THE RELATION BETWEEN DIABETES WITH HYPERTENSION AND OTHER CARDIOVASCULAR RISK FACTORS USING LOGIT AND PROBIT MODEL

BY

ABRAHAM YEBOAH ACHEAMPONG

A Thesis submitted to the Department of Mathematics,

Kwame Nkrumah University of Science Technology

In partial fulfillment of the requirements for the degree of

MASTER OF PHILOSOPHY

College of Science

JUNE 2011

DECLARATION

I hereby declare that apart from certain documentary and other sources which have cited and duly acknowledged, this submission is the result of my own research and that it has neither in whole nor in part been presented to any institution or organization anywhere for award of degree. I also declare that my supervisor nor any other person but the author alone is responsible for whatsoever errors and omissions that might appear in the work.

ABRAHAM YEBOAH ACHEAMPONG	…….	……
(20069271)	Signature	Date

Certified by:

MR. K. F. DARKWAH	…….	……
Supervisor's Name	Signature	Date

DR. BENJAMIN A. EGHAN	……	……...
Co-Supervisor's Name	Signature	Date

MR. F. K. DARKWAH	……..	……...
Head of Department	Signature	Date

ABSTRACT

The study is aimed at finding out the relationship between diabetes with hypertension and other cardiovascular risk factors among diabetic patients at Komfo Anokye Teaching Hospital, Kumasi. The main purpose is to model the relation between hypertensive diabetic and significant risk factors. Among the 310 subjects randomly selected, 198 were hypertensive diabetics with 147(74%) females. The data analysis was executed by SPSS with logit and probit analysis used for modeling. Findings revealed that the best predictor variables for diabetes with hypertension using the above analysis are age, body mass index (BMI) and blood glucose level (BGL) with their respective odds ratios 1.113, 1.266 and 0.937. Based on this, two similar models- logit and probit are developed for assessing the risks associated with hypertensive diabetic and also to make future predictions. Logit (Y) = 1n $(\frac{\pi}{1-\pi})$ = -10.581+ 0.107Age + 0.236BMI + (-0.065BGL) and

$$\text{Probit}(\pi_i) = \Phi[(-3.758) + 0.026\text{Age} + 0.058\text{BMI} + (-0.024\text{BGL})]$$

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

ADA	American Diabetes Association
AHA	American Heart Association
BGL	Blood Glucose Level
BMI	Body Mass Index
CHD	Coronary Heart Disease
CVD	Cardiovascular Disease
DBP	Diastolic Blood Pressure
DM	Diabetes Mellitus
FBG	Fasting Blood Glucose
FHD	Family History of Diabetes
GDA	Ghana Diabetes Association
HDL-C	High-density lipoprotein- Cholesterol
IFG	Impaired Fasting Glucose
IGT	Impaired Glucose Tolerance
IDF	International Diabetes Federation
KATH	Komfo Anokye Teaching Hospital
LDL-C	Low-density Lipoprotein- Cholesterol
OR	Odd Ratio
SBP	Systolic Blood Pressure
TG	Triglyceride
WHO	World Health Organization
WHR	Waist Hip Ratio

DEDICATION

This piece of work is dedicated to all diabetic patients who attend clinic at Komfo Anokye Teaching Hospital and members of Ghana Diabetes Association.

ACKNOWLEDGEMENT

My first appreciation goes to the Almighty God who has blessed me with life and wisdom to come this far. This research would not have been successful without my supervisors, Mr. K. F. Darkwah and Dr. B. A. Eghan. I am grateful to them for their supervision, encouragement, suggestions, guidance and discussions which influenced the completion of this thesis. For my able lecturers, I wish to express my sincere appreciation for their knowledge and experiences they shared with me for the period.

My profound gratitude goes to the chairman, Osomfour Prof. Sir J. W. Acheampong and members of Committee on Human Research Publication and Ethics of School of Medical Sciences/Komfo Anokye Teaching Hospital, Kumasi for their cooperation.

I acknowledge the doctors, nurses and all the diabetic screening team of Komfo Anokye Teaching Hospital for their generous assistance especially Mad. Esther Adufah Mad Sunkwa.

I am highly indebted to all the members of my family for their financial support and encouragement. It is not possible to credit all who have contributed in diverse ways towards the accomplishment of this research. However, I would like to give a particular recognition to my course mates and supportive friends, who were cooperative and helpful.

May God bless you!

CHAPTER 1

INTRODUCTION

1.1 Background to the Study

The study is about making inferences of data on patients suffering from diabetes with hypertension and its related cardiovascular risk factors. For this purpose, it is appropriate to commence with review of the background of these medical conditions. This discussion gives insight into the nature of the conditions, their symptoms, causes and treatment.

Cardiovascular disease (CVD) is a critical public health issue, nationally and internationally. It was responsible for less than 10% of all global deaths at the beginning of the 20th century, but in 2005 that number was 30%. About 80% of these deaths were in low- and middle- income countries. (WHO, 2007) According to American Heart Association (AHA) and WHO, a study by Gakidou et al. (2010), the prevalence of arterial hypertension in people with diabetes was high in all surveys conducted in seven countries. The rate of diagnosis was distinctly higher in England, Scotland, and the United States than in Colombia, Iran, Thailand, or Mexico. There are many risk factors for cardiovascular diseases, the coronary Heart Disease CHD alone has more than 200 risk factors but the most significant risk factors are: diabetes, hypertension, obesity, dyslipidaemia, physical inactivity, smoking, diet, family history and genetic influence. (Hobbs, 2004)

1.1.1 Diabetes

Diabetes mellitus, commonly referred to as diabetes is a group of metabolic disorder that present the phenotype of hyperglycemia. The etiologies of diabetes mellitus are complex interaction of genetics, environmental factors and life-style choices. The

pathogenesis of diabetes include reduce insulin secretion, decrease glucose usage and increase glucose production. (Pickup & Williams, 2003) WHO defines diabetes as having fasting blood glucose level (BGL) of 126 milligrams per deciliters (mg/dl) or random plasma glucose > 200 such that for symptomatic individuals, one abnormal value is diagnostic but asymptomatic individuals two values are required. Glucose is the end product of carbohydrate (starches and sugars) foods. The pancreas, an organ near the stomach, makes a hormone called insulin to help glucose get into the body's cells. The glucose transported into other organs in the body provides the body with energy. When the blood glucose elevates, insulin is released from the pancreas to normalize it. In patients with diabetes, the absence or insufficient production of insulin causes hyperglycemia. The BGL and amount of insulin production classify diabetes as follows:

- **Pre-diabetes**

  Pre-diabetes is a condition in which blood glucose levels are higher than normal (a fasting glucose level between 100 and 125mg/dl) but not yet diabetic. In this condition, the cells in the body are becoming resistant to insulin or the pancreas is not producing sufficient insulin as required. This is also known as impaired fasting glucose (IFG) or impaired glucose tolerance (IGT). These are intermediate state of abnormal glucose regulation between how a body normally uses glucose and diabetes. People with IFG or IGT are at risk of developing type 2 diabetes, although this is not predictable. (Asare, 2008)

- **Type 1 Diabetes**

  The other form of diabetes mellitus is type 1 diabetes, which follows immunologic destruction of pancreatic b-cells. Type 1 diabetes usually begins

early in life and is often called juvenile diabetes. This form of diabetes frequently produces microvascular complications, nephropathy, and retinopathy,but it also predisposes to CHD. (Unger et al. 1998) In type 1 diabetes, the pancreas produces little or no insulin. The origin of type 1 diabetes is not fully understood, and there are several theories. To some it can be caused by genetic disorder. To others, the pathologic process in nearly all patients with type 1 diabetes is autoimmune, destruction of the pancreatic islet beta cells with absolute loss of insulin secretion. (Kumar, 2002) But all these possible causes converge to the same point; the pancreas produces little or no insulin at all and so the glucose stays in the blood. For now there is no cure that exists for diabetes. Without daily injection of insulin, people with type 1 diabetes will not survive. For this reason it is also known as insulin- dependent diabetes. (Pickup and Williams, 2003)

- **Type 2 Diabetes**

  Type 2 diabetes, formerly known as non-insulin-dependent diabetes mellitus (NIDDM) and adult-onset diabetes is most common form of diabetes (90% to 95% of people with diabetes) and appears most often in middle-aged adults. But today, the rate at which adolescents and young adults are developing type 2 diabetes is alarming. Type 2 diabetes occurs from variable combinations of insulin resistance and insulin secretion defects (beta-cell dysfunction), with one or other abnormality predominating in a given patient.(Kumar, 2002) Thus the body of the patient produces insulin, but the system has resistance to it or cannot make efficient use of the insulin it produces.

- **Gestational Diabetes**

Gestational diabetes is hyperglycemia which is recognized firstly during pregnancy. It has the same symptoms as in type 2 diabetes. Usually diagnosed through prenatal screening, rather than reported symptoms. (WHO, 2008)

- **Some Symptoms of Diabetes**

Increased or Extreme Thirst: The extreme thirst or unusual demand of drinking water may suggest a sign of diabetes, especially when it is compounded with frequent urination. The pancreas of the diabetic patient produces little or no insulin which makes the body absorb extra water out of the blood to dilute the body glucose. As a result of this, the body becomes dehydrated and therefore demands more water to replace the one that is losing.

Increased or Frequent Urination: This results from high blood glucose level. When pancreas secretes little or no insulin, the kidney cannot filter glucose back to the blood. For kidney to function effectively, the system absorbs more water to dilute the glucose. This keeps the bladder full and therefore frequent urination.

Unusual Weight Loss: This is mostly identified among type 1 diabetic patients, where the pathologic process in nearly all the patients is autoimmune destruction of pancreatic islet beta cell with absolute loss of insulin secretion. In this situation, glucose (energy source) cannot be transported into the body's cells and so the body demands for energy source breaks down muscle tissues and fat for energy. Therefore, the wear out tissues and fat contribute to the weight loss.

Increased Fatigue: This is as a result of lack of energy in the body. When the body's source of energy, glucose, enters into the bloodstream, it is assisted by insulin to be transported into the cell of the body where energy is produced. The situation where there is little or no secretion of insulin will make glucose remain in the bloodstream. The cells are then unable to produce energy for activities of the body.

Other symptoms include tingling or numbness in limbs, blurred vision, skin itching, frequent infections or cut and bruises that take long time to heal. (ADA, 2008)

- **Causes of Diabetes**

Scientists are unsure of the exact cause, although researchers are investigating a combination of genetic and environmental factors. So far researchers have identified 20 genes involved in Type 1 diabetes, and they are working to determine each gene's role in causing the disease. The inheritance patterns of Type 1 diabetes are complicated, with many different genes influencing a person's risk. For instance, a gene known as DR plays a role in Type 1 diabetes. Two forms of this gene, called DR3 and DR4, are present in 95 percent of people with Type 1 diabetes. People who inherit DR3 alone develop diabetes at an older age and have antibodies that destroy insulin-producing beta cells. Those who inherit DR4 tend to develop diabetes earlier in life and have antibodies that destroy insulin. A person with both DR3 and DR4 typically develops diabetes at a very young age and has the highest level of insulin-destroying antibodies. (Pickup and Williams, 2003)

In the year 2000, researchers were surprised to find that a variation of a gene called Caplain-10, which is not involved in glucose metabolism, is associated with the development of Type 2 diabetes. One form of this gene produces a small amount of protein, and researchers are studying how this decrease in protein increases a person's risk for diabetes. Other genetic studies indicate that certain genes cause a variation of Type 2 diabetes called maturity onset diabetes of the young (MODY), which develops in people under the age of 25. Although scientists do not yet understand how these genes cause MODY, the genes are known to be active in the liver, intestine, kidney, and pancreas. (Microsoft Encarta, 2009)

Other scientists hope to identify the environmental factors that trigger Type 1 diabetes in people with a genetic predisposition for the disease. If they can determine what causes the immune system to attack the cells that produce insulin, they may discover how to prevent the condition from developing. For instance, studies suggest that certain viruses such as coxsackie B, rubella, and mumps, may trigger an immune reaction against beta cells or in some cases directly infect and destroy these cells. (Microsoft Encarta, 2009)

Researchers attribute most cases of Type 2 diabetes to obesity. Studies show that the risk for developing Type 2 diabetes increases by 4 percent for every pound of excess weight a person carries. Researchers are investigating the exact role that extra weight plays in preventing the proper utilization of insulin and why some overweight people develop the disease while others do not. (Microsoft WHO, 2008)

- **Treatment of Diabetes**

At present no cure exists for diabetes, experts can only manage it. Research also focuses on transplanting a healthy pancreas or its insulin-producing beta cells into a person with Type 1 diabetes to provide a natural source of insulin. Some patients who have received pancreas transplants have experienced considerable improvements in their health, but positive, long-term results with beta-cell transplants have not yet occurred. In both types of transplants recipients must take drugs that suppress their immune systems so the body will not reject the new pancreas or cells. These drugs can cause life-threatening side effects because the patient's body can no longer protect itself from other harmful substances. In most people with diabetes, these drugs pose a greater risk to health than living with diabetes. Scientists are also studying the development of an artificial pancreas and ways to genetically manipulate non-insulin-producing cells into making insulin. (Unger et al, 1998)

New methods for accurately measuring blood glucose levels may improve the quality of life for many individuals with diabetes. New techniques include the use of laser beams and infrared technology. For example, a tiny computer using infrared light can be used to measure a person's blood sugar level. The computer automatically delivers the reading to an insulin pump carried on the diabetic's body that injects the appropriate amount of insulin. (Microsoft Encarta, 2009)

Other advances include new drugs that control blood sugar. In April 2000 the United States Food and Drug Administration (FDA) approved glargine, an insulin drug that needs to be injected only once a day. Sold under the brand name

Lantus, this drug can be used by people with Type 1 diabetes, as well as by those with Type 2 diabetes who require insulin injections. And, as mentioned earlier, in 2006 the FDA approved a form of insulin that can be inhaled. Physicians have long known that some insulin-dependent diabetics fail to take the drug as often as needed because of the discomfort of injections. Doctors hoped the inhalant form of insulin would lead to better patient compliance. (Microsoft Encarta, 2009)

A number of drugs have been developed to help people with Type 2 diabetes. Examples include acarbose, (sold under the brand name Precose), which controls blood sugar by slowing the digestion of carbohydrates; and metformin (sold under the brand name Glucophage), which controls liver production of sugar, causes weight loss, and reduces total cholesterol. Pioglitazone (brand name, Actos) and rosiglitazone (brand name, Avandia) are drugs that make the cells more sensitive to insulin. A study published in 2007 found that Avandia increases the risk of heart attacks. Soon afterward the FDA told the manufacturers of Avandia and Actos, which had also been shown to carry a heart risk, adding prominent safety warnings to the drugs' labels. (Gakidou et al, 2011)

1.1.2 Hypertension

Blood pressure is the force of blood against the arteries when the heart beats (systolic pressure) and rests (diastolic pressure). It is measured in millimeters of mercury (mmHg). This is normally represented as two numbers in the form of fraction. The numerator or the upper number is called systolic pressure which indicates the highest pressure in the arteries generated when the heart beats. The denominator or the lower number is the diastolic pressure. That is pressure in the

arteries when the heart is relaxing between the heart beats. It gives a clue as to how much resistance in the small arteries is putting up to the flow of blood. (American Heart Association, 2008)

Blood pressure of a particular person may vary depending on situations like physical activities, emotional conditions and time of day or night. However, there are two factors that determine blood pressure in a body. These are the amount of blood that the blood pumps and the diameter of the arteries receiving blood from the heart. When the diameter of arteries is small, it increases the resistance to blood flow. The more blood the heart pumps and the smaller the arteries, the higher the blood pressure. This high pressure is called Hypertension.

The American Heart Association categorized blood pressure as in Table 1.1

Blood Pressure Category	Systolic(mmHg) (upper #)	Diastolic(mmHg) (lower #)
Normal	< 120	< 80
Pre-hypertension	120 -139	80 - 89
High blood pressure(hypertension) stage 1	140 - 159	90 - 99
High blood pressure(hypertension) stage 2	≥ 160	≥ 100
Hypertensive crisis(Emergency care needed)	≥ 180	≥ 110

(American Heart Association, 2008)

- **Symptoms of Hypertension**

  There are no physiological symptoms of blood pressure. Many people falsely believe that high blood pressure has to do with being tense, nervous or hyperactive. One can be calm or relaxed and still have blood pressure. Therefore,

it is important for everyone to have his/her blood pressure measured or checked.(American Heart Association, 2008)

- **Causes of Hypertension**

The causes of hypertension are not fully explicable by cardiologists. Several factors and conditions may play a role in it development, including: smoking, lack of physical activity, being overweight too much salt in the diet, too much alcohol consumption, stress, old age, genetics, family history of high blood pressure, chronic kidney disease, adrenal thyroid disorders. There is no clear cause that can be identified for nearly 95% of the cases. This type of high blood pressure is known as essential hypertension, and scientists suspect that genetic factors may play a role in its development. In about 5% of cases high blood pressure develops as a result of other medical disorders, such as kidney or liver disease, or as side effect of certain medications. This type of high blood pressure is known as secondary hypertension.

