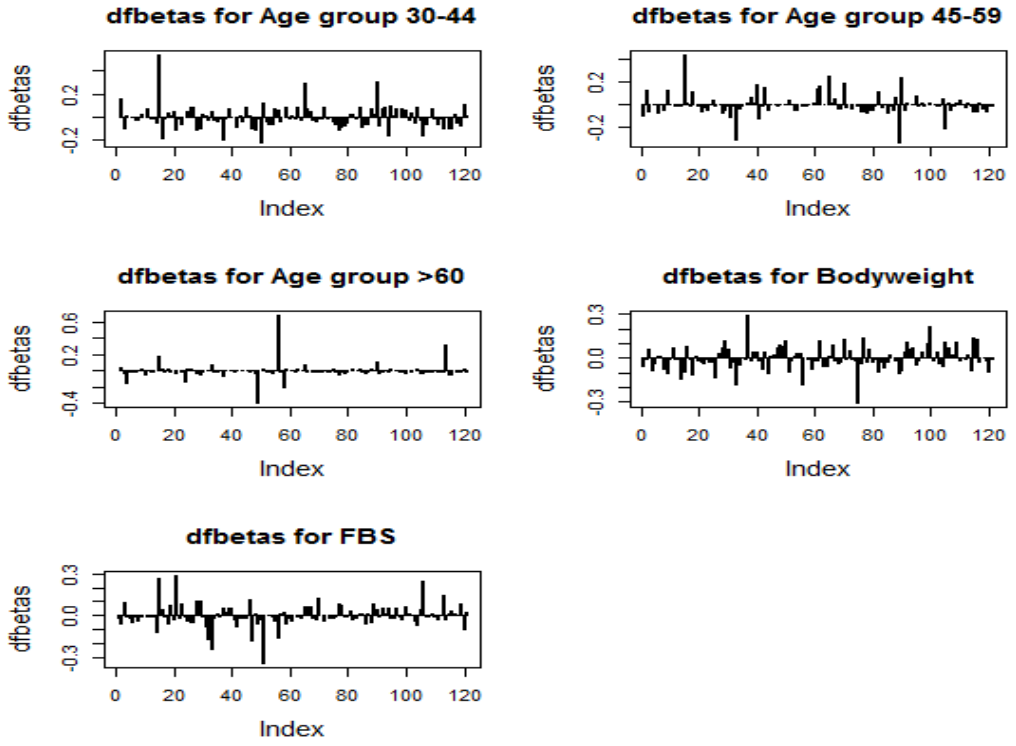


Figure 24: Influential observations for Cox PH with shared gamma frailty model in type-I DM

The index plots produced in figure 24, comparing the magnitudes of the largest dfbeta values to the regression coefficients suggests that none of the observations is terribly influential individually for type-I DM.

4.5.2. **Modeling recovery time for type-II Diabetic Mellitus (DM)**

From table 13; age group(30-44, p-value=3.0e-02,45-59, p-value=2.6e-04, 60-74, p-value=1.7e-09 and>74, p-value=8.3e-07), bodyweight(p-value=6.9e-09), sex(p-value=1.6e-07) and FBS (P-value=1.6e-05) at baseline are significant with recovery time in type-II diabetic mellitus at significance level of 5% after controlling and accounting frailty. Then we selected among variables significant at 10% in the univariate analysis (Appendix-II D).

Hence; age group, bodyweight, FBS at baseline and sex are the significant covariates associated with the recovery time of type-II DM same to Cox PH without frailty in multivariate analysis. Rates of recovery time takes longer time in females than in male patients (table 13) in type-II DM same in Cox PH. After controlling for other prognostic factors and accounting for frailty, patients with age group 30-44, 45-59, 60-74 and >74 (years) who has not yet recovered by a certain time has 0.687, 0.543, 0.288 and 0.165 times the chance of being recovered at the next point in time compared to someone in the age group 18-29 (years) in type-II DM. Thus, the lower the ages the faster the rate of the blood sugar level returning to normal range in type-II DM.

Table 13: Multivariable Cox-PH with shared gamma frailty model for type-II DM

Covariates	Coef$\left(\hat{\beta}\right)$	se(Coef$\left(\hat{\beta}\right)$)	P-value	Hazard Ratio(HR)	95%C.I for HR
Age					
30-44 years	-0.3759	0.17331	3.0e-02*	0.687	(0.4889, 0.964)
45-59 years	-0.6113	0.16733	2.6e-04*	0.543	(0.3909, 0.753)
60-74 years	-1.2443	0.20668	1.7e-09*	0.288	(0.1922, 0.432)
>74 years	-1.8038	0.36596	8.3e-07*	0.165	(0.0804, 0.337)
Bodyweight(kg)	-0.0216	0.0037	6.9e-09*	0.979	(0.9716, 0.986)
Sex (Male)	0.7058	0.13455	1.6e-07*	2.026	(1.5560, 2.637)
FBS(mg/dl)	-0.00324	0.00075	1.6e-05*	0.997	(0.9953, 0.998)
Theta (θ)= 0.0661, $\bar{\chi}_{01}^{2}$ = 12.8, p-value= 0.00017, 49 Newton-Raphson, I-likelihood = -1451.1, Likelihood ratio test= 207, p=0.0001, Wald test = 132, p=0.0001, AIC=2911.5					

*Note: standard errors of hazard ratios are conditional on theta, * Significant at 0.05 level.*

The LR of a Cox-PH model without frailty and gamma frailty model is 2(1457.50-1451.1) =2*6.4=12. 8, it's P-value = 0.00017, there is a significant frailty effect, implies the correlation within district cannot be ignored. In type-II DM, gamma frailty models indicating that frailty

variable (districts) is very highly significantly related to the time to first recovery of type-I DM. Thus, there is much evidence pointing towards a population that is indicating heterogeneity.