Kidney plays a major role in the regulation of blood pressure. It secretes the hormone rennin, which causes arteries to contract, thereby raising blood pressure. The kidneys also control the fluid volume of blood, either by retaining salt or excreting salt into urine. When kidneys retain salt in the bloodstream, the salt absorbs water, increasing the fluid volume of blood. As higher volume of blood passes through arteries, the blood pressure increases.(Asare, 2008)

- **Treatment of Hypertension**

Diet and change of lifestyle, like losing weight and stopping smoking as recommended by physicians prevent increase in blood pressure in people with prehypertension. Some patients can lower the blood pressure by limiting salt in their diet. Increasing physical activity and reducing alcohol consumption to less than two drinks per day for men and one drink per day for women may also lower blood pressure.

For those with stage 1 and stage 2 hypertension, a physician may prescribe diet and lifestyle changes, as well as one or more drugs known as antihypertensive. Diuretics are antihypertensive that promote excess salt and water excretion, reducing the amount of fluid in the bloodstream and relieving pressure on the walls of blood vessel. Beta blockers reduce heart rate and amount of blood the heart pumps. ACE inhibitors prevent the narrowing the walls of blood vessel to control blood pressure. Calcium channel blockers slow heart rate and relax blood vessels. Studies show that two drugs are more effective than one drug at lowering blood pressure to less than 140/90 mmHg.(Asare, 2008).

1.1.3 Obesity

The prevalence of obesity in the globe is in ascendency. According to the International Obesity Task Force, more than 1.1 billion adults worldwide are overweight (16.6%) and 312 million of them are obese (4.7%). In 2005, the prevalence of obesity in U.S adults (above 18 years old) was 23.9%. This trend may have far-reaching consequence, as plays important etiologic in variety of diseases, including hypertension, type 2 diabetes, CHD and stroke.(Lahham, 2009)

Obesity can be classified as mildly, moderate or severe depending on the persons Body Mass Index (BMI). BMI is one of the most commonly used indicators of obesity, but it is not an ideal one as it does not account for the body fat distribution. It is the weight in kilograms divided by the square of the height in metres. (Oster et al., 1999) According to the World Health Organization (WHO) definition 'overweight' is BMS ≥ 25kg/m^2 and 'obesity' is a BMS ≥ 30kg/m^2 . This study categorize BMI into four group; underweight (BMI <18.5), normal (18.5 ≤BMI≤ 24.9), overweight (25≤ BMI≤29.9) which is consistent with US dietary guidelines for Americans. The fourth obesity is defined as BMI of 30kg/m^2 or greater (consistent with criteria of WHO).

Waist -to-Hip Ratio is an easy and convenient measure of abdominal obesity and an indicator of the hidden fat in the abdomen. The waist and hip circumference (measurement) is obtained and the ratio is calculated by dividing the waist measurement by hip measurement. A higher ratio implies larger waist measure which reflects high amount of abdominal fat. But a lower ratio means larger hip measure which is good as it may reflect the amount of lower body muscle.

This fat in large amount can interfere with normal metabolic functions causing high blood sugar, harmful cholesterol levels and increase the risk of heart diseases and diabetes. According to WHO the cut off for the cardiovascular risk factors is > 0.90 in men and > 0.85 in women. (WHO, 2006)

Obesity is a major risk factor for cardiovascular diseases and has been strongly associated with insulin resistance. Insulin resistance may be a mechanism by which obesity leads to CVD. Weight loss can improve cardiovascular risk, decrease insulin sensitivity. Obesity and insulin resistance also has been associated with other risk factors,

such as high blood pressure. Exercising and weight loss can prevent or delay the onset of type 2 diabetes, reduce blood pressure and help to reduce the risk for heart attach and stroke. (WHO, 2004)

1.1.4 Dyslipidemia

Atherogenic dyslipidemia, often called diabetic dyslipidemia in people with diabetes, is a condition associated with insulin resistance. Diabetes tends to lower 'good' cholesterol and raise triglyceride and 'bad' cholesterol levels, which increases the risk for heart diseases and stroke. This type of dyslipidemia is characterized by high levels of triglycerides (hypertriglyceridemia), high level of small low-density-lipoprotein (LDL) particles and low levels of high-density-lipoprotein (HDL). This lipid triad often occurs in patient with premature CHD. Growing evidence suggest that all of the components of the lipid triad can contribute to the development of atherosclerosis (fatty buildup in artery walls) and can be considered a risk factor. Even though most patients with diabetes do not have marked elevations of LDL cholesterol, their levels are high enough to support the development of atherosclerosis. Eating a diet low in saturated fat, *trans* fat and cholesterol can help avoid this LDL cholesterol.(Nesto, 2008)

1.1.5 Physical Inactivity

Physical inactivity is another modifiable major risk factor for insulin resistance and CVD. It is likely that any type of physical activity-whether sports, household work, gardening or work-related physical activity is beneficial. Despite the debate about the amount, intensity, frequency and duration of the activity for optimal health, researchers concur that physical activity is necessary for metabolic and cardiovascular benefits.

Physical activity can slow the initiation and development of diabetes and the sequence of CVD through its effect on body weight, insulin sensitivity, glycemic control, blood pressure, fibrinolysis, endothelial function and inflammatory defense systems. (Bassuk and Manson, 2005) Regular participation in physical activity has the potential to protect against the most common chronic diseases. (Warburton et al, 2007)

People who are physically inactive have an increased risk of developing breast cancer, colon cancer, hypertension, coronary artery disease, type 2 diabetes, osteoporosis or succumbing to stroke relative to their physical active counterpart. (Katzmarzyk and Janssen, 2004). To prevent or treat type 2 diabetes, aerobic and resistance activities are recommended and weight bearing and impact exercise are suggested to prevent osteoporosis. (Warburton et al, 2007)

1.1.6 Family History

It is a recognized risk factor for CVD which represents genetic, environmental and behavioural elements, in addition to the interaction among them. As DNA analysis for susceptible genes is not yet necessary, family history can be helpful public health tool for the prevention of diseases as it can reflect genes susceptibility and other risk factors. (Harrison, 2003)The American Diabetes Association, the American Heart Association and the international Cholesterol Education Programs including family history in their guidelines has considered family history as a factor to be assessed and made decision about treatment.

1.1.7 Smoking

The recent studies shows cigarette smoking is the leading avoidable cause of death in US, accounting for more than 440000 deaths year. It is also the most important modifiable cause of premature death; 33.5% of these deaths are cardiovascular-related. For instance, about 22% of adults with diabetes smoke. Smoking assumed to increase the risk for developing insulin resistance, a condition in which the body does not efficiently use the insulin it produces. Insulin resistance predisposes a person to both diabetes and CVD. It may also decrease high-density-lipoprotein (HDL), cholesterol in the blood, raising the risk of a heart attack. In addition, smoking causes coronary spasms by increasing catecholamine release. (Nesto, 2009)

1.2 Study Area

The Komfo Anokye Teaching Hospital (KATH) in Kumasi, Ashanti Region being a centre in terms of hospitals and transit point across the country (Ghana), was chosen for the study. The hospital was established in 1955 and became a Teaching Hospital in 1975 for training of medical students from Kwame Nkrumah University of Science and Technology (KNUST), School of Medical Science (SMS), Kumasi. The diabetes centre was set up in the year 2000 to diagnose, treat and manage diabetes patients. Information from the centre shown that there are hundreds of diabetes patients registered. All of them have been going through diabetic care program such as medications, dietary education and screening for diabetic complications.

1.3 Problem Statement

It is evident that the prevalence of diabetes is a life threatening medical condition all over the globe. According to Ghana Diabetes Association (GDA), diabetes epidemic threatens to overwhelm most healthcare systems worldwide and would hit the disadvantaged and vulnerable people. Sub-Saharan Africa (SSA) has been identified as the region with the highest prevalence rate of diabetes claiming about 3.2 million lives every year. (GDA, 2007).

Many scientists all over the world are investigating a combination of genetic and environmental factors that cause the disease. The fluctuations of clinical indicators such as blood hemoglobin (Hb), blood pressure (BP) and blood glucose level (BGL) may lead to corresponding change in the health of a person. The fact that these indicators determine the presence or absence of a particular disease does not mean one should ignore the effect of some factors like age, gender, height, or weight of a person. These factors sometimes expose people to acquisition of some cardiovascular risk factors, including diabetes, hypertension or obesity.

Sometimes, people acquire certain disease as a result of their lifestyle (i.e. the way they live). Some of these practices are excessive drinking of alcohol, smoking, eating habit or lack of regular exercise. Those people may or may not be aware and prepare a fertile ground for some diseases to thrive.

Researchers continue to find risk factor associated with various cardiovascular diseases but key question that remain unanswered is why people are living with only diabetes and some others with both diabetes and other risk factors like hypertension and obesity. This is why the study seeks to come out with a model to assist physicians to make future predictions.

1.4 Objectives

The objectives of the study are to

1. use logit and probit models for assessing the relation between cardiovascular risk factors and hypertensive diabetics
2. use the models for prediction

1.5 Justification

There have been many studies on diabetes prevalence worldwide and Sub-Saharan Africa has been identified as the region with the highest prevalence rate claiming about 3.2 million lives every year. (Ofori, 2007) However, the information (data) on the prevalence of diabetes and hypertension in Ghana is scanty. (Amoah et al., 2007) For this reason:

1. more information about diabetes with hypertension is required to apply better strategies for prevention and control of diabetes and its sequences which this study envisages.
2. this study deems it necessary to assess the available information and present pictorial interpretation to the understanding of patients to avoid further complications
3. the study would model the associated cardiovascular risk factors in diabetes with hypertension to enable physicians to predict its development and complications in individuals in Kumasi and its environs.
4. the study would serve as base for further academic research to improve upon the existing information and the quality of life among Ghanaians.

1.6 Methodology

The target population of the study comprises the diabetic patients who attend clinic at the Komfo Anokye Teaching Hospital for their medication. The problem being research into cardiovascular risk factors associated with diabetes compounded with hypertension. The suitable statistical model that would be applied are logit and probit. The theories of the two regression analysis have been reviewed in Chapter three. In this study, the data to be used would be collected on patients of 40 years old and above. Statistical Package for Social Sciences (SPSS) software would be used for the analysis of the data. The resources available to ensure the successful completion of this study would be KNUST Mathematics Library, main Library and internet facility.

1.7 Ethical consideration and Quality assurance

The study has received an approval from the Committee on Human Research Publication and Ethics of School of Medical Sciences, Kwame Nkrumah University of Science and Technology, Kumasi-Ghana. Confidentiality was seriously taken into consideration in the data collection process especially, information obtained from patients' medical records and interview conducted. No participant in this research was included unless an informed consent for them to take part was received. The objectives and the processes involved in the survey were explained to all the subjects.

The hospital from which the research was conducted is one of the biggest in the country and is a referral centre for hospitals in the northern sector. The KATH diabetic a specialist clinic and supervised by well trained medical consultants. As a teaching a hospital, the quality of professionals cannot be overemphasized. It has well trained doctors and nurses to diagnose,

manage and educate clients on diabetes, hypertension and other risk factors. For this reason the quality of information obtained was assured.

1.8 Organization of the study

The report is organized into five chapters. The first chapter gives insight into the background, mainly, medical background of cardiovascular risk factors (the nature, the symptoms, causes and management), the problem statement, the objectives of the study, justification, methodology, ethical consideration and quality assurance, and organization of the thesis. Chapter 2 is the literature review. Chapter 3 is the methodology which reviewed the relevant mathematical and statistical methods. The data analysis and modeling is Chapter 4 and finally Chapter 5 is the conclusions and recommendations of the study.

CHAPTER 2

LITERATURE REVIEW

This chapter seeks to review briefly some studies that have been carried out on cardiovascular diseases which diabetes and hypertension are included by some researchers.

King et al. (1998) conducted a research on prevalence of diabetes in adults worldwide. According to the report, the number of adults with diabetes in the globe will rise from 135 million in the year 1995 to 300 million in 2025. These figures represent 4.0% of the world population in 1995 which is expected to rise to 5.4% in 2025. The report stated that diabetes is prevalence in the developed than the developing countries. It was revealed that the first three countries with the highest number of people with diabetes are India, China and US.

A similar research was done by Wild et al. (2004). The aim of the study was to estimate the prevalence of diabetes for all age-groups from the year 2000 to 2030 worldwide. The total number of people with diabetes was projected to rise from 171 million in 2000 to 366 million in 2030. This projection in percentage was 2.8% in 2000 and 4.4% in 2030 indicating that the prevalence of diabetes is expected to almost double between 2000 and 2030. The data on diabetes prevalence by gender indicate that diabetic prevalence is higher in men than women, but there are more women with diabetes than men. Urban and rural populations were considered separately for developing countries and the urban population with diabetes is projected to double between 2000 and 2030. Most expected population growth between 2000 and 2030 would be concentrated in the urban areas of the world. According to the report, when the age-specific prevalence remains constant, the number of people with diabetes in the world is expected to approximately double between 2000 and 2030, based solely upon demographic changes. The

most important demographic influence on diabetes prevalence across the globe appears to be the increase in the proportion of people > 65 years of age. The greatest relative increases would occur in the Middle Eastern Crescent, Sub-Saharan Africa, and India. The report revealed that the absolute increase in the number of people with diabetes would be India, follow by China, Indonesia, Japan, Pakistan, Russia, Brazil, Italy, Bangladesh and so on.

Lahham (2009), researched into Cardiovascular Diseases and Risk Factors among Diabetic patients in Nablus District, West Bank, Palestine. It was observed that all cardiovascular risk factors except systolic blood pressure, diastolic blood pressure and smoking, were significantly higher in diabetic patients than nondiabetic controls. The best predictors of diabetes using logistic regression analysis were total cholesterol, family history of diabetes, waist circumference and triglyceride. About 78 % of diabetic patients had at least one diabetes complications; hypertension was in diabetic patients as twice as nondiabetics (55.2% vs 27.0%). Diabetic females were more obese and less physically active. But diabetic males were more ex- and current smokers than diabetic females. Frequencies of diabetes complications were more in diabetic male than diabetic females. Diabetic patients living in the city showed higher prevalence of almost all cardiovascular risk factors. Similarly, the diabetic complications were higher in diabetics living in the city except systolic blood pressure and body mass index which were higher in diabetics living in the refugee camps. Awareness of cardiovascular complications was comparable and low in both diabetic patients 56.9% and nondiabetic controls 57.2%, (p=1.000).The study concluded that urbanization and obesity are high in Palestine and awareness is low, more risk factors and complications of diabetes are expected in the ensuing years. So effective interventions must be developed and implemented in the national level.

In their study, Storving et al. (2003) observed an increase in prevalence of obesity in many countries in the globe and the importance of obesity as a risk factor for diabetes. The report stated that number of cases of diabetes in the year 2030 may be considerably higher than expected. Increasing evidence of effective intervention, including changes in diet and physical activity or pharmacological treatment to reduce prevalence of diabetes, provides an impetus for wider introduction of preventive approaches. However, improved survival may contribute to increasing prevalence of diabetes in future especially in developed countries.

A study was carried out by Barcelo and Rajpathak (2001) to estimate the prevalence rate of diabetes mellitus from the years 2000 to 2025 in Americas. The results were that the number of people with diabetes were 35 million in 2000 and is expected to increase to 64 million by 2025. Whereas currently 52% of these people from Americas live in Latin Americas and the Caribbean; by 2025 the percentage will have reached 62%, representing 40 million persons. In the North America, the recent estimate of the prevalence of diagnosed diabetes among adults in Canada was 3.2%. The report stated that in the United States, the prevalence rate of diabetes increase from 11.4% in 1976-1980 to 14.3% in 1988-1994. Diabetes prevalence rates for Mexican-Americans were twice as high as for non-Hispanic whites. About 20% of non-Hispanic blacks in the United States were affected by diabetes. The prevalence rate of this group was the second highest after that of Mexican-Americans. The report concluded that the Pima Indians from the state of Arizona have shown the highest prevalence of diabetes in the Americas and one of the highest in the world.

Ghaffar et al. (2004) researched into the prevalence of diabetes in South Asia. The report stated that the diabetes prevalence has risen more rapidly in South Asia than any other large region in the globe. India has the highest number of people with diabetes than any other country,

with estimates ranging from 19.4 million in 1995 to 32.7 million in 2000. A projection for 2020 based on a model for estimation by WHO shows a marked escalation related burden in South Asia. The number of people with diabetes is expected to rise by 19.5% in India from 1995 to 2025 representing 57.2 million people. Pakistan is expected to have about 14.5 million people with diabetes by that year.

Mohan et al. (2006) cited the World Health Organization (WHO, 2003) reports which indicate that 32 million people had diabetes in the year 2000. The International Diabetes Federation (IDF) estimates the total number of diabetes to be around 40.9 million in India and this further set to rise to 60.9 million by the year 2025. The National Urban Diabetes Survey (NUDS) reported that the prevalence of type 2 diabetes in people aged 20 years and over was 12.1% and the prevalence of impaired glucose tolerance test (IGT) was 14%.

Melidonis et al. (2006) reported that the prevalence of diabetes in a representative rural area of Greece was 7.8%, with known diabetic being 5.3% of the population and undiagnosed diabetes being 2.5%. The prevalence was not difference between males and females. There was a significant increase in the prevalence of diabetes with increasing age; impaired glucose tolerance was diagnosed in 3.9% while impaired fasting glucose was diagnosed in 1.9%. Age, obesity, family history of diabetes, arterial hypertension and evaluated triglyceride level were significantly associated with the presence of diabetes.