The multivariate results of a Cox PH with gamma frailty model fitted to this dataset were obtained on table 13. The final model is then given by:-

$$h(t_{ij} / Z_i) = h_0(t_{ij}) Z_i \exp(-0.3759 * Age_{30-44ij} - 0.6113 * Age_{45-59ij} - 1.2443 * Age_{60-74ij} - 1.8038 * Age_{>74ij} - 0.0216 * bodyweight_{ij} + 0.7058 * Sex_{ij} - 0.00324 * FBS_{ij})$$

and $$f_z(z) = \frac{z^{\frac{1}{0.0661}-1} \exp(-z/0.0661)}{0.0661^{\frac{1}{0.0661}} \Gamma(\frac{1}{0.0661})}$$

Data reveals moderate dependence $(\theta \sim 0.0661, kendall's\ \tau = 0.032)$. Since, Kendall's τ is 0.032 for Cox-PH with gamma frailty, thus there is on average a positive correlation of 0.032 between the recovery times of type-II diabetic patients within district.

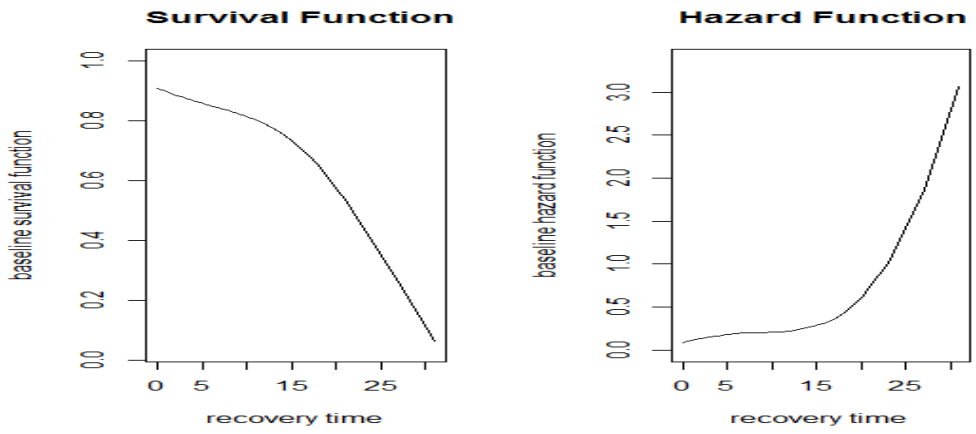


Figure 25: Estimated baseline survival and hazard in the PHs models for type-II diabetic patients of Cox PH with gamma frailty model.

From figure 25 above, the patients' chance of recovery time up to 15 months in districts slowly decreases and increases at baseline survival and hazard function respectively.

4.5.2.1. **Model Diagnosis for Type-II DM Patients**

A plot of the Cox-Snell residuals against the cumulative hazard is presented in figure 26. The hazard function follows the 45 degree line very closely except for very large values of time. It is

very common for models with censored data to have some wiggling at large values of time and it is not something which should cause much concern. This plot reveals that there is little evidence of a systematic deviation from the straight line which gives us only some concern about the adequacy of the fitted models. Overall, the final model fits the data very well. Again, the plot shows that the line related to the Cox-Snell residuals of the Cox PH with frailty model were nearest to the line through the origin as compared to Cox PH without frailty (figure 13) indicating that this model describes the type-II DM dataset well.

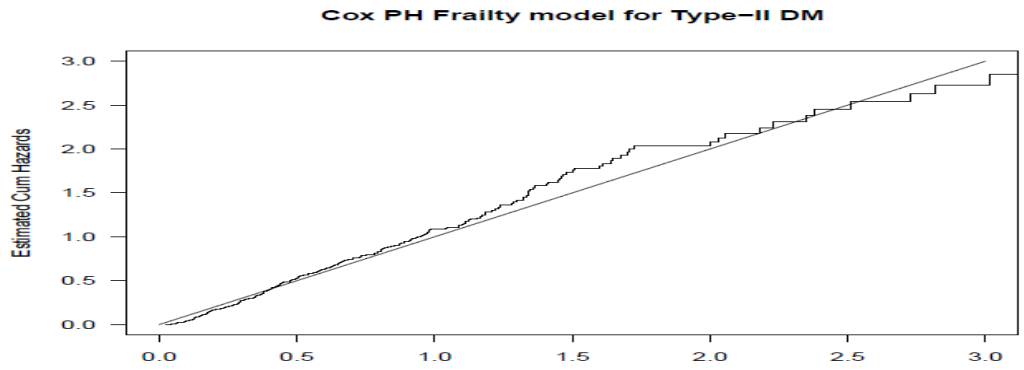


Figure 26: Cox PH with Shared Gamma Frailty Model for type-II DM

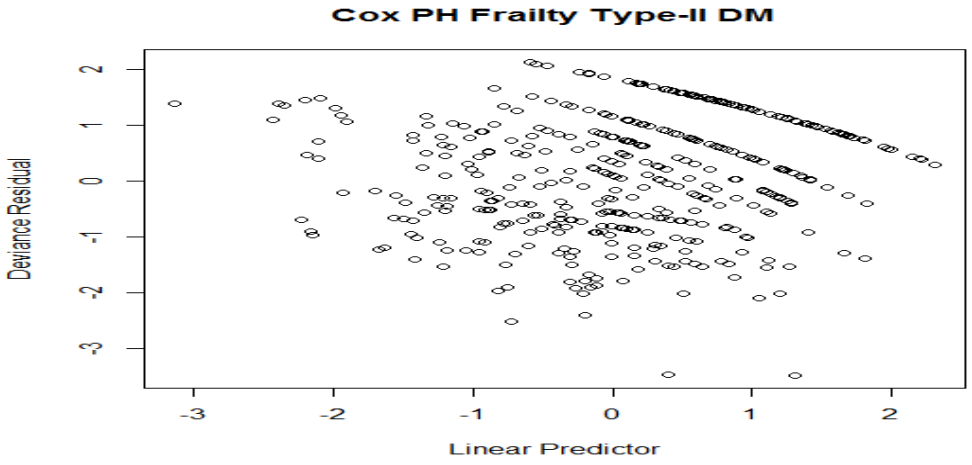


Figure 27: Deviance Residuals for Cox PH with Shared Gamma Frailty Model for type-II DM

The plot of deviance residual (Figure 27) shows that the deviance residuals seem to be approximately symmetrically distributed about zero and there exists not as such clearly outlying

observation. There is a pattern in the residuals that do not necessarily indicate any problems with the model. Therefore, we have almost some concern about the adequacy of the fitted Cox PH with gamma frailty model for type-II DM.

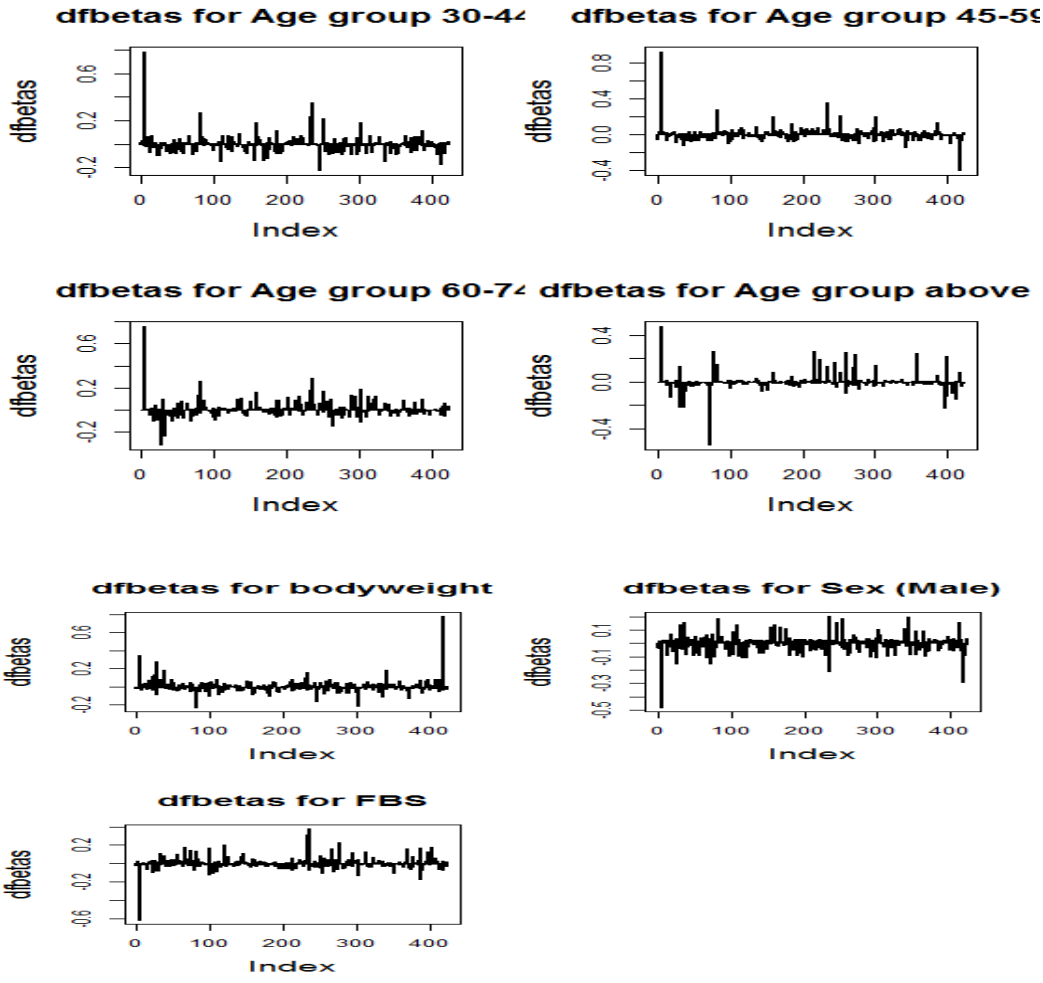


Figure 28: Influential observations for Cox PH with shared gamma frailty model in type-II DM

Also, delta-beta statistic for Cox PH with shared gamma frailty is used to measure the influential observations on the model as a whole (Figure 28). That is, comparing the magnitudes of the largest dfbeta values to the regression coefficients suggests that none of the observations is terribly influential individually. Therefore, we do not remove them from the dataset and conclude that there are no influential observations.

4.6. **Comparison of Cox PH versus Shared Gamma Frailty Models**

In this study, in order to compare the efficiency of the models the AIC (Akaike's Information Criterion) was used.

Table 14. Comparison of Cox PH and Shared Gamma Frailty Models for separate and inseparable DM

	Types of DM		Types of Diabetic Mellitus Separately				
Disease	inseparably		Type-I DM		Type-II DM		
Model	Log-like (model)	AIC	Log-like (model)	AIC	Log-like (model)	AIC	Rank
Cox-PH without frailty	*-2061.395*	*4138.8*	*-396.9*	*803.8*	*-1457.50*	*2929*	*2*
Cox-PH with gamma frailty	*-2043.38*	*4117.6*	*-385.55*	*794.73*	*-1441.677*	*2911.5*	**1**

DM: Diabetic Mellitus, Log-like: Log-likelihood, AIC: Akaike's Information Criterion

The AIC is a criterion that assesses goodness of fit of a statistical model, and the lower value of AIC suggests a better model. Table 14 gives the log-likelihood and AIC values of the two models telling the Cox with gamma frailty to be the most powerful one in predicting recovery time of DM types separately and generally when compared to Cox without frailty model.

5. DISCUSSION AND CONCLUSION

5.1. Discussion

Diabetes Mellitus (DM) is a life-long challenge that requires behavioral change and adequate self-care practices for better glycaemic control. In the absence of appropriate self-care practice, the desired therapy targets are difficult, or even impossible to achieve. Glucose control is almost entirely in the hands of the patient who lives with this condition. Several factors are known in various studies as influencing factors. In the literature, there are many studies on the field of diabetic, but researchers tend to examine the effects of covariates on patients using logistic regression (Endalew H. et al (2012)) and chi-square (Charlton, et al (2005) and Endalew H. et al (2012)) model. If patients' recovery time differs, results of Cox model are seriously under question. Hougaard (1995) points out that the impact of unmeasured covariates can lead to transformation of the hazard function and the coefficients of the measured covariates. A model that is becoming increasingly popular for modeling association between individual survival times within subgroups is the use of a frailty model. Beard (1959), Vaupel, et al. (1979), and Lancaster (1979) suggested a random effects model in order to account for the unobserved heterogeneity due to unobserved covariates. Here, the frailty represents the total effect on survival of the covariates not measured when collecting information on group of subjects. The frailty distributions most often applied are the gamma distribution.