According to Aekplakorn et al. (2003), the estimated national prevalence of diabetes in Thai adults aged 35 years and over was 9.6% which included 4.8% previously diagnosed and 4.8% undetected before the survey. The prevalence of impaired (borderline high) fasting glucose was an additional 5.4%. The duration of disease in those with known diabetes was 6.4 years. The

prevalence of type 1 diabetes was estimated to be 0.2%. The number of diabetic Thai adults aged 35 years and over in 2000 was estimated to be 2.4 million.

A study by King and Henry (1997) was to estimate the global number of excess death due to diabetes in the year 2000. According to the report, global excess mortality attributable to diabetes was estimated at 2.9 million deaths (1.4 million men and 1.5 million women), which was equivalent to 5.2% of world all-cause mortality in the year 2000; 1 million deaths in developed countries and 1.9 million deaths in developing countries. In countries with a high prevalence of diabetes in younger age–groups (Southeast Asia Region, SEAR D; Arabian Peninsula, Eastern Mediterranean Region, EMR B; and Western Pacific Region, WPRB3), the percentage of excess deaths peaked at 50-54 years of age. In the rest of the world, the percentage of excess deaths due to diabetes was highest in people aged 55-59 year. The report stated that 7.5 million people with diabetes are estimated to have died in the year 2000. This includes 4.6 million people with diabetes assumed have died from causes other than diabetes, plus the excess 2.9 million that died because of diabetes. In individuals with diabetes below 35 years, 75% of all deaths were attributable to diabetes; in individuals with diabetes aged 35-64 years, 59% of deaths were attributable to diabetes; while in individuals with diabetes and older than 64 years, 29% of all death were attributable to diabetes. The conclusion was that diabetes is likely to be the fifth leading cause of death globally.

WHO and IDF (2006) conducted a study in Europe and Asia on definition and diagnoses of diabetes. From the report, the definition and classification of diabetes in Europe (DECODE) data shown that of all European people with IFG defined by a fasting plasma glucose of 6.1-6.9mmol/L alone, 64.8% have isolated IFG, 28.6% have IGT and 6.6% have diabetes. In a similar definition and classification of diabetes in Asia (DECODA) the data shown that Asia

people with IFG defined by fasting plasma glucose of 6.1-6.9mmol/L alone, 45.9% have isolated IFG, 35.2% have IGT and 18.9% have diabetes. According to the report, the prevalence of IFG varies between populations and across different age groups within populations. Overall prevalence rates in the order of 5% or more are common. IFG is typically more common in men than in women. The DECODE study showed an increase in prevalence of isolated IFG from 5.2% in 30-39 year old men up to 10.1% in 50-59 year old men and then decrease to 3.2% in 80-89 year old men, whereas in women prevalence increased from 2.6% in 30-39 year old to 5.9% in 70-79 year old. In Asia populations, prevalence of isolated IFG generally increases with age, except in the India population where prevalence does not change much with age. Data from Mauritius indicate that in people with IFG at baseline, 40% reverted to normal, 15% remained as IFG, 20% changed to IGT and 25% developed diabetes.

Amoah et al. (2003) carried out a study to assess risk factors for cardiovascular disease in the Ashanti Region of Ghana. The report stated that 14 out of 931 individuals in the study had diabetes mellitus giving a prevalence of diabetes for the whole population of 1.5%. About 71.4% of the diabetics were from the semi-urban population (prevalence 2.1%) and 28.6% were from the rural population (prevalence 0.9%). 57.1% of the diabetics were female. The majority of diabetic (64.3%) were from the 60 - <70 age group. Another 28.6% came from the < 50 age group. All diabetics were from the semi-urban population. The diabetics were older (mean age 58.4 [SD=9.2]) than the mean age for the whole population. The mean serum glucose for study subjects who were diabetic was 10.9mmol/l (SD=4.8). According to the report the diabetics were heavier than the general population (mean BMI= 23.3 [SD=5.4] vs. 21.1 [SD=4.2]) and had higher mean systolic and diastolic blood pressures compared to the general population

(137.8mmHg [SD=4.1] vs125.6mmHg [SD=26.0] and78.7mmHg [SD=8.7] vs. 74.5mmHg [SD=13.6]).

A study conducted by Al-Moosa et al. (2006) among the people age 20 years and over in Oman, the over all prevalence of diabetes was 11.6%. This varied according to urban or rural residence, age, marital status, educational level, smoking status, measure of obesity, cholesterol and systolic blood pressure. The prevalence of hypertension is 21.1%. The report stated that urban residence was significantly associated with diabetes (adjusted odds ratio (OR)=1.7, 95% confidence interval(CI): 1.4-2.1), as was age (OR=1.2, 95% CI: 1.1-1.2). At 95% confidence interval, 1.5 – 2.1with OR = 1.8 obesity was associated with diabetes. Systolic blood pressure (SBP) 120-139 (OR=1.4, CI: 1.04-1.8), SBP 140-159 (OR=1.9, 95% CI: 1.4-2.6), SBP≥ 160 (OR=1.7, CI: 1.2-2.5) was also significant. Stratified analysis revealed higher education was associated with reduced likelihood of diabetes in rural areas (OR=0.6, 95% CI: 0.4-0.9).

A prospective cohort study by Hwang et al. (2006) in Taiwan assessed the relationship between overweight and obese status and incidence of type 2 diabetes. The report stated that the prevalence of overweight and obesity were 17.6% and 14.5% respectively. Obese subjects with baseline BMI ≥ 25kg/m^2 had a significant multivariate-adjusted relative risk of 14.8 (unadjusted RR = 12.22) for diabetes and the relative risk was higher in women than in men.

A cross-sectional survey by Kim et al.(2004) in South Korea showed that overweight (25.0≤ BMI<30.0) and obesity (BMI ≥30) were positively and statistically significantly associated with diabetes (OR=2.34 and 2.64 in men respectively, and OR=3.20 and 4.03 in women respectively).

Hodge et al. (2004) carried out a research about the association between type 2 diabetes and fiber, glycemic load (GL), dietary glycemic index (GI), and fiber-rich food. The report

indicated that the odds ratio (OR) for the highest quartile of white bread intake compared with the lowest was significant. Intake of carbohydrate (OR per 200g/day=0.58), sugars (OR per 100g/day=0.61), and magnesium (OR per 500g/day=0.62) were inversely associated with incidence of diabetes, whereas intake of starch (OR per 100g/day=1.47, within the confidence interval, 1.06-2.05) and dietary GI (OR per 10 units= 1.32, within the confidence interval, 1.05-1.66) were positively associated with diabetes.

Montonen et al. (2004), conducted cohort study in Finland supported the hypothesis that development of type 2 diabetes may reduced by the intake of antioxidants in the diet. Intake of vitamin E, α-tocopherol, γ-tocopherol, δ-tocopherol, and β-tocotrienol were inversely related to a risk of type 2 diabetes. Among single carotenoids, β-cryptoxanthin intake was significantly associated with a reduced risk of type 2 diabetes. No association was evident between intake of vitamin C and type 2 diabetes.

Ruangwatcharin (2007) cited Oslo (2000) a study that investigate the relationship between leisure time physical activity and smoking measured in middle age, and the occurrence of the metabolic syndrome and diabetes in men. According to the report, leisure time physical activity adjusted for age and educational attendance was significant predictor of both the metabolic syndrome and diabetes in 2000. The odds ratio was 0.65 [95% CI: 0.54-0.80] for the metabolic syndrome and 0.68 [0.52-0.90] for diabetes, $p< 0.05$). Smoking was associated with the metabolic syndrome but not with diabetes in 2000.

Another study conducted by Nakanishi et al. (2003) about alcohol consumption and risk for development of impaired fasting glucose or type 2 diabetes in middle-aged Japanese men. There was U-shape association between alcohol consumption and the incidence of IFG or type 2

diabetes, with the lowest incidence at alcohol intake of 23.0-45.9g ethanol/day. The report indicated that moderate alcohol consumption among healthy Japanese men is associated with reduced risk for development of IFG or type 2 diabetes.

Ruangwatcharin (2007) conducted research into prevalence of diabetes mellitus and related factors among people aged 40 years and above in Kiriratnikom District, Surat Thani province, Thailand. The research was a cross-sectional study to ascertain the prevalence of diabetes and assess associations of diabetes risk with variety of potential risk factors. 483 subjects were selected by stratified random sampling which ensured proportional representation for each of the 8 sub-districts. The study considered fasting blood sugar level ≥ 126 mg/dl and 100-125 mg/dl for subjects with diabetes and pre-diabetes respectively. Prevalence of diabetes and pre-diabetes were, respectively, 5.2% (6.3% in men and 4.2% in women) and 9.5% (9.5% in men and 9.6% in women). More than half of those with diabetes had not be previously diagnosed (prevalence of diagnosed and undiagnosed diabetes were 2.5% and 2.7% respectively). The report stated that hypertension was significantly associated with diabetes (adjusted odd ratio (OR) = 14.49, 95% CI: 1.11- 1.28). Alcohol consumption was significantly associated with diabetes. It was concluded that diabetic screening program should be conducted due to the large number of undetected diabetes in the district and also health promotion for reducing BP and overweight is useful for prevention of diabetes.

According to Foy et al. (2005), the study among US adults, smoking shares a robust association with incidence of diabetes. The study divided the sample into 5 groups by number of cigarettes smoked daily and years of smoking: never, former smokers with < 20 packs-years, former smokers with ≥ 20 pack-years, current smokers with < 20 pack-years and current smokers

with ≥ 20 pack-years. Current smokers exhibited increased incidence of diabetes compared with never smokers (OR=2.66, p=0.001).

Bowden et al. (2006) conducted a research to explore the individual risks for diabetes and for heart disease, and the relationships between them. Cardiovascular disease risks include insulin resistance, obesity, high HDL cholesterol, and hypertension, and these risk factors are thought to be increased in diabetes. It was also stated that hypertension, hyercholesterolaemia, and smoking constitute the major risk factors for coronary heart disease and cardiovascular disease. The alternative view is that the increased risk of coronary heart disease is not caused by diabetes, but that "these two conditions share common antecedents". Metabolic syndrome (characterised by insulin resistance) was suggested to be one of these common antecedents, while another is an inherent genetic susceptibility to both diseases.

A study on the characteristics of hypertensive and diabetic by Mensah et al. (2006) revealed a very interesting statistics. Between 2000 and 2005, about 75% of the hypertensive patients admitted to KATH wards had their body mass index, an indication of obese, above normal weight. About 67% of them were above 60 years. On the other hand, 72% of the diabetic patients were above the age 60 years, while 65% were over weight.

Micah and Agyenim (2003) conducted a study on hypertension and renal failure at KATH, Kumasi. The objectives of the study were to assess: (a) renal disease and blood pressure related admissions and deaths among acute medical admissions to KATH during an 8-month period; (b) the burden of renal disease among out-patient hypertensive at the same hospital. According to the study the Ward admission book for two periods were studied, 1st January to 30th April 1995 and 1st January to 30th April 1996. The death rate among the admissions was 22.9%

(758 among 3319 admissions); 28.5% of these were among those with high BP and its complications, or renal failure. When men and women were examined separately, men had a higher death rate than women. No one died with a diagnosis of 'heart attack'. Out of 496 referrals to the hospital, 48 patients (9.7%) were excluded because they neither had their BP recorded nor seen a doctor. A total of 448 (163 men and 258 women) patients were available for analysis. The report stated that body weight in men and women was similar, as were systolic and diastolic BP. Men had higher plasma creatinine values than women, but the proportions with proteinuria in the two groups were similar. Hundred and ninety-three (17.9%) of 3317 acute medical admissions were ascribable to a cardiovascular cause (hypertension, heart failure, stroke); 171 of these died. Eighty- nine of the 448 had a diastolic blood pressure >115mmHg; in this group 38 (42.7%) had a plasma creatinine of 140mmol/L(and 18 or 20.2%>400mmol/L).

A study conducted by Rosemary et al. (2006), was to determine the burden of illness in a representative sampling of adult urban women. A positive review of systems for hypertension was reported in 309 (23.7%) with only 52.5% using anti-hypertensive medication and only 4.4% had a normal blood pressure. The blood pressure measurement was elevated in 712/1303 (54.6%). Significant risk factors for an elevated blood pressure included age ≥ 50 years, BM I ≥ 30.00, parity of three or more children, menopause before age 50years, elevated fasting blood glucose , elevated fasting cholesterol and no formal education and first degree family history of hypertension. There was no association with income level, diet, or activity. Significant protective factors include young age, null parity, normal BMI, and a lower than normal fasting LDL-cholesterol. The conclusion was that the prevalence of presumptive hypertension is greater than anticipated. Public health initiatives to increase awareness of hypertension and to initiate and

maintain treatment regimens will serve to improve the health of the women and urban community.

Francesco et al. (2006) researched into prevalence, detection, management, and control of hypertension in 12 villages in Ashanti, Ghana. Out of 1013 sample, 385 were men while 628 were women, and 532 lived in semi-urban while 481 in rural villages. According to the report, women were heavier but shorter than men and therefore, had a higher BMI. Participants in semi-urban areas were of comparable height to rural dwellers but were heavier than rural dwellers. They had higher BP than in rural dwellers. The semi-urban dwellers also had a significantly higher BMI compared to the rural dwellers. Systolic BP increased with age, while diastolic BP peaked in the 55 to 60 year group than reached plateau. This pattern was seen in both rural and semi-urban participants. Diastolic, but not systolic BP, was higher in men than in women. Both systolic and diastolic BPs was higher in the semi-urban participants. These differences were confirmed after adjustment for the confounding effects of age, gender, and BMI. The overall prevalence of hypertension was 28.7% in both men and women, but higher in semi-urban and villages and increased with age. Hypertension is common in adults in central Ghana, particularly in urban areas. Detection rates are suboptimal in both men and women, especially in rural areas. Adequate treatment of high BP is at a very low level. The report remarked that there is urgent need for preventive strategies on hypertension control in Ghana.

Asare (2008), conducted a research to model the Hypertensive and non-hypertensive diabetic patients at Komfo Anokye Teaching Hospital, Kumasi. According to the report, out of 260 diabetic patients studied the females were more diabetic (73%) than the males (27%). Among these diabetic patients, 144 (55%) were hypertensive while 116 (45%) were non hypertensive. Out of 144 hypertensive diabetic patients, 111(77%) were females and the rest

were males. It was therefore observed that females were more diabetic and more hypertensive than the male patients. The report stated that the number of patients at 60 years and above diagnosed as hypertensive diabetic outweighed those between 36 and 59 years and this range also outweighed 35 years and below. Moreover, there was positive correlation between age and systolic BP as well as diastolic BP for both hypertensive and non hypertensive diabetic patients. The clients suffering from diabetes with hypertension reported significantly higher levels of BMI compared with that of the clients suffering from only diabetes. However, the blood glucose levels reported by the two groups were not statistically significant.

According to WHO (2010), a study, Management of diabetes and associated Cardiovascular risk factors in seven countries was conducted by Institute for Health Metrics and Evaluation (IHME) researchers. Of the seven countries studied (Colombia, England, Iran, Mexico, Scotland, Thailand, and the United States), Mexico had the highest prevalence of diagnosed diabetes, with 24% of males and 21% of females over the age of 35 with the disease. England and Scotland have the lowest prevalence rates. Prevalence is higher among men in Colombia, England, Mexico, and the US, and among women in Iran and Thailand. Diagnosis rates were higher for women than for men. In all countries, only a small percentage of people with diabetes met the International Diabetes Federation's treatment targets. The US had the highest rates, while England and Scotland had the lowest. The prevalence of arterial hypertension in people with diabetes was high in all surveys, and the rate of diagnosis was distinctly higher in England, Scotland, and the United States than in Colombia, Iran, Thailand, or Mexico. Treatment rates for arterial hypertension followed a similar pattern and were higher in the more developed countries. The United States was the best performer, with 38% of men and 25% of women with diabetes and hypertension meeting their treatment targets for hypertension.

The prevalence of high cholesterol among individuals with diabetes was above 55% in all surveys except in Mexico, where it was only about 35%. Countries had a lower proportion of people with diabetes being treated for high cholesterol than for high blood pressure or blood glucose, but of those receiving treatment for high cholesterol, higher proportions were meeting treatment targets compared to those receiving treatment for high blood pressure or blood glucose. Overall, the proportion of individuals with diabetes reaching treatment targets for blood glucose, arterial blood pressure, and serum cholesterol was very low, ranging from 1% of male patients in Mexico to 13% for men in the US. Income and education were not significantly related to the rates of diagnosis anywhere except Thailand. However, in the three countries with available data (Colombia, Mexico, and the US), insurance status was a strong predictor of diagnosis and effective management, especially in the US. In the US, insured individuals were almost twice as likely to be effectively managed as those without insurance.

Daousi et al. (2006) carried out a study to "determine the prevalence of overweight and obesity among patients with type 1 and type 2 diabetes mellitus…and to assess the impact of overweight and obesity on glycaemic control and cardiovascular risk factors in patients with type 2 diabetes" The study clearly confirm the link between obesity/overweight and type 2 diabetes, unlike type 1 diabetes, but their most significant finding is a correlation between obesity and the increased risk of cardiovascular disease. According to the report, obesity in type 2 diabetes was associated with poorer glycaemic control, blood pressure, and lipid profiles, and increased use of lipid lowering and antihypertensive drugs. It concluded that there is a link between obesity as a risk factor for both diabetes and cardiovascular disease.