The main goal of this study was modeling time-to-first recovery of adult diabetic patients of Jimma University Specialized Hospital using Cox PH and shared gamma frailty models using numerous factors such as gender, age, bodyweight, diabetic type , upper (systolic) blood pressure, lower (diastolic) blood pressure and fasting blood sugar (FBS) at baseline using district as a cluster. The outcome response is time to first recovery.

From the total of 544, 404(74.26%) experienced the event and the rest 140 (25.74%) loss to follow-up from the study. In type-I diabetic from the total of 121, 103 (85.12%) experienced the event and the rest 18(14.88%) loss to follow-up and in type-II diabetic from the total of 423, 301 (71.16%) experienced the event and the rest 122 (28.84%) loss to follow-up from the study. The 95% C.I for type-I and type-II diabetic were 1&2 and 4&5 months respectively but, it varies depending on the covariates included in the study. The minimum and the maximum recovery time

of type-I diabetic were 1 and 6 months and for type-II diabetic 1 and 31 months respectively. More men than women are in diabetes; 69.42 per cent compared with 30.58 per cent in those with type-I diabetes and 63.12 percent compared with 36.88 percent in those with type-II diabetes this is consistent with DM in state of UK (2011/12).

The mean bodyweight, age, and FBS at baseline for diabetic mellitus patients in type-I were 55.52kg, 36 years, & 209.34 mg/dl respectively whereas, in type-II were 63.38kg, 47 years, and 217.999 mg/dl respectively. Patients with type-I diabetic are young (36 years), lower bodyweight (55.52kg) and lower FBS (209.34mg/dl) at baseline as compared to type-II diabetic (47 years old, 63.38 kg and 217.999 mg/dl). These results showed the age and FBS at baseline for type-II and type-I DM are older and higher respectively as compared to cross- sectional study which was conducted in Jimma, South Western Ethiopia and Mekele, Northern Ethiopia. This may reflect different nutritional and environmental influences as well as methods of data collection and analysis may vary.

The PH assumption checking with graphical method based on the Schoenfeld residuals have been described and included in the model. The graphs for some of the categorical variables displayed using Kaplain Meir (figure 1) appeared were parallel; implying that the proportional-hazards assumptions for all variables have not been violated.

In Cox-PH with and without shared gamma frailty models of inseparable diabetic types; age group, bodyweight, diabetic type, FBS, and sex of patients at baseline shows a statistically significant association with time to first recovery to normal blood sugar level. In univariable and multivariable analysis of Cox-PH with and without frailty models, the types of diabetic was a strong and independent prognostic factor, indicating better recovery time for type-I patients controlling other factors in the model. This means that patients with type-II getting affected by diabetic mellitus prolonged recovery time as compared to type-I; these findings are consistent with those done in Uganda countries by Olive, D., et al., (2007).

Olive, D., et al., (2007), showed that the five factors determined as significant by *p*-value were age, bodyweight, sex, type of diabetic and family history of diabetic. In this study; age group, bodyweight, sex, FBS at baseline and diabetic type for two models are found to be significant

with the Wald statistics 222.2 (P-value=0.0001) for the Cox PHs model (2) without frailty and *202 (P-value= 0.0001)* for the gamma frailty model (7).

In separate analysis of type-II DM, Cox-PH with and without shared gamma frailty models; age group, bodyweight, FBS and sex of patients at baseline shows a statistically significant association with time to first recovery whereas in type-I DM, age group, bodyweight and fasting blood sugar of patients at baseline shows a statistically significant association with time to recovery. The results further reveal that the upper (systolic) and lower (diastolic) blood pressure has no significant effect for type-II and type-I diabetic patients, while for type-I diabetic; sex is not significantly associated. In Cox-PH with and without shared gamma frailty models, the lower the ages, bodyweight and FBS at baseline the faster the rate of the blood sugar level returning to normal range in type-I & II DM.

In Cox-PH with shared gamma frailty model (7) the recovery time of an individual suffering from type-II diabetes who has not recovered yet has 0.586 (HR=0.586, 36.95% chance of the type-II diabetic patients recovered first) times as compared to type-I DM. Being female prolonged the recovery time as compared to males (HR=2.026, 66.953% chance of the male diabetic patients recovered first) in type-II DM. In Cox-PH with and without shared gamma frailty models gender was a strong and independent prognostic factor in univariable and multivariable analysis, indicating males are better recovering to normal blood sugar as compared to females' inseparable diabetic types and type-II diabetic. This means that females getting affected by diabetic mellitus (DM) have a slightly takes longer time to recover to normal blood glucose level than males, these findings are consistent with those again obtained in Ugandan countries by Olive, D., et al., (2007). When the frailty is ignored, the estimate for β and its estimated error is smaller compared to the shared gamma frailty model (7). This is expected as the frailty model account for the extra variance associated with unmeasured risk factors.

The heterogeneity parameter θ estimated and kendall's τ for each type of diabetic and inseparable DM are $(\theta \sim 0.0526, \tau = 0.02563)$ (inseparable diabetic), $(\theta \sim 0.195, \tau = 0.089)$ (in type-I), and $(\theta \sim 0.0661, \tau = 0.032)$ (type-II) and the likelihood-ratio test of, $H_0 : \theta = 0$ are rejected with P-value 0.01797(type-I), 0.00017(type-II) and 2.088e-05 (inseparable diabetic) , implies that the correlation within district cannot be ignored.

The goodness of fit by residual plots assessed using the Cox-Snell residuals, deviance residuals and influential observations. Overall, the final model fits the data very well. In this study, in order to compare the efficiency of models the AIC (Akaike Information Criterion) (section 3.4.6.2) and log-likelihood were used. The AIC is a criterion that assesses goodness of fit of a statistical model, and the lower value of AIC suggests a better model. Cox with shared gamma frailty model is the smallest AIC as compared to without shared gamma frailty model in inseparable, type-I and II DM patients. Additionally, the Cox-Snell residuals of the Cox-PH with shared gamma frailty models are nearest to the line through the origin as compared to without frailty, indicating that these models fit the data best. Therefore, the Cox-PH with shared gamma frailty model is the most powerful one in predicting recovery time of diabetic patients when compared to without frailty in diabetic mellitus inseparably, in type-I and type-II.

5.2. Conclusion

The study considers diabetic mellitus inseparably (when types of diabetic are in one data set) & individual types of Diabetic Mellitus (when types of diabetic are separately analyzed) at JUSH. The aim of this study was to suggest a better model to analyze recovery time for types of diabetic mellitus (DM) patients among Cox-PH with and without shared gamma frailty using districts (weredas) as clustering.

The proportional hazards assumptions were hold, indicating a good fit. Cox-PH with shared gamma frailty models indicates that frailty variable (districts) is very highly significantly related to the time to first recovery for DM in both univariable and multivariable analysis. In univariable and multivariable analysis of Cox-PH with and without shared gamma frailty models the type of diabetic was a strong and independent prognostic factor, indicating better recovery time for type-I patients accounting and controlling other factors. This means that patients with type-II getting affected by diabetic mellitus delayed the recovery time as compared to type-I.

AIC and log-likelihood were used to evaluate the performance among models. Based on AIC and log-likelihood, the Cox-PH with shared gamma frailty model provides a suitable choice for the life time model of first recovery time for Diabetic Mellitus (DM) as compared to Cox-PH without frailty model. Also, the Cox-Snell residuals of the Cox-PH with shared gamma frailty models are nearest to the line through the origin as compared to without frailty, indicating that these models

fit the data best. Therefore, the Cox-PH with shared gamma frailty model is the most powerful one in predicting recovery time of diabetic patients when compared to without frailty in diabetic mellitus inseparably, in type-I and type-II. For Cox PH with and without shared gamma frailty models influential observations in the magnitudes of the largest delta-beta statistics suggests that none of the observations terribly influential, individually.

In type-I diabetic a significant factor associated with first recovery time were age group, bodyweight and FBS at baseline and in type-II diabetic patients significant factors associated with first recovery time were age group, sex, FBS and bodyweight at baseline both in Cox-PH with and without shared gamma frailty models. These covariates are important factors that should take into consideration when selecting a treatment method for both types of DM.

Covariates significant in the Cox-PH model (2) are also significant in the Cox-PH with shared gamma frailty model (7) both in univariable and multivariable analysis. The covariates that increased time to recovery in Cox-PH model also increased in Cox-PH with shared gamma frailty model and vise versa. Being old, female, higher FBS and overweight at baseline prolonged the recovery time. In Cox-PH model the estimate for β and its estimated error is smaller compared to the shared gamma frailty models (7). This is expected as the frailty model account for the extra variance associated with unmeasured risk factors.

I. REFERENCES

1. Alberti K.G.M.M, P.Z. Zimmet for the WHO Consultation, (1998). Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications Part1: Diagnosis and Classification of Diabetes Mellitus Provisional Report of a WHO Consultation.
2. Andersen, P. K., Klein, J. P., Knudsen, K. M., & y Palacios, R. T. (2013). Estimation of variance in Cox's regression model with shared gamma frailties. *Biometrics.*
3. Beard, R.E. (1959). Note on some mathematical mortality models, in: C.E.W. Wolstenholme, Connor (eds.). The life span of animals. CIBA Foundation Colloquia on Ageing.
4. Charlton, W. (2005). Diabetes outcomes in the Indian health system during the Era of the special diabetes program for Indians and the government performance the results act.
5. Chaudhry, Gannon, Nuttall, (2006). Conducted a thesis on stability of body weight in type 2 diabetes.
6. Clayton, D., Cuzick, J. (1985). The semi-parametric Pareto model for regression analysis of survival times. Proceedings of the Centenary Session of the International Statistical Institute, Amsterdam.
7. Commenges, D., Andersen, PK. (1995). Score test of homogeneity for survival data. Lifetime Data Analysis.
8. Connor, H. and Boulton, A. J. M. (1989). Diabetes in Practice. Chichester: Wiley
9. Cox, (1972). Regression models and life tables (with discussion) Journal of the Royal Statistical Society B.
10. Cox and Snell, (1968). A general definition of residuals (with discussion).
11. Cox, D.R. and Oakes, D. (1984). Analysis of survival data. London: Chapman and Hall.
12. Dereje, A. (2005). Diabetes Mellitus, Diploma program for the Ethiopian health center team.
13. Duchateau & Janssen, (2007). The frailty Model.
14. Duchateau, L. and Janssen, P. (2004). Penalized partial likelihood for frailties and smoothing splines in time to first insemination models for dairy cows. Biometrics.
15. Efron, B. (1982). The Jackknife, the Bootstrap and Other Re-sampling Plans, Society for Industrial and Applied Mathematics, Philadelphia, PA.