CHAPTER 3

METHODOLOGY

3.0 Introduction

There are more established as well as emerging techniques of multivariate analysis. The selection of any of these techniques is based on the research objectives and the nature of data available. This chapter contains the review of the theories associated with the statistical techniques that will be employed in the next chapter.

3.1 General Linear Model

Suppose that we have k predictor or independent variables, $X_1, X_2, \ldots, X_k$ with unknown parameters $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$. It is assumed that the response variable is a linear combination (linear function) of the set the predictors. For response variable y_i and the set of k predictor variables, we have the following:

$$y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon_i \quad \text{or}$$

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ji} + \varepsilon_i \tag{3.1}$$

for all $j = 1, 2, \ldots, k$

If $X_{0i} = 1$, *for all* $i = 1, 2, \ldots, n$. The equation (3.1) can be written as:

$$y_i = \sum_{j=0}^{k} \beta_j X_{ji} \varepsilon_i \tag{3.2}$$

where β_j is the regression coefficient for each corresponding predictor variable, X_j, and ε_i is the error of the prediction. Note that the expectation $E[\varepsilon_i] = 0$, variance $(\varepsilon_i) = \sigma^2$ and therefore,

$$\hat{y} = E[y_i] = \sum_{j=0}^{k} \hat{\beta}_j x_{ji} \tag{3.3}$$

To fit the above model (3.3) to a dataset (y_i , x_i), one has to estimate the parameters β_i. The most commonly used method of estimation is least square. Its application to linear regression model is extensively used in Milton and Arnold (1995), Walpole et al. (1987) and many other sources.

However, in the situation whereby the response variable y takes one of only two possible values representing success and failure, or more generally the presence or absence of an attribute of interest, the model (3.3) would not be appropriate.

Example 3.1: Suppose we want to study the labour-force participation of adult males as a function of unemployment rate, average wage rate, family income education etc. A person is in labour-force or not. Hence the response variable, labour-force participation, can take only two values: 1 person is in labour-force and 0 if he /she is not.

There are several examples of dataset where the dependent variable is dichotomous. We can consider the following examples; a study of a union membership of a college professor as a function of several quantitative or qualitative variables, a certain drug is effective in curing an illness or not, decision of a firm to declare dividend or not, a president to veto a bill or not.

A unique feature of all these examples is that the dependent variable is of type which elicit as yes or no response (binary data). There is special estimation or inference problem

associated with such models. The most commonly used approaches to estimating such models are the Linear Probability model, the Logit and Probit model. (Vasisht, 2001)

3.2 Linear Probability Model

Suppose we want to model the labour-force participation of adult males cited above, the response y_i is binary, assuming only two values that for convenience we code as 1 and 0. For example we could define

$$y_i = \begin{cases} 1, & \text{if the i-th male adult is in labour-force} \\ 0, & \text{otherwise} \end{cases}$$

We view y_i as a realization of a random variable Y_i that can take the values 1 and 0 with respective probabilities:

$$\Pr(Y_i) = \Pr(Y = 1/X_i = x_i) = \pi_i \quad \text{and } \Pr(Y_i) = \Pr(Y = 0/X_i = x_i) = 1 - \pi_i \tag{3.4}$$

The distribution Y_i is a Bernoulli distribution with parameter π_i, and can be written as:

$$\Pr(Y_i = y_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i} \tag{3.5}$$

for $y_i = 0, 1$ and if $y_i = 1$ we obtain π_i and if $y_{i,} = 0$ we obtain $1 - \pi_{i.}$ As a Bernoulli distribution, the variance and the expected value of Y_i are;

$$\text{Var}(Y_i) = \sigma_i^2 = \pi_i(1 - \pi_i), \quad \text{and} \quad E[Y_i] = \mu_i = \pi_i \tag{3.6}$$

Let us consider linear regression model (3.1) with the usual assumptions:

$$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ji} + \varepsilon_i$$

where $\varepsilon_i \sim N(0, \sigma_\epsilon^2)$, and ε_i and ε_j are independent for $i \neq j$. If X is random, then we assume that it is independent of ε. Therefore,

$$E[Y_i] = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ji} \qquad (3.7)$$

Comparing equations (3.6) and (3.7), we have;

$$\pi_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ji} \qquad (3.8)$$

The model (3.8) above is known as Linear Probability Model, it has dummy response variable which is a linear function of the predictors. (Rush, 2001)

This model is untenable, because of the following reasons associated with it:

a) Non-normality: Because Y_i can take on only the values 0 and 1, the error ε_i is dichotomous as well as – not normally distributed; if Y_i =1 which occurs with probability π_i, then $\varepsilon_i = 1 - E[Y_i]$

$$= 1 - (\beta_0 + \sum_{j=1}^{k} \beta_j X_{ji})$$

$$= 1 - \pi_i$$

Alternatively, if $Y_i = 0$, which occurs with probability $1 - \pi_i$, then

$$\varepsilon_i = 0 - E[Y_i]$$

$$= 0 - (\beta_0 + \sum_{j=1}^{k} \beta_j X_{ji})$$

$$= 0 - \pi_i$$

Because of the central-limit theorem, however, the assumption of normality is not critical to least-square estimation of the normal- probability model.

b) Non-constant error variance: if the assumption of linearity holds over the range of the data, then $E(\varepsilon_i) = 0$. Using the relations just noted,

$$Var(\varepsilon_i) = \pi_i (1- \pi_i)^2 + (1- \pi_i) (-\pi_i)^2$$

$$= \pi_i (1- \pi_i)$$

The heteroscedasticity of the errors bodes ill for ordinary- least-squares estimation of the linear probability model, but only if the probabilities π_i get close to 0 or 1. This suggests that a linear model that allows the predictors to affect the mean but assume that the variance is constant will not be adequate for the analysis of binary data.

c) Nonlinearity: the assumption that $E[\varepsilon_i] = 0$ (i.e. the assumption of linearity) is only tenable over a limited range of X-values. If the range of X_i is sufficiently broad, then the linear specification cannot confine π_i to the unit interval of [0,1] (Fox, 2010)

Therefore, there is a need of probability model that has the following two features:

i) As X increases, π_i increases but never step outside the interval [0,1] or map the probabilities from [0,1] to entire real line, and

ii) The relationship between π_i and X_i is non-linear, that is, the function approaches zero at slower and slower rates as X_i gets small and approaches one at slower and slower rates as X_i gets very large.(Fox, 2010)

3.3 Logit Model

Logit analysis is univariate or multivariate technique which allows for estimating the probability that an event occurs or not, by predicting binary dependent outcome from a set of independent variables. Let us consider the linear probability model in (3.8)

$$\pi_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ji} \qquad \text{and let} \qquad Z = \pi_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ji}$$

$$\text{implies} \qquad Z = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ji} \tag{3.9}$$

substituting the above equation in a logistic function, $f(z) = \frac{1}{1+e^{-z}}$ which shows the mathematical form on which logistic function is based. This function provides an appealing elongated S-shaped and approaches 0 as $Z \rightarrow -\infty$ or 1 as $Z \rightarrow \infty$.

.

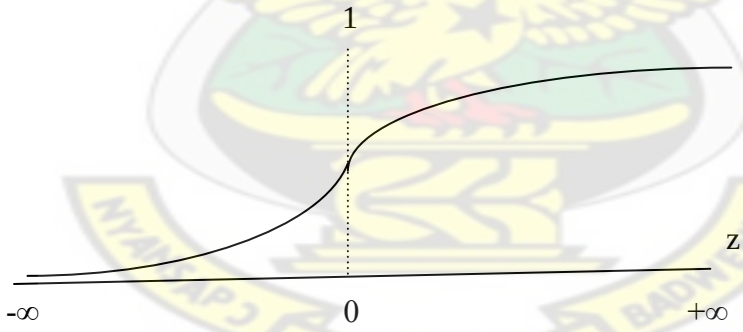

Figure 3.1: The graph of logistic function

It also provides estimates that must lie in the range of zero and one. That is when Z approaches -∞ $f(-\infty) = \frac{1}{1+e^{-(-\infty)}} \rightarrow 0$ and Z approaches +∞, $f(\infty) = \frac{1}{1+e^{-(\infty)}} \rightarrow 1$. Therefore, the range $0 \leq f(z) \leq 1$ regardless of the value of Z. These two features satisfied the conditions the linear probability model could not. (Kleinbaum and Klein, 1998)

However, an estimation problem has been created because π_i is nonlinear not only in X_i but also β's. This means that one cannot use ordinary least square procedure to estimate the parameters.

Let us consider epidemiologic study framework: if the probability of an individual with a disease (D) is 1 and without disease is 0, then the probability being modeled can be denoted by the conditional probability statement;

$$P(D = 1/ X_1, X_2, \ldots, X_k) = \frac{1}{1+e^{-z}}$$

For notational convenience, we will denote the probability statement $P(D = 1/ X_1, X_2, \ldots, X_k)$ as P(X), where X is all Xs. Therefore,

$$P(X) = \frac{1}{1+e^{-(\beta 0+\sum \beta_j X_{ji})}} \quad (3.10)$$

Example 3.2: A study carried out to find out how well people with traumatic brain injury, TBI can be classified shortly after the injury is sustained either having made sufficient good recovery to be back at work 6 months (work = 1) or not having recovered sufficiently to be back at work at 6 months (work = 0). The study was carried out on 54 people with TBI who have Glasgow coma score below 12. Data are obtained on the following covariates: (1) an EEG-derived score (EEG), (2) the coma score (COM) and bivariate (yes/no) people reactivity (RET) as in the Table 3.1 below:

WORK	EEG	COM	RET
1	8	11	Yes
1	4	8	Yes
1	4	9	No
1	7	10	Yes

.	.	.	.
.	.	.	.
.	.	.	.
0	4	5	No
0	5	4	No
0	5	6	Yes
0	7	6	No

To illustrate the use of the above model (3.10), suppose we denote the variable of interest, a person able to recover back to work as D, as coded 1 if able to come to work and 0 if not, and independent variables: X_1 = EEG, X_2 = COM and X_3 = RET (0,1). We would fit the model with unknown parameters β_0, β_1, β_2 and β_3.

$$P(X) = P(D = 1/X_i) = \frac{1}{1+e^{-[\beta_0+\beta_1(EEG)+\beta_2(COM)+\beta_3(RET)]}}$$

Suppose the results of our model fitting yield the estimated parameters; $\beta_0 = 11.471$, $\beta_1 = 1.336$, $\beta_2 = -2.691$ and $\beta_3 = 3.109$. Our fitted model becomes

$$\hat{P}(X) = P(D = 1/X_i) = \frac{1}{1+e^{-[11.471+1.336(EEG)-2.691(COM)+3.109(RET)]}}$$

3.3.1 Logit Transformation

We could recall that the linear probability model in (3.8) the probability π_i on the left-hand-side has to be the range of [0,1], but the linear predictor on the right- hand –side can take any real value. To avert this problem logistic function was used, but the function became elongated S-shaped. A simple solution to this problem is to transform the probability to remove the range restrictions, and model the transformation as a linear function of the covariates.

Suppose that we have a binary response variable y that takes the value 1 if a sampled individual is diseased and takes the value 0 if not diseased. Let the probability that the individual is diseased be $\Pr(D) = \Pr(y = 1) = \pi$ and the probability that the individual is not diseased be $\Pr(\bar{D}) = \Pr(y=0) = 1 - \pi$. Then the odds($\tau$) defined as the ratio of the probability to its complement, or the ratio of favourable (diseased) to unfavourable (not diseased) case is given by:

$$\text{Odds}(\tau) = \frac{\pi}{1-\pi} \tag{3.11}$$

Also if X_i is a risk factor (predictor variable) for $Y_i = 1$ if an individual with disease and $Y_i = 0$ if an individual without disease, then the respective conditional probabilities can be expressed using Equation (3.10) as follows:

$$\pi = \Pr(Y_i = 1/X_i = x_i) = \frac{1}{1+e^{-\left(\beta_0+\sum\beta_j X_{ji}\right)}} \tag{3.12}$$

for probability of an individual with disease and

$$1-\pi = \Pr(Y_i = 0/X_i = x_i) = 1- \frac{1}{1+e^{-\left(\beta_0+\sum\beta_j X_{ji}\right)}} \tag{3.13}$$

for probability of an individual without disease. Simplifying Equation (3.13) we obtained

$$1-\pi = \Pr(Y_i = 1/X_i = x_i) = \frac{e^{-\left(\beta_0 + \sum \beta_0 X_{ji}\right)}}{1 + e^{-\left(\beta_0 + \sum \beta_j X_{ji}\right)}} \tag{3.14}$$

This implies that the Odds ratio given by $\frac{\pi}{1-\pi}$ is obtained from Equation (3.13) and Equation (3.14) as

$$\text{Odds ratio } (\tau) = \frac{\frac{1}{1 + e^{-\left(\beta_0 + \sum \beta_j X_{ji}\right)}}}{\frac{e^{-\left(\beta_0 + \sum \beta_j X_{ji}\right)}}{1 + e^{-\left(\beta_0 + \sum \beta_j X_{ji}\right)}}} = e^{\left(\beta_0 + \sum \beta_j X_{ji}\right)} \tag{3.15}$$

$$\text{Therefore, the Odds ratio } (\tau) = \frac{\pi}{1-\pi} = e^{\left(\beta_0 + \sum \beta_j X_{ji}\right)} \tag{3.16}$$

In general, we consider Y_i as a random sample of n successes out of n_i trials, and let the corresponding probabilities of success be π_i. If we wish to express the probability π_i as a function of the explanatory variable $x_{1j}, x_{2j}, \ldots, x_{kj}$, then Equation (3.16) can be generalized as;

$$\text{Odds } (\tau) = \frac{\pi}{1-\pi} = e^{\sum_{i=0}^{k} \beta i \, Xi} \tag{3.17}$$

where $x_{01} = 1$ for all j = 1, 2, ……n. Taking the natural logarithm through Equation (3.17), we obtained

$$1n(\tau) = 1n\left(\frac{\pi}{1-\pi}\right) = \sum_{i=0}^{k} \beta_i X_{ij} \tag{3.18}$$

The logarithm of the odds ratio is not only linear in X_i, but also the parameters. The above model is called Logit of π, is therefore the log of the odds that Y is 1 rather than 0. This model is a positive monotone (i.e. non decreasing) function that maps the linear predictors $\sum_{i=0}^{k} \beta_i X_{ij}$ into the unit interval. That is, it retains the fundamental linear structure of the model while avoiding probabilities below 0 and above 1.

In Example 3.2 suppose we want to use our fitted model to obtain the prediction that a certain individual would recover back to work, we would need to specify the values of the independent variables (EEG, COM and RET) for this individual. Let us consider an individual with EEG=7, COM = 10 and RET = 1 and the predicted model;

$$\pi_i = \hat{P}(X) = P(D = 1/X_i) = \frac{1}{1+e^{-[11.471 + 1.336(EEG) - 2.691(COM) + 3.109(RET)]}}$$

Substituting the values into the model we heve;

$$\pi_i = \hat{P}(X) = \frac{1}{1+e^{-[11.471 + 1.336(7) - 2.691(10) + 3.109(1)]}}$$

$$\pi_i = \hat{P}(X) = \frac{1}{1+e^{-[-2.978]}}$$

$$\pi_i = \hat{P}(X) = \frac{1}{1+19.648}$$

$$= 0.0484$$

Therefore, the odds of this individual $(\tau) = \frac{\pi}{1-\pi}$

$$(\tau) = \frac{0.0484}{1-0.0484}$$

$$(\tau) = 0.0586$$

The logarithm of odds, logit = $1n\ (\tau) = 1n\ (\frac{\pi}{1-\pi}) = \sum_{i=0}^{k} \beta_i X_{ij}$

$$1n\ (\tau) = 1n\ (0.0586) = \sum_{i=0}^{k} \beta_i X_{ij}$$

$$1n\ (\tau) = -\ 2.837 = \sum_{i=0}^{k} \beta_i X_{ij}$$

This value describes the odds for coming back to work for an individual with specific set of independent valuable X's (EEG, COM and RET). The logit (value) is symmetric around 0, and unbounded both above and below (i.e. it goes from $-\infty$ to ∞), though the probabilities lies [0,1], making the logit a preferred for the response-variable side of a linear model.

The logit model is also a multiplicative for the odds. From Equation (3.17), if $\sum_{i=0}^{k} \beta_i X_{ij}$ is expressed in the form $\beta_0 + \beta\ x_i$, then the Odds $(\tau) = \frac{\pi}{1-\pi} = e^{\beta_0 + \beta\ x_i}$.

$$= e^{\beta_0} . e^{\beta x_i}$$

$$= e^{\beta_0} (e^{\beta})^{x_i}$$

So, increasing X_i by 1changes the logit by β and multiplies the odds by e^{β}. For instance, if β =2, then increasing X by 1 increases the odds by a factor $e^2 \approx 2.718^2 = 7.389$. It is therefore important to understand the parameters, β in the logit model. One way of doing this is to consider the slope of the relationship between the probability π and X. Since this relationship is nonlinear, the slope is not constant; the slope is βπ(1- π), and hence is at a maximum when π = 0.5, where the slope is β/4. (Fox, 2001)

Recall from Example 3.2 the β for people reactivity is 3.109 therefore, increasing people reactivity by one unit multiplies the odds of coming back to work by $e^{3.109} = 22.399$ and the slope of the relationship between the fitted probability of coming back to work and the people reactivity at π =0.5 is 3.109/4 = 0.78.