16. Emmanuel, V., Sean, HC., Stephen, PM., Fracp (1991-2005). Incidences, Treatments, Outcomes, and Sex Effect on Survival in Patients with End Stage Renal Disease by Diabetes Status in Australia and New Zealand.
17. Endalew, H., H. Mariam, W., Belachew, T., Birhanu, Z. (2012). Self-care practice and glycaemic control amongst adults with diabetes at the Jimma University Specialized Hospital in south-west Ethiopia: A cross-sectional study.
18. Erik, (1998). Asymptotic Theory for the Correlated Gamma-Frailty Model.
19. Flinn, C.J. and Heckman, J.J., (1982). New methods for analyzing individual event histories.
20. Gebregziabher, M., Egede, LE., Lynch, CP., Echols, C., Zhao, Y.(2010). Effect of Trajectories of Glycemic Control on Mortality in Type II Diabetes.
21. Ghadimi, (2011). Prognostic factors for the survival of patients with esophageal cancer in Northern Iran.
22. Green, P. J. (1987). Penalized likelihood for general semi-parametric regression models. International Statistical Review.
23. Greenwood, M., Yule, G.U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. Journal of the Royal Statistical Society.
24. Gutierrez, R. G., S. Carter, and D. M. Drukker. (2001). on boundary-value likelihood-ratio tests.
25. Gutierrez, R. (2001). Frailty Models on STATA.
26. Gyimah, (2003). Frailty models with applications to the study of infant deaths on birth timing in Ghana and Kenya.
27. Heckman and Singer (1982). The identification problem in econometric models for the duration data.
28. Henderson, R. and Oman, J. P. (1999). Effect of frailty on marginal regression estimates in survival. Journal of the Royal Statistical Society.
29. Hipszer, B. R. (2001). Type 1 diabetic model (A).
30. Hougaard, P.(1986). Survival models for heterogeneous populations derived from stable distributions.

31. Hougaard, P. (1995). Frailty models for survival data. Lifetime Data Analysis.
32. Hougaard, P., Myglegaard, P., and Borch-Johnsen, K. (1994). Heterogeneity models of disease susceptibility, with application to diabetic nephropathy.
33. Guariguata, L., Whiting, D. R., Hambleton, I., Beagley, J., Linnenkamp, U., & Shaw, J. E. (2013). Global estimates of diabetes prevalence for 2013 and projections for 2035 for the IDF Diabetes Atlas. *Diabetes Research and Clinical Practice*.
34. James, N., Moser, Marvin, (2002). Undiagnosed or untreated type1 diabetes of people loses bodyweight and increase blood pressure.
35. John Fox, (2002). Cox Proportional-Hazards Regression for Survival Data Appendix to an R and S-PLUS Companion to Applied Regression.
36. Joly, P., Commenges, D., Letenneur, L. (1998). A penalized likelihood approach for arbitrarily censored and truncated data:
37. Kains, H. & Bontempi, (2006). Data Analysis and Modeling Techniques Survival Analysis.
38. Kalbfleisch, J. D., and Prentice, R. L. (1963). Marginal likelihoods based on Coxís regression and life model.
39. Kamoi, K. (2011). Prevention for Micro-and Macro-Vascular Complications in Diabetic Patients.
40. Kaplan and Meier. (1958). Non-parametric estimation from incomplete observations. Journal of American Statistical Association.
41. Keyfitz, N. and Littman, G. (1979). Mortality in a heterogeneous population.
42. Klein, JP. (1992). Survival analysis: State of the art.Kluwer Academic; Boston, Massachusetts: Estimating random effects in the Framingham heart study.
43. Lancaster, T. (1979). Econometric Methods for the Duration of Unemployment. Econometrica.
44. Leong, W.Y. (2007). Survival analysis of Diabetic patients in queen Elizabeth hospital, kota kinabalu.
45. Lester, (1980). The Clinical Pattern of Diabetes Mellitus in Ethiopians.
46. Nielsen, GG., Gill, RD., Andersen, PK., Sorensen, T. (1992). A counting process approach to maximum likelihood estimation in frailty models.

47. O'Sullivan, F. (1988). Fast computation of fully automated log-density and log-hazard estimators.
48. Olive D., Odwee O and Atuhaire, K. (2007). The determinants of recovery time of diabetic patients from three hospitals in Uganda.
49. Pickles, A., Bolton, P., Macdonald, H., Bailey, A., Le CA, Sim CH. (1995). Latent-class analysis of recurrence risks for complex phenotypes with selection and measurement error: A twin and family history study of autism.
50. Rondeau, V., and Commenges, D. (2000). Relationship between aluminum concentrations in drinking water and Alzheimer's disease: an 8-year follow-up study.
51. Rozenfeld, Y., Jacquelyn, S., Plauschinat, C., and Ken, S. (2008). Oral Antidiabetic Medication Adherence and Glycemic Control in Managed Care.
52. Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model.
53. Therneau, T.M and P.M. Grambsch (2000). Modelling Survival Data: Extending the Cox Model. Springer-Verlag, New York.
54. Therneau, T.M. and P.M. Grambsch (2002). Penalized Cox models and frailty. Working manuscript.
55. Trussel and Richards (1983). Correction for unobserved heterogeneity in hazard models.
56. Tuma (1976). Rewards, Resources and rate of mobility. A non-stationary multivariate stochastic model.
57. Tuma (1979). Dynamic analysis of event histories.
58. Type-I Diabetes (www.thelancet.com, 2010).
59. Ulviya, A. (2013). Frailty Models for Modeling Heterogeneity.
60. Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography.
61. Yashin, A., Iachine, A., Alexander, Z., and Vaupel, W. (2000). Hidden Frailty: Myths and Reality.
62. Zhou, Hanson and Jara (2010). A covariate-adjusted frailty Cox model for clustered failure time data.

APPENDIX-I

Modeling Recovery Time for Inseparable Diabetic Mellitus (Dm)

A. Univariable Analysis using Cox PH Model for DM Dataset

Covariates	Coef$\left(\hat{\beta}\right)$	se(Coef$\left(\hat{\beta}\right)$)	P-value	Hazard Ratio(HR)	95%C.I for HR
Age					
30-44 years	-0.4266	0.1342	0.00148 *	0.6527	(0.50170, 0.8491)
45-59 years	-0.7926	0.1360	5.62e-09 *	0.4527	(0.34674, 0.5909)
60-74 years	-1.5320	0.1817	< 2e-16 *	0.2161	(0.15137, 0.3085)
>74 years	-2.2596	0.3365	1.89e-11 *	0.1044	(0.05398, 0.2019)
Bodyweight(kg)	-0.024278	0.003061	2.22e-15 *	0.976	(0.9702, 0.9819)
Systolic blood pressure(SBP)					
110-130 mm/Hg	-0.06583	0.13147	0.617	0.9363	(0.7236, 1.211)
>130 mm/Hg	-0.15208	0.16564	0.359	0.85961	(0.6208, 1.188)
Diastolic blood Pressure(DBP)					
60-80 mm/Hg	-0.0817	0.4127	0.843	0.9215	(0.4104, 2.069)
>80 mm/Hg	-0.0868	0.4229	0.837	0.9169	(0.4003, 2.100)
Diabetic Type (Type-II)	-0.9227	0.1194	1.07e-14 *	0.3974	(0.3145, 0.5022)
Sex (Male)	0.6605	0.1097	1.71e-09 *	1.936	(1.561, 2.4)
FBS(mg/dl)	-0.0035119	0.0006365	3.44e-08 *	0.9965	(0.9953, 0.9977)

Modeling Recovery Time for Type-I Diabetic Mellitus (Dm)

B. Univariable Analysis using Cox PH Model for Type-I DM

Covariates	Coef$\left(\hat{\beta}\right)$	se(Coef$\left(\hat{\beta}\right)$)	P-value	Hazard Ratio(HR)	95%C.I for HR
Age					
30-44 years	-0.6505	0.22342	0.0036 *	0.5218	(0.33677, 0.8085)
45-59 years	-0.5357	0.2823	0.0577*	0.585	(0.33655, 1.0177)
>60 years	-2.32013	0.7309	0.0015 *	0.09826	(0.02345, 0.4116)
Bodyweight(kg)	-0.016107	0.007117	0.0236 *	0.984	(0.9704, 0.9978)
Systolic blood pressure(SBP)					
110-130 mm/Hg	-0.3030	0.2653	0.253	0.7386	(0.4391, 1.242)
>130 mm/Hg	-0.2628	0.3577	0.462	0.7689	(0.3814, 1.550)
Diastolic blood Pressure(DBP)					
60-80 mm/Hg	-1.3791	0.7251	0.0572*	0.2518	(0.06079, 1.043)
>80 mm/Hg	-1.2776	0.7614	0.0934.	0.2787	(0.06266, 1.240)
FBS	-0.00181	0.00111	0.100*	0.9982	(0.996, 1)
Sex (Male)	-0.155	0.2132	0.467	0.8564	(0.5639, 1.301)

C. Multivariate Cox-PH Model for the Selected Variable

Covariates	Coef$\left(\hat{\beta}\right)$	se(Coef$\left(\hat{\beta}\right)$)	P-value	Hazard Ratio(HR)	95%C.I for HR
Age					
30-44 years	-0.578017	0.22583	0.01048 *	0.561	(0.3604, 0.8734)
45-59 years	-0.6124	0.28624	0.03239*	0.542	(0.3093, 0.9499)
60-74 years	-2.04026	0.7412	0.00591 *	0.130	(0.0304, 0.5556)
Bodyweight(kg)	-0.018406	0.0084	0.02757 *	0.9818	(0.9658, 0.998)
FBS	-0.00283	0.00119	0.01756 *	0.9972	(0.9948, 0.9995)
Diastolic blood Pressure(DBP)					
60-80 mm/Hg	-1.2020	0.7277	0.09857.	0.3006	(0.0722, 1.2513)
>80 mm/Hg	-1.071	0.76644	0.16230	0.3427	(0.07629, 1.539)
Likelihood ratio test= 31.91, P-value=4.213e-05, Wald test = 26.71 p-value=0.0003758, Score (logrank) test = 30.59, p-value=7.41e-05					

*Coef: coefficient for covariate, HR: hazard ratio; p-value: probability value, 95%C.I HR: 95% confidence interval for HR, * Significant at 0.05 level.*

Modeling Recovery Time for Type-II Diabetic Mellitus (Dm)

D. Univariable Analysis using Cox PH Model for Type-II DM Data set

Covariates	Coef$\left(\hat{\beta}\right)$	se(Coef$\left(\hat{\beta}\right)$)	P-value	Hazard Ratio(HR)	95%C.I for HR
Age					
30-44 years	-0.3297	0.1682	0.05.	0.7192	(0.51721, 1.000)
45-59 years	-0.6470	0.1614	6.13e-05 *	0.5236	(0.38158, 0.7185)
60-74 years	-1.3266	0.1992	2.77e-11 *	0.2654	(0.17958, 0.3921)
>74 years	-2.0219	0.3451	4.65e-09 *	0.1324	(0.06732, 0.2604)
Bodyweight(kg)	-0.022119	0.003445	1.37e-10 *	0.9781	(0.9715, 0.9848)
Systolic blood pressure(SBP)					
110-130 mm/Hg	-0.05834	0.15171	0.701	0.9433	(0.7007, 1.270)
>130 mm/Hg	-0.07328	0.18753	0.696	0.9293	(0.6435, 1.342)
Diastolic blood Pressure(DBP)					
60-80 mm/Hg	0.05913	0.50530	0.907	1.061	(0.3941, 2.856)
>80 mm/Hg	0.13318	0.51471	0.796	1.142	(0.4166, 3.133)
Sex (Male)	0.8075	0.1267	1.83e-10 *	2.242	(1.749, 2.874)
FBS(mg/dl)	-0.0038	0.00075	3.92e-07 *	0.9962	(0.9947, 0.9977)

APPENDIX-II

Modeling Recovery Time for Inseparable Diabetic Mellitus (Dm)