3.3.2 Estimating the Model Parameter

A model to a data set is obtained when the model parameters $\beta_0, \beta_1, \ldots, \beta_j$ are estimated. The maximum likelihood method is used to estimate these parameters. For a set of observations (x_i, y_i), let us consider y_i as a realization of a random variable Y_i that can take the values one and zero, and x_i as vector of predictor variables, $X_1, X_2, \ldots, X_n$ for i^{th} subject. The distribution Y_i is a binomial distribution with probabilities (likelihood contributions) $\pi(x_i)$, where y_i =1 and 1- $\pi(x_i)$, where y_i = 0 and can be written in compact form as:

$$\Psi(x_i) = \binom{n_i}{y_i} \pi(x_i)^{y_i} [1 - \pi(x_i)]^{n_i - y_i} \qquad (3.19)$$

We want to construct a likelihood function, which expresses the probability of the observed data as a function of the unknown parameters. Since Equation (3.19) account for only one set of observations, the observations are assumed to be independent of each other so we multiply their likelihood contribution to obtain the complete likelihood function:

$$L(B) = \prod_{i=1}^{n} \psi(x_i) \tag{3.20}$$

Where B is the collection of parameters, $\beta_0, \beta_1, \ldots, \beta_j$ and L(B) is the likelihood function of B, which can be written in the form; $L(B) = \prod_{i=1}^{n} \binom{n_i}{y_i} p_i^{y_i} q_i^{n_i - y_i}$

$$= \prod_{i=1}^{n} \binom{n_i}{y_i} (e^{z_i})^{y_i} \left(\frac{1}{1+e^{z_i}}\right)^{n_i} \tag{3.21}$$

Now we obtained the log likelihood function $\ell(B)$ by taking the natural log of equation (3.21), we have;

$$\ell(B) = \sum_{i=1}^{n} \langle y_i z_i - n_i \, 1n(1 + e^{z_i}) \rangle \tag{3.22}$$

But maximum likelihood estimates can be obtained by calculating B such that $\ell(B)$ is maximized. So we employ the technique of calculus to determine the value of B that maximizes $\ell(B)$. Differentiating Equation (3.22) with respect to β_t, we obtain a score function;

$$\ell_t = \frac{\partial \ell(\beta)}{\partial (\beta_t)} = \sum_{i=1}^{n} \langle y_i x_{ti} - n_i x_{ti} e^{z_i} (1 - e^{z_i})^{-1} \rangle \tag{3.23}$$

where t = 0, 1, 2, . . .,j. Setting the resulting derivative (score function) equal to zero we have the following likelihood equations, and there are j + 1 such equations;

$$\sum_{i=1}^{n} (y_i - n_i p_i) = 0$$

for the intercept, β_0, and

$$\sum_{i=1}^{n} x_{ti} (y_i - n_i p_i) = 0$$

for the predictors, $\beta_1, \beta_2 \ldots, \beta_j$. The solution to the likelihood equations is the maximum likelihood estimate, $\hat{B}$. (Rush, 2001)

We can also find the second derivative of the likelihood function (3.22), and by the theorem (under regularity condition), the information function, I, equal to the negative expectation of the second derivative. That is,

$$I = - E\left[\frac{\partial^2 \ell(\beta)}{\partial \beta_t \partial \beta_s}\right] = \sum_{i=1}^{n} n_i x_{ti} x_{si} p_i (1 - p_i), \quad (3.24)$$

and the large sample variance-covariance matrix is I^{-1}. The solution to the above equations can be obtained by using computer programs. (Asare, 2008)

3.3.3 Interpretation of Logit Model

When it comes to interpretation or making inference of a model, so many questions are posed. Two prominent Statisticians, Hosmer and Lemeshow (1989) in the process of testing the significance of predictor variables included in a model posed this question. 'Does the model that includes variables in question tell us more about the outcome (or response) variable than does a model that does not include that variables?' This means that in modeling one must assess the contribution of the predictor variables before they are included in the model.

One of the ways to achieve this is testing of significance of predictor variables. With the help of statistical packages like SPSS, SAS or Stata, we obtain the quantities; maximum likelihood estimates, $\hat{L}$, and their standard error to enable us to do this. The following test statistics would be employed:

The Likelihood Ratio Test

This is the differences between log likelihood statistics for two models, one of which is special case of the other, and has an approximate chi-square distribution. Let us consider the following logit models; model 1: logit $\pi_1(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ and

$$\text{model 2:} \quad \text{logit } \pi_2(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

with their respective maximum likelihood $\hat{L}_1$ and $\hat{L}_2$, where model 1 (reduced model) is a special case (subset) of model 2 (full model). This implies $\hat{L}_2 > \hat{L}_1$.

The reduced model is obtained by setting certain parameters in the full model equal to zero. The set of parameters (β_3) in the full model that is set equal to zero specify the null hypotheses being tested. Therefore, the model 1 against model 2, the null and the alternative hypotheses are ;

$H_0 : \beta_3 = 0$

$H_1: \beta_3 \neq 0$

The above hypotheses are tested with generalized likelihood ratio test statistic;

$$G^2 = -2 \ln \hat{L}_1 - (-2 \ln \hat{L}_2) \quad \text{or}$$

$$G^2 = -2 \ln\left(\frac{\hat{L}1}{\hat{L}2}\right)$$

which is approximately chi-square distribution for large sample size with degree of freedom, df = number of parameters set equal to zero. For this case the degree of freedom is 1. The large value of G^2 the high probability that the null hypotheses will be rejected. (Kleinbaum and Klein, 1998)

The Wald Test

This test is usually done when there is a single parameter being tested. From the above example the hypotheses will be:

$H_0 : \beta_3 = 0$

$H_1: \beta_3 \neq 0$

The Ward test statistic is computed by dividing the estimated coefficient of interest by the standard error;

$$Z = \frac{\hat{\beta}_3}{S_\beta}$$

which is approximately normal distribution N(0, σ^2). The square of Z statistic is approximately a chi-square statistic with one degree of freedom (i.e. $Z^2 \approx X^2$). The computed Z can be compared to percentage points from standard normal table or the square of Z can be compared to percentage points from chi-square distribution with one degree of freedom.

Deviance

This is a measure of the extent to which a particular model differs from the saturated model for a data set. It is defined explicitly in terms of likelihoods of the two models; L_0, the model containing only the constant and L_1, the full model. (Everitt, 2006)

The likelihood ratio is given by:

$$D = -2\ 1n\ L$$

The statistic, D, is known as the deviance can measure the degree to which using the predictor variables improves predictability of the response. With this the deviance of the model without any of the predictor variables is determined and then compare this value with that of the model consisting of different combinations of variables. The deviance always decreases with the addition of more variables, but the more it decreases, the more that particular predictor variable is related to the response variable. As we add variables, we can evaluate the p-value of the deviance, which test for significance of that particular combination of predictor variables. A low p-value ($< \alpha = 0.05$) justifies the rejection of null hypothesis which state that all the β's are equal to zero. The rejection of the null hypotheses means that variables included in the model are significant. (Rush, 2001)

3.4.0 Probit Model

In order to explain the behaviour of a dichotomous response variable we have to use a suitably chosen Cumulative Distributive Function (CDF). The logit model uses the cumulative logistic function, but this is not the only CDF. The estimating model that emerges from the normal CDF is known as Probit Model. (Fox, 2010)

Let us assume that in labour-force participating example, decision of i-th male adult to be in labour-force or not depends on unobserved on latent response variable y_i^* such that the larger the value of y_i^*, the greater the probability of a male adult in the labour-force. Consider the latent variable, y_i^* which is assumed to a linear function of a vector of predictor variable, X and the unobservable error variable, ε_i:

$$y_i^* = \beta_0 + \beta X_i + \varepsilon_i \tag{3.25}$$

where β is a k-vector parameters. That linearly depends on X_i and the error term, $\varepsilon_i \sim N(0, \sigma^2)$. Choosing the case $y_i = 1$ if the latent variable is positive and or otherwise, we have expression of the form;

$$y_i = \begin{cases} 1 & y_i^* > 0 \\ 0 & y_i^* < 0 \end{cases}$$

The latent variable interpreted as the utility differences between choosing $y_i = 1$ and 0. The probability, π_i that $y_i = 1$ can be derived from the latent variable;

$$\pi_i = \Pr(Y_i = 1/ X_i) = \Pr(y_i^* > 0)$$

$$= \Pr(\beta_0 + \beta X_i + \varepsilon_i > 0$$

$$= \Pr[\varepsilon_i > -(\beta_0 + \beta X_i)]$$

$$= 1 - \Phi\left(-\frac{\beta_0 + \beta X_i}{\sigma}\right)$$

implies

$$\pi_i = \Phi\left(\frac{\beta_0 + \beta X_i}{\sigma}\right) \tag{3.26}$$

Assuming that the error term has a standard normal distribution $\varepsilon_i \sim N(0,1)$. The Equation (3.26) becomes

$$\pi_i = \Phi(\beta_0 + \beta X_i) = \Phi(\eta_i) \quad (3.27)$$

where Φ is the standard normal CDF. The Equation (3.27) above is the Probit Model. (Ahmed, 2005)

3.4.1 The Probit Transformation

In insuring that π_i stays within an interval [0,1], we require a positive monotone (non-decreasing) function that maps the 'linear predictor', $\eta_i = \beta_0 + \beta X_i$ into the unit interval. The cumulative distribution function in Equation (3.27) meets this requirement.

However, we obtained the constrained linear-probability model; which is not smooth and symmetric and not approach $\pi = 0$ and $\pi = 1$ as asymptotes. So it is advantageous to write the inverse transformation which gives the linear predictor as a function of the probability:

$$\eta_i = \Phi^{-1}(\pi_i) \quad (3.28)$$

The transformation function in probit model is the CDF of the standard normal distribution;

$$\pi_i = \Pr(Y_i = 1/ X_i) = \Phi(\eta_i)$$

$$= \int_{-\infty}^{\eta_i} \Phi(z)dz$$

3.4.2 Estimating the Model Parameters

Suppose we have probit model for the probability of successes;

$$\pi_i = \Phi(\beta_0 + \beta_i X_i)$$

for observation i and covariates X_i, where β_0 and β_i are unknown parameters to be estimated and Φ is the standard normal cumulative distribution function. Consider the likelihood function for binomial distribution;

$$L = \prod_{i=1}^{n} \binom{n_i}{r_i} \pi_i^{r_i} (1 - \pi)^{n_i - r_i} \quad (3.29)$$

where n_i is the number of trials and r_i is the number of successes on the observation i. Since the binomial coefficient has no effect on the parameter estimates or covariates, we obtain the log likelihood function

$$\ell = \sum_i [r_i \, 1n(\pi_i) + (n_i - r_i) 1n(1 - \pi_i)] \quad (3.30)$$

We compute the score function, S, and the information function, I, which are the first and second derivatives of the log likelihood function respectively. For any parameter β, the derivatives are:

$$S = \frac{\partial \ell}{\partial \beta_k} = \sum_i \left[\frac{r_i}{\pi_i} - \frac{n_i - r_i}{1 - \pi_i}\right] \frac{\partial \pi_i}{\partial \beta_k} \quad (3.31)$$

and

$$I = \frac{\partial^2 \ell}{\partial \beta_k \partial \beta_{k'}} = \sum_i \left[-\left(\frac{r_i}{\pi_i^2} + \frac{n_i - r_i}{(1 - \pi_i)^2}\right) \frac{\partial \pi_i \partial \pi_i}{\partial \beta_k \partial \beta_{k'}} + \left(\frac{r_i}{\pi_i} - \frac{n_i - r_i}{1 - \pi_i}\right) \frac{\partial^2 \ell}{\partial \beta_k \partial \beta_{k'}}\right] \quad (3.32)$$

Note that, from the definition of the probit model;

$$\frac{\partial \pi_i}{\partial \beta_0} = \Phi\,(\beta_0 + \beta X_i)$$

$$\frac{\partial \pi_i}{\partial \beta_1} = X_i \Phi\,(\beta_0 + \beta X_i)$$

$$\frac{\partial^2 \pi_i}{\partial \beta_0} = -\,(\beta_0 + \beta X_i)\,\Phi\,(\beta_0 + \beta X_i) \qquad \text{and so on,}$$

where Φ is the standard normal probability density function. Putting these into the expressions for the derivatives of the log likelihood function and replacing β_k with their current estimates gives rise to systems of (k + 1) Equations in equation (3.31). The maximum likelihood estimates

for β can be found by setting each of the (k +1) equations equals to zero and solve for each β_k, which is a vector. After that the matrix of second partial derivative is negative definite, and that solution is the global maximum rather than local maximum, then we can conclude that this vector contains the parameter estimates for which the observed data would have the highest probability of occurrence. (Czepeiel, 2005)

However, the solution cannot be derived algebraically as in the case of linear equations; it is numerically estimated using an iterative process like Newton-Raphson method. This is the basic iteration step of algorithm that computer package like SAS uses to fit a probit model. (Ahmed, 2005). Newton's method is general procedure for finding the roots of an equation $f(\theta) = 0$ and the algorithm is based on the recursion;

$$\theta_{t+1} = \theta_t - \frac{f'(\theta_t)}{f''(\theta_t)} \tag{3.33}$$

In this case, the equations whose roots we want to solve are those in Equation (3.31), the first derivative of the log- likelihood function. (Mahadeven, 2006) We can write Equation (3.31) as $\ell'(\beta)$ and let $\beta^{(0)}$ represent the vector of initial approximation for each β_k, then the first step of Newton-Raphson algorithm can be expressed as;

$$\beta^{(1)} = \beta^{(0)} + \left[-\frac{l'\beta^{(0)}}{l''\beta^{(0)}}\right] \tag{3.34}$$

The process will continue until there is essentially no change between the elements of β, from one iteration to the next. At that point, the maximum likelihood estimates are said to have converged, and the covariance matrix of the parameter estimates is obtained with the inverse of the observed information matrix, evaluated at the least iteration.

3.4.3 Interpretation of the Model

As already discussed, the logit model denotes the cumulative standard logistic probability distribution function while the probit model uses a cumulative normal probability distribution function. Despite different probability distribution functions used, the logit and probit models produce almost the same results in reality. (Long, 1997) The logit model produces roughly 1.8 times (standard deviation of the standard logistic function) larger coefficients than the probit model. But Z and Wald (chi-squared) statistic are equivalent (almost the same p-values) indicating that the binary logit and probit models present similar results in different ways. Therefore, the inference of the probit model is not different from that of the logit.

CHAPTER 4

DATA ANALYSIS AND RESULTS

4.0 Data Collection

This chapter deals with descriptive analysis of the data, inferential analysis and modeling based on the study objectives and the information gathered. The data was collected from Komfo Anokye Teaching Hospital diabetes centre in Kumasi. In every month, K.A.T.H diabetes centre is scheduled to attend to hundreds of patients in appointment basis. The centre holds three clinics, Tuesday, Wednesday and Friday in a week where clients are attended to by physicians. No client except unusual cases, visit the centre more than one in the month. As a result of that, the data collection was carried out in the month of May, 2011which the number of clients constituted the population size, N. 350 diabetics were randomly interviewed and their folders were traced to capture other information with the help of a structured questionnaire. (Appendix I) This figure represented the sample size, n, that can be considered as large enough to be investigated according to the sample size theory and estimation (Cochran, 1963) since 350 >> 30.

The data collected were organized, summarized and coded on SPSS spreadsheet. Out of 350 respondents randomly selected, 310 were included in the study constituted 88.6% of the sample. The rest of the respondents were not included in the analysis due to misinformation and missing values. Table 4.1 shows the summary of the data used in the study. The first column of the table contains the variables, the second, third and the fourth containing the various categories of the variables, the frequency and the percentages respectively. Appendix II shows the full data.

Table 4.1 Summary of the data collected

Variable	Categories	Frequency	Percent
Gender:	Male	92	29.7%
	Female	218	70.3%
Residence:	Urban	223	71.9%
	Peri-urban	51	16.5%
	Rural	36	11.6%
Level of Education:	Not at all	88	28.4%
	Basic	156	50.3%
	Secondary	27	8.7%
	Tertiary	39	12.6%
Age:	40 – 49(years)	78	25.2%
	50 – 59(years)	101	32.6%
	60 – 69(years)	76	24.5%
	70 and above	55	17.7%
Physical Exercise	Regular (30min/day)	5	4.8%
	Sometimes	124	40.0%
	Not at all	171	55.2%
BMI	< 18(under weight)	9	2.9%
	18-24.9(normal)	115	37.1%
	25-29.9(over weight)	130	41.9%
	≥ 30 (obese)	56	18.1%

Occupation	unemployed	57	18.4%
	Formal	48	15.5%
	Trading	142	45.8%
	Farming	45	14.5%
	Retired	18	5.8%
Smoking	non smokers	300	96.8%
	Smokers	10	3.2%
Alcohol	non drunkards	209	67.4%
	Drunkards	101	32.6%
Salt intake	not at all	29	9.3%
	As in food	234	75.5%
	At table	47	15.2%

From the data in table 4.1, the percentage of female clients is 70.3% and that of the male counterparts is 29.7%. Table 4.1 revealed that 71.9% of the patients were urban dwellers while the rest were peri-urban and rural dwellers. It is also observed from Table 4.1 that greater number of diabetic patients who visit the K.A.T.H attained the lower level of formal education. About 50.3% reached the basic level while 28.4% have not had formal education at all. A greater number of these patients do not have time for physical exercise. Only 4.8% perform physical exercise for about 30 minutes daily, 55.2% do not perform any exercise at all. Most of these clients (45.8%) are traders. 75.5% take salt as contain in food and only 15.2% do not take at all. Considering smoking and alcoholic drinks, 3.2% smoke and 32.6% take alcoholic drinks.

4.1 Descriptive Statistics

Table 4.2 shows the comparative analysis of the two conditions with the continuous variables. The first column of the table described the two conditions, diabetes with hypertension and diabetes without hypertension with their respective sample size. The second column is the continuous variables with their corresponding minimum value, maximum value, mean, and standard deviation in third, fourth, fifth and sixth column in that order. The full table is placed on Appendix III.

Table 4.2:Comparative Analysis of the two conditions with continuous variables.

Condition	Variable	Minimum value	Maximum value	Mean	Std Deviation
Diabetes with Hypertension (sample size=198)	Height(metres)	1.34	1.87	1.62	0.08
	Weight (Kg)	42.0	103.0	70.24	11.94
	SBP	110	210	147.5	16.9
	DBP	60	190	85.2	14.11
	BGL	2.6	32.5	10.2	4.97
	BMI	16.0	38.0	26.6	3.79
	Age	40	87	61	10.2
Diabetes without Hypertension (sample size=112)	Height(metres)	1.40	1.87	1.63	0.09
	Weight (Kg)	35.6	97.7	62.3	13.1
	SBP	100	140	118.9	9.3
	DBP	60	100	74.2	7.7
	BGL	4.0	32.5	12.08	5.57
	BMI	14.0	35.0	23.3	4.28
	Age	40	90	51	11.2

It is observed from the Table 4.2 that the mean weight of the clients diagnosed with diabetes with hypertension is 70.2 as compare to 62.3 for non hypertensive diabetics. Similarly, the mean body mass index (BMI) for diabetes with hypertension patients is 26.6 kgm^{-2} as against 23.3 kgm^{-2} for the corresponding non hypertensive diabetics. Also the mean age of the diabetes with hypertension patients is 61years as compared to 51years for those without hypertension.

However, there is no much difference between the averages of the blood glucose level (BGL) of the clients for the two conditions as the mean level for hypertensive diabetics is 10.2 and that of the non hypertensive is 12.8. Again the mean height of the clients for both conditions is not different, 1.62 for hypertensive diabetic and 1.63 for non hypertensive diabetic. Considering the blood pressure (BP), the mean systolic BP for diabetes with hypertension is 147 mmHg while diabetes without hypertension clients is 118.9 mmHg.

Graphical Analysis on Occupation

Figure 4.1 displays the comparative analysis of the two cases by occupation. The number of the patients is on the vertical axis against the various occupations on the horizontal axis. The shaded bars represent diabetes with hypertension.

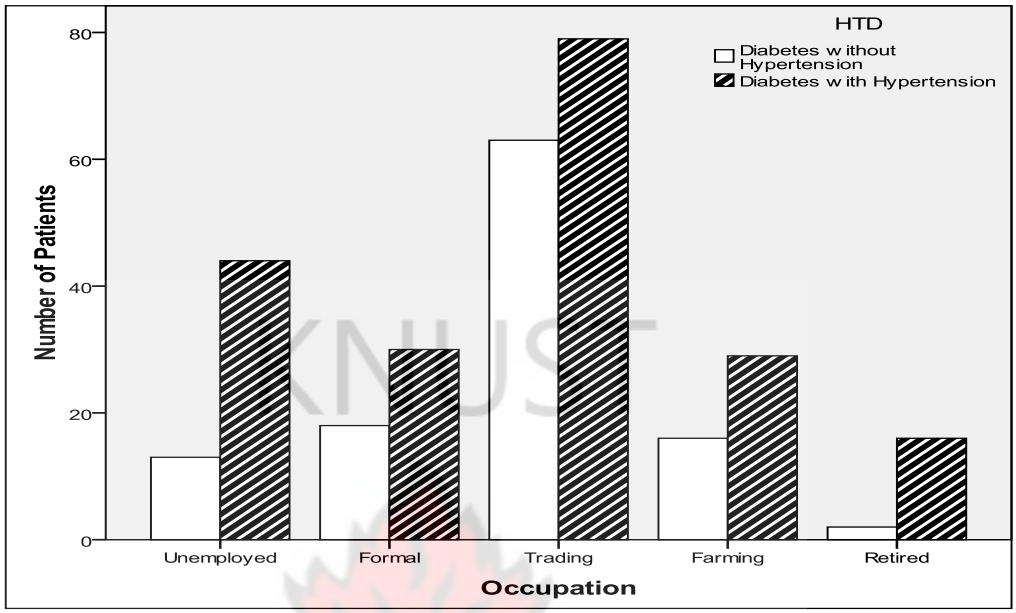

Figure 4.1: Comparative Analysis of the two cases by Occupation

From Figure 4.1 above, it is visible that most of the clients are traders in both diabetes with hypertension and diabetes without hypertension. However, diabetes with hypertension clients out numbered those without hypertension in all categories of occupation. For unemployed and retired clients, diabetes with hypertension is more than doubled the number of those without hypertension.

Graphical Analysis on Age Groups

Figure 4.2 shows the comparative analysis of the two cases by age groups. The vertical axis indicates the number of clients against the various age groups on the horizontal axis. The shaded bars represent diabetes with hypertension.

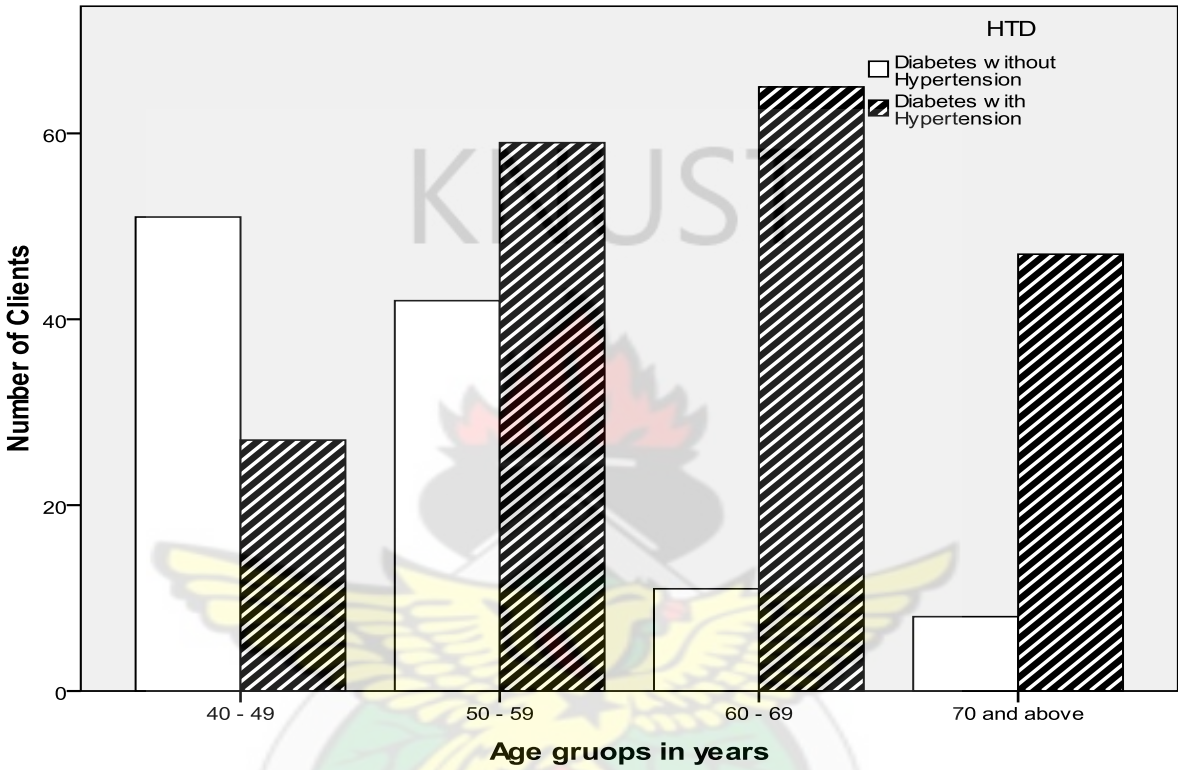

Figure 4.2: Comparative Analysis of the two cases by Age groups

Figure 4.2 above clearly shows that the number of non hypertensive diabetic clients decreases as the ages of patients increases from 40 years. With the exception of 40 – 49 years age group, the number of patients with diabetes with hypertension becomes larger as the age rises from 50 years with the majority of them within 50 – 69 years old. At 60 years and above, the number of hypertensive diabetic patients is more than doubled that of the diabetic without hypertension. This made it appears that at older age the probability of developing diabetes with hypertension is high as compared with that of diabetes without hypertension. Further investigation would be

conducted to find out whether the rise in age or the growth of clients can be a risk factor for developing diabetes with hypertension as the chart depicts.

Graphical Analysis on Body Mass Index

Figure 4.3 displays the comparative analysis of the two cases by body mass index (BMI). The number of the clients is on the vertical axis against the various groups of body mass index on the horizontal axis. The shaded bars represent diabetes with hypertension.

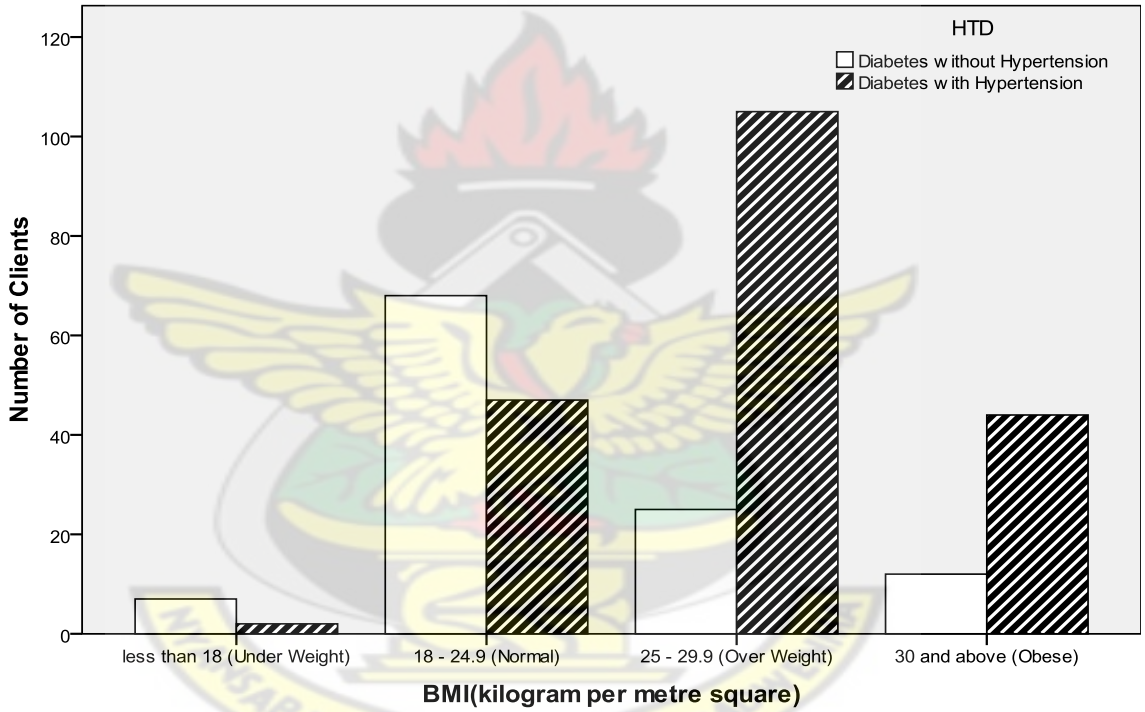

Figure 4.3: Comparative Analysis of the two cases by BMI

Figure 4.3 above give pictorial view of categories of BMI with respect to the number of patients with diabetes only and diabetes with hypertension. It could be observed from the chart that the number of hypertensive diabetic patients rises from BMI of 25kg/m^2 (over weight) to obese as those diagnosed with diabetes without hypertension decreases. Most of over weight patients are

hypertensive. This observation indicates that there is direct variation between the two variables, BMI and hypertension which would be investigated in the next section.

Graphical Analysis on Education

Figure 4.4 shows the comparative analysis of the two cases by age education. The vertical axis indicates the number of clients against the various levels of education on the horizontal axis. The shaded bars represent diabetes with hypertension.

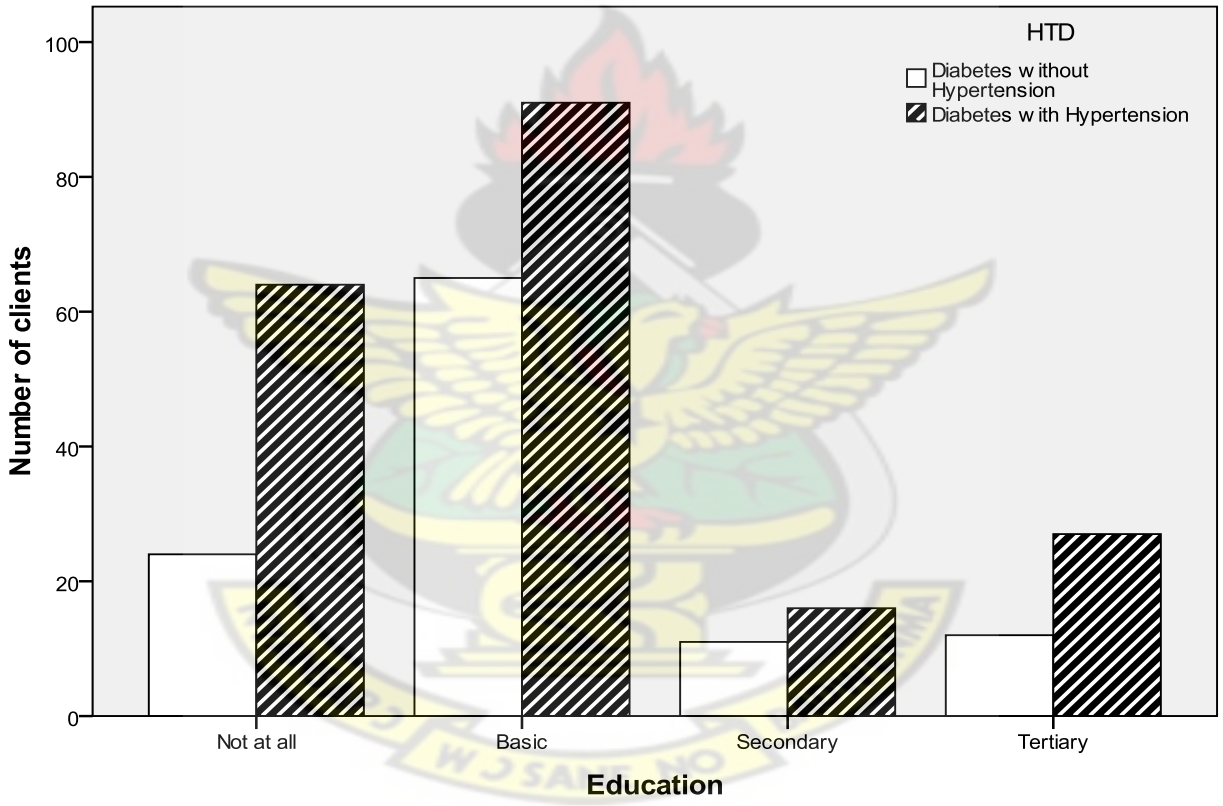

Figure 4.4: Comparative Analysis of the two conditions by Education

The Figure 4.4 above shows that higher proportion of the clients interviewed at the K.A.T.H attained only basic education or has not had formal education at all. A close look at these two

levels of education reveals that patients with diabetes with hypertension are concentrated there. This suggests that there is a relationship between hypertensive diabetic and formal education which should be investigated.

Line Graph of Mean Systolic Blood Pressure against Age Groups

Figure 4.5 shows the impact of age on systolic blood pressure by the two medical conditions, diabetes with hypertension and diabetes without hypertension. The mean systolic blood pressure on the vertical axis is plotted against the various age groups on the horizontal axis. The continuous line illustrates patients with diabetes with hypertension.