A. Univariable Analysis Using Cox PH with Shared Gamma Frailty Model for DM Dataset

Covariates	Coef$\left(\hat{\beta}\right)$	se(Coef$\left(\hat{\beta}\right)$)	Chisq	P-value	Hazard Ratio(HR)	95%C.I for HR
Age						
30-44 years	-0.455	0.138		1.0e-03	0.635	(0.4842, 0.832)
45-59 years	-0.769	0.140		4.1e-08	0.464	(0.3523,0.610)
60-74 years	-1.477	0.186		1.9e-15	0.228	(0.1586, 0.329)
>74 years	-2.157	0.342		3.0e-10	0.116	(0.0591, 0.226)
Theta	0.0735		39.1	9.7e-06		
Bodyweight(kg)	-0.0248	0.00324		1.9e-14	0.976	(0.969, 0.982)
Theta	0.146		58.8	1.7e-08		
Systolic blood pressure(SBP)						
110-130 mm/Hg	-0.0977	0.133		4.6e-01	0.907	(0.698, 1.18)
>130 mm/Hg	-0.2598	0.169		1.2e-01	0.771	(0.554, 1.07)
Theta	0.173		64.72	2.1e-09		
Diastolic blood Pressure(DBP)						
60-80 mm/Hg	-0.029	0.418		9.4e-01	0.971	(0.428, 2.21)
>80 mm/Hg	-0.125	0.430		7.7e-01	0.882	(0.380, 2.05)
Theta	0.169		64.06	2.6e-09		
Family History present	0.942	0.122		9.9e-15	2.56	(2.02, 3.26)
Theta	0.17		64.9	1.9e-09		
Diabetic Type (Type-II)	-0.942	0.122		9.9e-15	0.39	(0.307, 0.495)
Theta	0.17		64.9	1.9e-09		
Sex (Male)	0.574	0.113		3.8e-07	1.77	(1.42, 2.21)
Theta	0.111		50.1	3.3e-07		
FBS	-0.0031	0.00065	22.4	2.2e-06	0.997	(0.996, 0.998)
Theta	0.124		53.0	1.3e-07		

Modeling Recovery Time for Type-I Diabetic Mellitus (Dm)

B. Univariable Analysis using Cox PH with Gamma Frailty Model for Type-I DM Dataset

Covariates	Coef$\left(\hat{\beta}\right)$	se(Coef$\left(\hat{\beta}\right)$)	Chisq	P-value	Hazard Ratio(HR)	95%C.I for HR
Age						
30-44 years	-0.685	0.234		0.0034*	0.5040	(0.3188, 0.797)
45-59 years	-0.626	0.309		0.0420*	0.5347	(0.2921, 0.979)
60-74 years	-2.377	0.751		0.0015*	0.0928	(0.0213, 0.404)
Theta	0.266		19.39	0.016		
FBS(mg/dl)	-0.00144	0.0012		0.10*	0.999	(0.996, 1)
Theta	0.233		17.62	0.0250*		
Bodyweight(kg)	-0.017	0.0074		0.021*	0.983	(0.969, 0.997)
Theta	0.255		17.62	0.014*		
Systolic blood pressure(SBP)						
110-130 mm/Hg	-0.316	0.281		0.260	0.729	(0.420, 1.26)
>130 mm/Hg	-0.254	0.382		0.510	0.776	(0.367, 1.64)
Theta	0.264		19.71	0.015*		
Diastolic blood Pressure(DBP)						
60-80 mm/Hg	-1.16	0.734		0.110	0.313	(0.0743, 1.32)
>80 mm/Hg	-1.05	0.780		0.180	0.350	(0.0759, 1.61)
Theta	0.25		18.79	0.019*		
Sex (Male)	-0.159	0.229		0.490	0.853	(0.545, 1.34)
Theta	0.268		19.54	0.017*		

C. Multivariable Analysis using Cox PH with Shared Gamma Frailty Model for Type-I DM Dataset

Covariates	Coef$\left(\hat{\beta}\right)$	se(Coef$\left(\hat{\beta}\right)$)	P-value	Hazard Ratio(HR)	95%C.I for HR
Age					
30-44 years	-0.56414	0.23827	0.0180*	0.569	(0.3566, 0.907)
45-59 years	-0.66536	0.30441	0.0290*	0.514	(0.2831, 0.934)
60-74 years	-2.10596	0.76255	0.0057*	0.122	(0.0273, 0.543)
FBS	-0.00243	0.00127	0.057*	0.998	(0.9951, 1.000)
Bodyweight(kg)	-0.01695	0.00855	0.047*	0.983	(0.9669, 1.000)
Diastolic blood Pressure(DBP)					
60-80 mm/Hg	-1.00816	0.73276	0.170.	0.365	(0.0868, 1.534)
>80 mm/Hg	-0.91881	0.78164	0.24.	0.399	(0.0862, 1.846)
30 Newton-Raphson, Theta(θ)=0.177, P-value=0.051, I-likelihood = -394, Likelihood ratio test= 53, p=1.07e-06, Wald test = 23.5, p=0.0393					

Modeling Recovery Time for Type-II Diabetic Mellitus (Dm)

D. Univariate Analysis using Cox PH with Shared Gamma Frailty Model for Type-II DM Data Set

Covariates	Coef$\left(\hat{\beta}\right)$	se(Coef$\left(\hat{\beta}\right)$)	Chisq	P-value	Hazard Ratio(HR)	95%C.I for HR
Age						
30-44 years	-0.369	0.176		3.6e-02*	0.691	(0.4899, 0.976)
45-59 years	-0.629	0.168		1.9e-04*	0.533	(0.3835, 0.742)
60-74 years	-1.268	0.206		7.4e-10*	0.281	(0.1878, 0.421)
>74 years	-1.965	0.355		3.3e-08*	0.140	(0.0699, 0.281)
Theta	0.115		41.50	5.5e-06		
Bodyweight(kg)	-0.0231	0.0037		4.5e-10*	0.977	(0.97, 0.984)
Theta	0.169		54.3	7.5e-08		
Systolic blood pressure(SBP)						
110-130 mm/Hg	-0.103	0.154		5.0e-01	0.902	(0.666, 1.22)
>130 mm/Hg	-0.219	0.192		2.5e-01	0.803	(0.552, 1.17)
Theta	0.182		58.59	1.5e-08		
Diastolic blood Pressure(DBP)						
60-80 mm/Hg	0.1524	0.513		7.7e-01	1.16	(0.426, 3.18)
>80 mm/Hg	0.0964	0.523		8.5e-01	1.10	(0.395, 3.07)
Theta	0.178		57.90	1.9e-08		
Sex (Male)	0.707	0.132		8.5e-08*	2.03	(1.57, 2.63)
Theta	0.104		40.3	6.8e-06		
FBS(mg/dl)	-0.0034	0.000773		1.1e-05*	0.997	(0.995, 0.998)
Theta	0.139		48.7	4.9e-07		