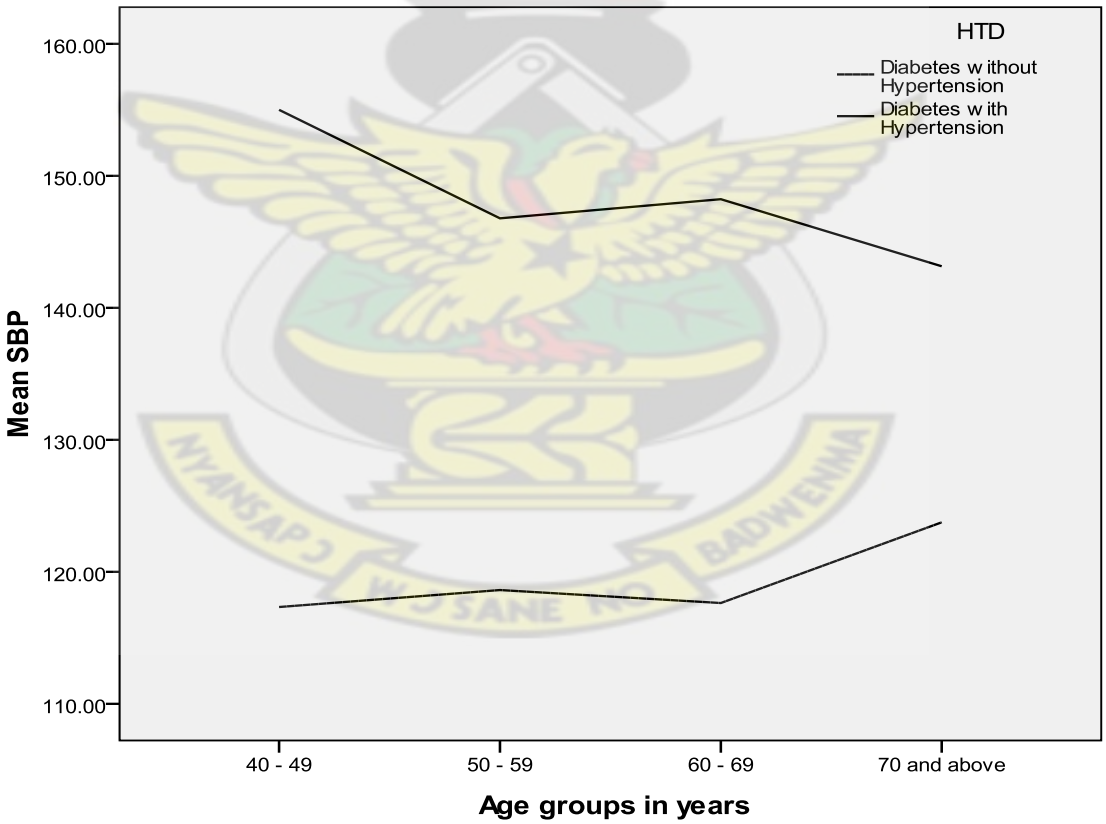

Figure 4.5: Impact of Age on systolic BP by the two conditions

Figure 4.5 above shows that the systolic BP of the patients with diabetes with hypertension is higher than those diagnosed with diabetes without hypertension. It appears that systolic BP among diabetes with hypertension patient decreases from age 40year to 59 and increases to 69 years and then decreases at older age. However, the systolic BP of the patients with diabetes without hypertension rises as the clients grow older. This indicates that there is positive correlation between systolic BP and age of the second group.

Line Graph of Mean Diastolic Blood Pressure against Age Groups

Figure 4.6 displays the impact of age on diastolic BP for each of the two medical conditions. The mean diastolic blood pressure on the vertical axis is plotted against the various age groups on the horizontal axis. The continuous line illustrates patients with diabetes with hypertension.

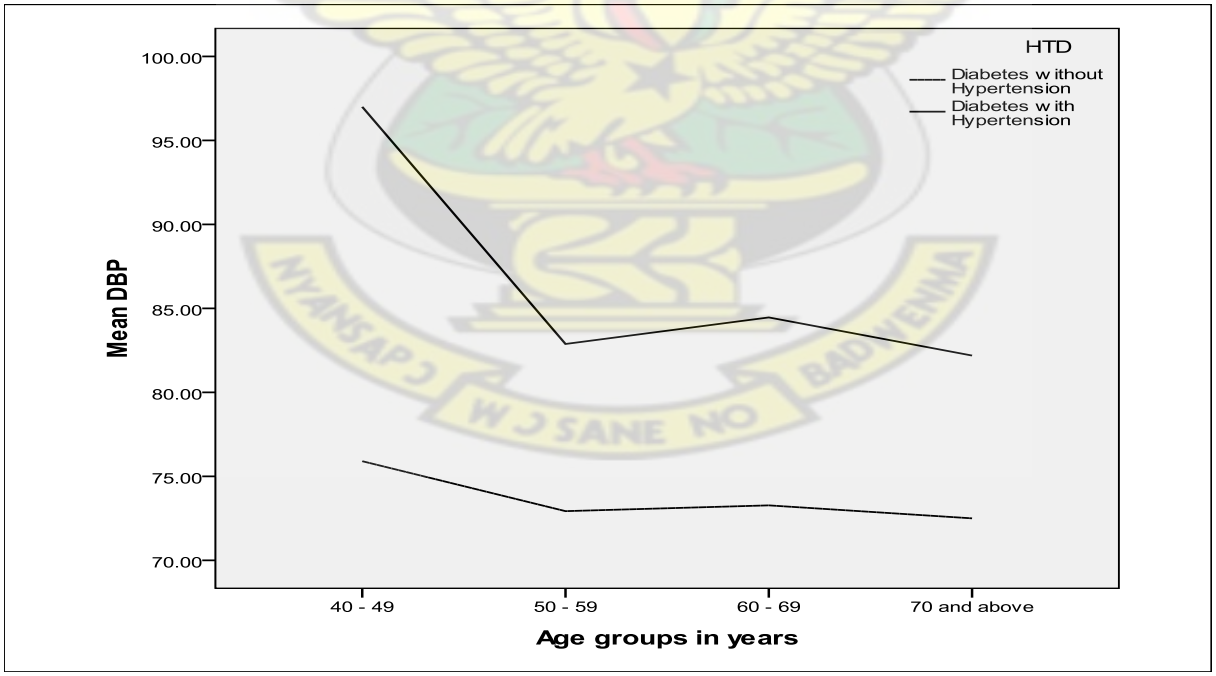

Figure 4.6: Impact of Age on Diastolic BP by the two Conditions

Likewise the systolic BP, the diastolic BP of the clients diagnosed with diabetes with hypertension is higher than those diagnosed with diabetes without hypertension. In both conditions, the diastolic BP decreases as the ages of patient increases from 40years and rise again as clients grow older.

Line Graph of Mean Body Mass Index against Age Groups

Figure 4.7 shows the impact of age on body mass index (BMI) by the two medical conditions, diabetes with hypertension and diabetes without hypertension. The mean BMI on the vertical axis is plotted against the various age groups on the horizontal axis. The continuous line illustrates patients with diabetes with hypertension.

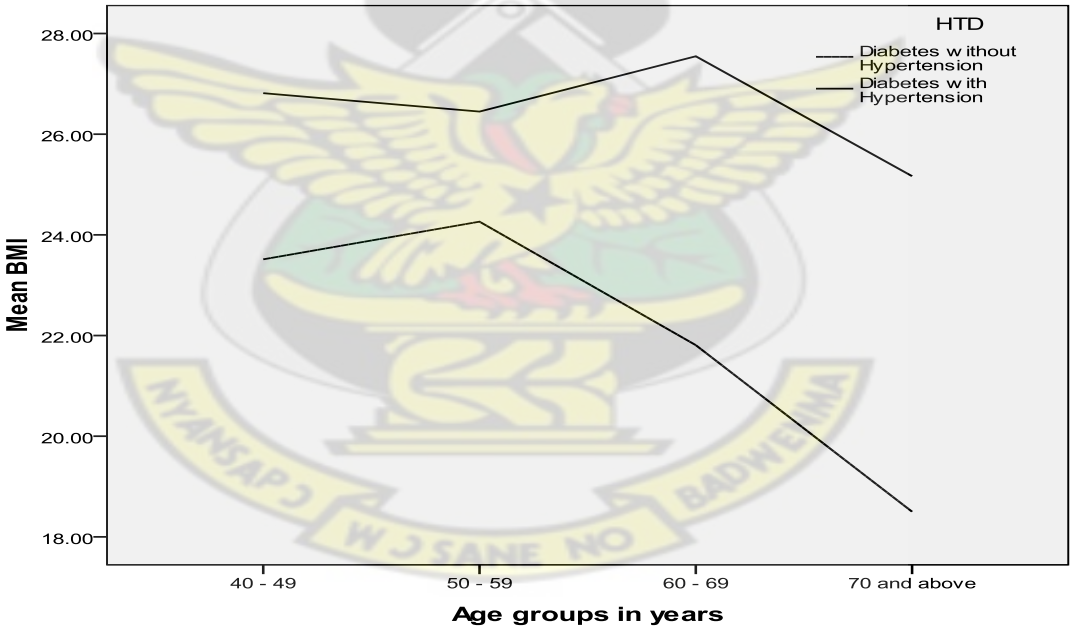

Figure 4.7: Impact of Age on BMI by the two conditions

From Figure 4.7 above, patients with diabetes with hypertension have higher body mass index compared with those with diabetes without hypertension. Considering the hypertensive diabetic patients, their BMI reduces from 40 – 49 year group to 50 – 59 years; it starts rising

again to 60 – 69 year group and decreases again. On the other hand, the BMI of non hypertensive diabetics rises up to 50 – 59 year group and decline as the patients becomes older.

Line Graph of Mean Body Mass Index against Age Groups (comparing gender)

Figure 4.8 presents the impact of age on body mass index (BMI) by gender. The mean BMI on the vertical axis is plotted against the various age groups on the horizontal axis. The continuous line illustrates the line graph of the female patients.

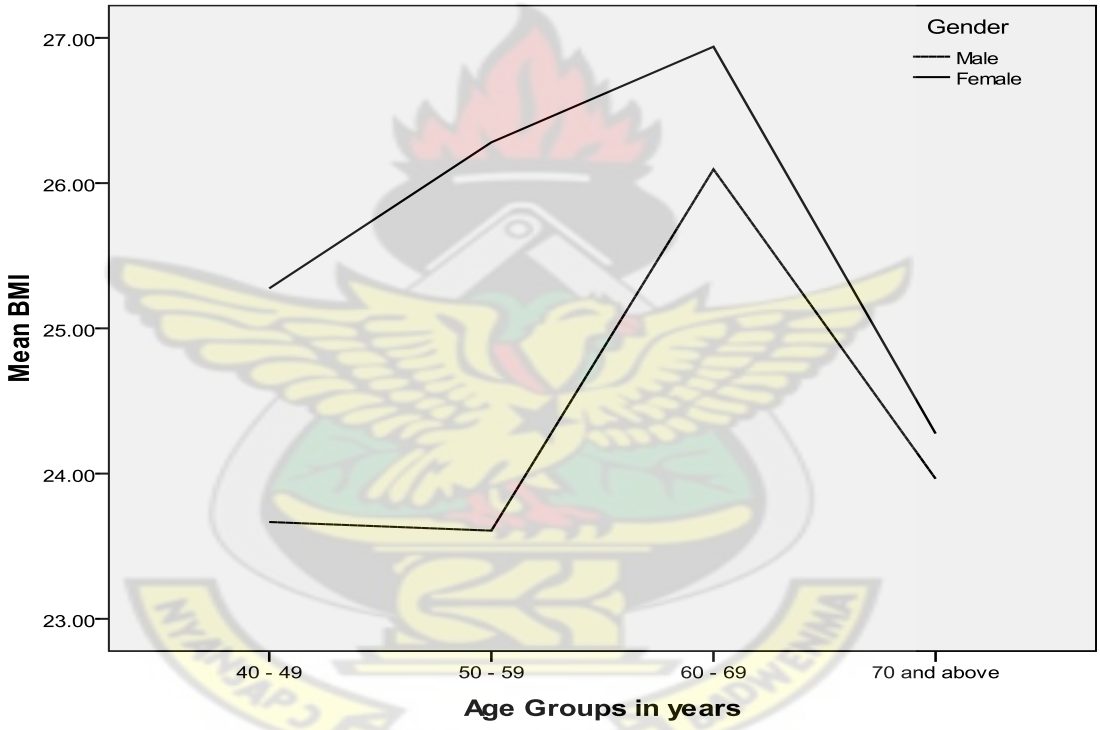

Figure 4.8: Impact of Age on BMI by Gender

A critical look at the Figure 4.8 above shows that females have higher BMI compared with the male patients. It seems in both male and female, BMI increases from 40 – 49 year age group through to 60 – 69 years and begins to decline as the patients become older. In both cases the BMI is higher in the 60 – 69 year group which is the retirement age.

Line Graph of Mean Blood Glucose Level against Age Groups

Figure 4.9 shows the impact of age on blood glucose level (BGL) by the two medical conditions, diabetes with hypertension and diabetes without hypertension. The mean BGL on the vertical axis is plotted against the various age groups on the horizontal axis. The continuous line illustrates patients with diabetes without hypertension

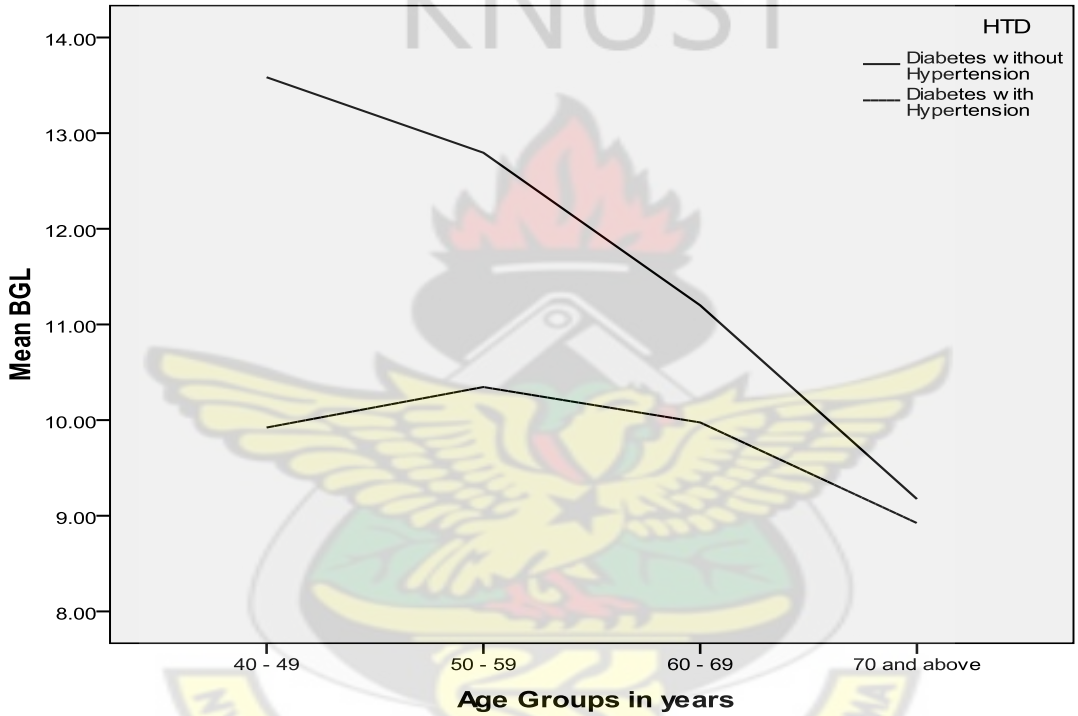

Figure 4.9: Impact of Age on BGL by the two conditions

The Figure 4.9 above shows that the patients who are hypertensive diabetic have lower blood glucose level than those without hypertension. In both cases, as the age of the patients rises, the BGL decreases. This picture suggests that there is negative correlation between BGL and hypertensive diabetics.

4.2 Statistical Analysis

This aspect of the analysis presents in detailed how the selected statistical techniques have been used to investigate the data and report the findings accordingly. The two statistical techniques the study deems it suitable for this investigation are binary logistic analysis (logit) and probit analysis as stated earlier. Detailed computer outputs of these two techniques would be observed and interpreted at 5% significance level (risk value).

4.2.1 Logit Analysis

Table 4.3 is the baseline classification table of the response variable, diabetes with hypertension when the predictor variables are not considered. It indicates the predictability of the model when it is fitted with only the constant.

Table 4.3: Classification Table

Observed			Predicted		
			HTD		
			Diabetes without Hypertension	Diabetes with Hypertension	Percentage Correct
Step 0	HTD	Diabetes without Hypertension	0	112	.0
		Diabetes with Hypertension	0	198	100.0
		Overall Percentage			63.9

The Table 4.3 above shows that fitting the model without any predictor variable gives the predictive ability of the model to be 63.9%. Thus, if one simple assume (guess) that all clients

are diagnosed as diabetes with hypertension, one would classify 63.9% of the clients correctly by chance. We expect this ability to improve when the predictors are included in the model.

A stepwise selection of the predictors into the model gives series of models. The full table is placed on Appendix IV. Table 4.4a and 4.4b displays Wald's Test of significance of the predictor variables in the initial and the final models respectively. The first column is the risk factors as predictor variables. The second column is the values of the unstandardized coefficients, β, column three is the Wald test, column four is the degree of freedom and column five is the significance level. The sixth column represents the odds ratio, which is the odds of diabetes with hypertension for each predictor against diabetes without hypertension and the seventh is the confidence interval.

Table 4.4a Wald's Test of Significance of Predictor Variables

Variable	Unst. Coeff.(B)	Wald	D.F	Sig	Exp(B)	95% C.I for Exp(B)	
						Lower	Upper
Initial Model							
Age	0.106	36.491	1	0.000	1.112	1.074	1.150
BMI	0.226	31.507	1	0.000	1.254	1.159	1.357
Occupation	0.010	0.005	1	0.945	1.010	0.757	1.349
BGL	-0.067	5.263	1	0.022	0.935	0.883	0.990
Exercise	0.412	2.269	1	0.132	1.510	0.883	2.583
Drinking	-0.109	0.120	1	0.729	0.897	0.484	1.662
Gender	0.216	0.399	1	0.528	1.242	0.634	2.431
Residence	-0.099	0.214	1	0.644	0.906	0.597	1.376
Education	0.264	2.098	1	0.147	1.303	0.911	1.863
Constant	-11.821	40.172	1	0.000	0.001		

Table 4.4b Wald's Test of Significance of Predictor Variables

Variable	Unst. Coeff.(B)	Wald	D.F	Sig	Exp(B)	95% C.I for Exp(B)	
						Lower	Upper
Final Model							
Age	0.107	41.646	1	0.000	1.113	1.077	1.150
BMI	0.236	35.590	1	0.000	1.266	1.171	1.367
BGL	-0.065	5.443	1	0.020	0.937	0.887	0.990
Constant	-10.581	45.620	1	0.000	0.001		

Considering the predictors with their respective parameters (β) from the above table after the stepwise method adopted, using Equation 3.18, the final model is:

$$\text{Logit}(Y) = 1n\left(\frac{\pi}{1-\pi}\right) = -10.581 + 0.107\text{Age} + 0.236\text{BMI} + (-0.065\text{BGL}) \qquad (4.1)$$

This implies that to determine or predict whether a client would be hypertensive diabetic or non hypertensive, the age, BMI and BGL of that client could be relevant factors. Therefore, a rise in these variables could cause the diabetic patient to develop hypertension in addition.

Prediction

Using the client number 14 who is 55 years old with BMI = 27 kgm^{-2} and BGL = 10.5 mmol, the probability that the client will be classified as hypertensive diabetic is calculated by

$$1n\left(\frac{\pi}{1-\pi}\right) = -10.581 + 0.107(55) + 0.236(27) + (-0.065(10.5))$$

$$1n\left(\frac{\pi}{1-\pi}\right) = -10.581 + 5.885 + 6.372 - 0.6825$$

$$1n\left(\frac{\pi}{1-\pi}\right) = 0.9935$$

Therefore, the resulting odds ratio (OR) is $e^{0.9935} = 2.70$, which implies the probability for that client to be classified as hypertensive diabetic is almost thrice.

Omnibus Test

Table 4.5 shows the goodness of fit test of the model. Omnibus Test of model coefficients indicate how well the initial model (model with the predictors) performs over and above the predictive ability of the model with none of the predictors entered into it.

Table 4.5: Omnibus Test of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	120.633	9	.000
	Block	120.633	9	.000
	Model	120.633	9	.000

From Table 4.5 above, the significant or probability value (p-value = 0.000) for the model is less than the risk (α-value = 0.05). Therefore, the model with set of predicted is better than SPSS original guess (model without predictors).

Hosmer and Lemeshow Test

Table 4.6 shows the Hosmer and Lemeshow Test which supports the omnibus test for the model. This is the one of the reliable test of model in SPSS output. But its test of significance is direct opposite to omnibus test. Thus, the model is well fit when the probability value or the significance value is greater than 0.05.

Table 4.6:Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	8.936	8	.348

From Table 4.6, the significance value of the Hosmer- Lemeshow test (0.348) is greater than 0.05. This is therefore enough evidence for the test to support the existing model.

Cox and Snell Test

Table 4.7 provides the pseudo R-square values ranging from 0 to 1. This provides the amount of variation of the response variable explained by the initial model.

Table 4.7: Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	284.944[a]	.322	.442

From Table 4.7, the values of Cox and Snell R square (0.322) and Nagelkerke R square (0.442) suggest that between 32.2% and 44.2% variability in the response variable is explained by the set of predictors in the initial model. However, when the stepwise method adopted for the variable selection for the final model, only age, BMI and BGL emerged as significant predictors. This model account for between 31.3% and 42.9% variability in the response explained by the three predictors, suggesting a reduction in power to explain variability in the response due to the drop of some predictors.

Classification Test

Table 4.8 displays the percentage of classification for the initial and the final model. It shows how well the model is able identify the correct category, diabetes with hypertension and

diabetes without hypertension using sensitivity analysis. The values in bracket are for the final model.

Table 4.8: Classification Table for initial and the final model

Observed		Predicted		
		Diagnosis		
		Diabetes without Hypertension	Diabetes with Hypertension	Percentage Correct
Diagnosis	Diabetes without Hypertension	71(72)	41(40)	63.4(64.3)
	Diabetes with Hypertension	26(21)	172(177)	86.9(89.4)
Overall Percentage				78.4(80.3)

From the Table 4.8 above, the percentage accuracy in classification (PAC) for the initial model is 78.4% while the final model is 80.3%. This indicates an improvement in classification when the final model is used. The two percentages can further be categorized into sensitivity of the model (for that of diabetes with hypertension) and the specificity of the model (for that of diabetes without hypertension) since the former has the characteristic of interest. Considering the final model, the sensitivity of model in classifying clients with diabetes with hypertension is 89.4% and the specificity of model to classify clients with diabetes without hypertension is 64.3%. Therefore the positive predictive value $=\left(\frac{177}{177+21}\right) = \frac{177}{198} = 89.4\%$ and

the negative predictive value $= \left(\frac{72}{72+40}\right) = \frac{72}{112} = 64.3\%$. This means the model is able to predict 89.6% of clients diagnosed with diabetes with hypertension and 64.3% of diabetes without hypertension.

4.2.2 Probit Analysis

Table 4.9 shows the parameter estimates of the probit model. The first column is the risk factors as predictor variables. The second column is the estimates of the parameter, β, column three is the standard error, column four is the Z test statistic and column five is the significance level. The last column represents the confidence interval.

Table 4.9:Parameter Estimates of the Probit Model

						95% Confidence Interval	
	Parameter	Estimate	Std. Error	Z	Sig.	Lower Bound	Upper Bound
PROBIT[a]	Age	.026	.005	4.816	.000	.016	.037
	BMI	.058	.013	4.344	.000	.032	.085
	Occupation	.001	.049	.021	.983	-.094	.096
	BGL	-.024	.011	-2.085	.037	-.046	-.001
	Drinking	-.009	.120	-.071	.943	-.243	.226
	Education	.077	.064	1.194	.233	-.049	.203
	Residence	-.020	.079	-.259	.796	-.176	.135
	Gender	.081	.132	.612	.541	-.178	.339
	Exercise	.111	.105	1.058	.290	-.095	.317
	Intercept	-3.758	.608	-6.184	.000	-4.366	-3.151

a. PROBIT model: PROBIT(p) = Intercept + BX

From Table 4.9, the significant predictors are age, BMI and BGL. Considering their respective estimates (β), using Equations (3.27) and (3.28) we have the probit model:

$$\text{Probit}(\pi_i) = \Phi[(-3.758) + 0.026\text{Age} + 0.058\text{BMI} + (-0.024\text{BGL})] \quad (4.2)$$

where Φ is the standard normal Cumulative distribution function.

Prediction

Using the client number 14 who is 55 years old with BMI = 27 kgm^{-2} and BGL = 10.5 mmol, the probability that the client will be classified as hypertensive diabetic is calculated by

$$\text{Probit}(\pi_i) = \Phi[(-3.758) + 0.026(55) + 0.058(27) + (-0.024(10.5))]$$

$$(\pi_i) = \Phi[(-3.758) + 1.43 + 1.566 - 0.252]$$

$$(\pi_i) = \Phi(-1.014)$$

$$\pi_i = 0.8729$$

Pearson Goodness of Fit Test of the Model

Table 4.10 shows the Person Goodness of Fit Test of the model. It has a Chi-Square distribution in the second column with degree of freedom in the third column. The model is well fit when the probability value or the significance value in column four is greater than 0.05.

Table 4.10: Chi-Square Tests

		Chi-Square	df[a]	Sig.
PROBIT	Pearson Goodness-of-Fit Test	137.838	300	1.000

From Table 4.10, the Chi-square value is 137.838 with degree of freedom 300. The significance value (1.000) is greater than 0.05. Therefore, there is enough evidence to conclude that the model is well fit.

4.3 Discussion

Among 198 (63.9%) hypertensive diabetic patients 147 (74.2%) were females, which forms 47.4% of the total sample. In general, females out numbered their male colleagues in both diabetes and diabetes with hypertension cases. These findings seem to be consistent with the work of many researchers on diabetes. According to WHO (2010) the prevalence of diabetes is higher among men in Colombia, England, Mexico, and the US, and among women in Iran and Thailand. But the diagnosis rates were higher in women than in men. Also King et al. (1998) conducted a research on prevalence of diabetes in adults worldwide. It was revealed that prevalence by gender indicates that diabetic prevalence is higher in men than women, but there are more women with diabetes than men.

The models are fitted using all the eligible risk factors obtained. The logit and probit analysis indicated that the two medical conditions do not differ so far as the following predictors: occupation, education, residence, gender, physical exercise and drinking are concerned. On other hand, Age, BMI and BGL contributed significantly to the development of the model. It is clearly shown in Figure 4.2 that the number of hypertensive diabetics raises as the ages of patients increases with the majority of them within 50 – 69 years old. At 60 years and above, the number of hypertensive diabetic patients is more than doubled that of the diabetic without hypertension. For instants, the numbers of hypertensive diabetics who fall within the ages 50-59, 60-69, 70 and above are 59, 65, and 47 respectively as compare to 42, 11, and 8 for the non hypertensive diabetics. It is therefore not surprising that age as a risk factor contributed positively to the modeling.

The pictorial view of categories of BMI in Figure 4.3 indicates that the BMI of hypertensive diabetics rises from BMI of 25kg/m^2 with most of over weight patients being

hypertensive. Similarly, Figure 4.7 presented the mean body mass index for patients with diabetes with hypertension and diabetes without hypertension. It shows patients with diabetes with hypertension have higher body mass index compared with those with diabetes without hypertension. This observation indicates that there is positive correlation between the two variables, BMI and hypertensive diabetics. This result is consistent with the study by Asare (2008), 'model characterizing the Hypertensive and non-hypertensive diabetic patients'. BMI and age were discriminating factor.

The patients who are hypertensive diabetic have lower blood glucose level than those without hypertension as indicated in Figure 4.9. In both cases, as the age of the patients rises, the BGL decreases. This picture suggests that there is negative correlation between BGL and hypertensive diabetics. However, in both logit and probit analysis, BGL proved to be significance to the modeling.

The study also revealed that education, occupation and place of residence are not significant predictors. It was observed that most of the clients have no formal education or only attained basic level. Also the majority of these clients were traders and due to the nature of the work, most of them do not perform any vigorous physical exercise.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.0 Conclusion

Among the predictor variables considered in the study to differentiate between the two medical conditions (diabetes with hypertension and diabetes without hypertension), Age, BMI and BGL emerged significance. In both logit and probit models, these risk factors are the significant predictors of hypertensive diabetics. This confirms the theoretical aspect of chapter three that logit and probit analysis produce similar results. The two well-fitted models are:

$$\text{Logit}(Y) = 1n\left(\frac{\pi}{1-\pi}\right) = -10.581 + 0.107\text{Age} + 0.236\text{BMI} + (-0.065\text{BGL}) \quad \text{and}$$

$$\text{Probit}(\pi_i) = \Phi[(-3.758) + 0.026\text{Age} + 0.058\text{BMI} + (-0.024\text{BGL})].$$

Using the model for prediction, a client who is 55 years old with BMI = 27 kgm^{-2} and BGL = 10.5 mmol, the probability that this client will be classified as hypertensive diabetic is calculated by

$$1n\left(\frac{\pi}{1-\pi}\right) = -10.581 + 0.107(55) + 0.236(27) + (-0.065(10.5))$$

$$1n\left(\frac{\pi}{1-\pi}\right) = -10.581 + 5.885 + 6.372 - 0.6825$$

$$1n\left(\frac{\pi}{1-\pi}\right) = 0.9935$$

$$\text{Probit}(\pi_i) = \Phi[(-3.758) + 0.026(55) + 0.058(27) + (-0.024(10.5))]$$

$$(\pi_i) = \Phi(-1.014)$$

$$\pi_i = 0.8729$$

5.1 Recommendations

The study came out with two models on logit and probit analysis. We hope that these models will go a long way to assist physicians to make future prediction or to determine who would be hypertensive diabetic or non hypertensive. This will give the health educators information to disseminate as part of their sensitization of general public on management of diabetes and diabetes with hypertension.

The researchers and health educators should make diabetes patients aware that ageing is one of the risk factors to develop diabetes with hypertension. It is unfortunate additional number of years gained among diabetes patients (which suppose to a blessing) could turn to be a risk factor. It is therefore necessary for non hypertensive diabetics to check their lifestyles so as to avoid developing hypertension.

Again, body mass index (BMI) is identified to be another risk factor to develop diabetes with hypertension. Since rise in weight lead to the corresponding rise in BMI, it should be made known to diabetes patients that extra weight gain could be a risk factor for diabetes with hypertension. For this reason diabetes patients should be introduced to some physical exercises that will enable them to maintain their normal weight.

REFERENCES

Aekplakom, W., Stolk, R. P., Neal, B., Suriyawongpial, P., Cheepudomwit, S. and Chongsuvivatwong, V. (2003). The prevalence and management of diabetes in Thia dults. **Diabetes care** 26(10), 2758-2663.

Ahmed, T.(2005). Logit Models for Bankruptcy Data Implementation in XploRe. Unpublished Master of Science Thesis, Humboldt-University, Berlin.

Al-Moosa, S., Allin, S., Jemiai, N., Al-Lawati, J. and Mossialos, M. (2006). Diabetes and Urbanization in the Omani population: an analysis of national survey data. **Population Health Matrice. 4.5.** http://www.pophealthmatrices.com/content/4/1/5 (assessed on June 6,2010)

American Diabetes Association and the American Heart Association. Circulation. 2006;113:2943-2946; Diabetes Care. 2006;29:1697-1699.

Amoah, K., Opoku, M. and Adu, K. E. (2003). Assessment of Risk Factors for Cardiovascular Diseases in Ashanti Region of Ghana. *American Heart Association*, 40, 48-53.

Asare, P. K. (2008). Model Characterizing Hypertensive and Non Hypertensive Diabetic patients. Unpublished Mphil thesis in Statistics. University of Cape Coast, Cape Coast.

Bassuk, S., Manson, J. (2005). Epidemiological evidence for the role of physical activity in reducing risk of type 2 diabetes and cardiovascular disease. *Journal of Applied Physiology*.

99: 1193-1204. Doi:10.1152/japplphysiol.00160.2005. 8750 -7587/05. (Accessed on 20 August 2010).

Bays, H. E., Chapman, R. H., and Grandy, S. (2007). The relationship of body mass index to Diabetes mellitus, hypertension dyslipidemia: comparison of data from two national *International Journal of Clinical Practice*. 61(5) 737-747.

Bierens, H. J. (2008). The Logit Model: Estimation, Testing and Implementation. http:www.med.logit.sav com. (Assessed on 15 March 2011)

Bowden, D.W., Rudoc, M., Ziegler, J. et al (2006) Coincident linkage of Type 2 diabetes, metabolic syndrome, and measures of cardiovascular disease in a genome scan of the diabetes heart study. *Diabetes* 5(5) 1985-1994.

Czepiel, S. A. (2005). Maximum Likelihood Estimation of Logistic Regression Models: Theory and Implementation. Published paper. Available on:http://czep.net/contact.html. (Assessed on April 5, 2011)

Daousi, C., Casson, I.F., Gill, G.V. et al (2006) Prevalence of obesity in type 2 diabetes in Secondary care: association with cardiovascular risk factors. *Postgraduate Medical Journal* 82 280-284. **Diabetes Care**. 26(10), 2758-2763.

Everitt, B. S. (2006). The Cambridge Dictionary of Statistics,Third Edition, Cambridge University Press.

Fox, J. and Spida, Y. (2010). Logit and Probit Model. http:www.med.logit.sav com. (Assessed

on 15 March 2011)

Foy, C. G., Bell, R. A., Farmer, D. F., Goff, D. C., and Wegenknecht, L. E. (2005). Smoking and Incidence of diabetes among US adults. **Diabetes Care**. 28(10) 2501-2507.

Francesco, M., Harld, K. P., and Frank, P. (2006). Prevalence, Management and Control of Hypertension. *Human Hypertension*, 50, 23-25.

Gakidou, E. et al.(2011). Management of diabetes and associated cardiovascular risk factors in seven countries: a comparison of data from national health examination surveys. *Bulletin of the World Health Organization*, 2011;89:172%u2013183 DOI: 10.2471/BLT.10.080820

Ghaffar, A., Reddy, K.S., and Singhi, M. (2004). Burden of non-communicable diseases in SouthAsia. BMJ. 328(7443), 807-810.

Hair J. F. Jr., Anderson R. E., Tatham R. L. Black W. C. and Babin B. J.,(2006), Data Analysis, Sixth Edition, Upper Saddle River, New Jersey, Prentice Hall.

Harrison, T.A., Hindorff, L. A., Kim, H., Wines, R.C., Bowen DJ, McGrath, B.B. et al. (2003). Family history of diabetes as a potential publichealth tool. American Journal of Preventive Medicine. 24 (2): 152-159. Doi: 10.1016/S0749-3797(02)00588-3 (Accessed on: 6 September, 2010).

Health Matrice. 4.5. http://www.pophealthmetrice.com/content/4/1/5. 09/11/2010.

Hobbs, F. (2004). Cardiovascular disease: different strategies for primary and secondary prevention*?*. *Heart.* 90(10):1217-1223. 99Doi: 10.1136/hrt.2003.027680 (Accessed on 10

August 2010)

Hosmer, D. W. and Lemeshow, S. (1989). Applied Logistic Regression, New York; Wiliey.

Johnson, G. J. (1993). A note on Fitting the Probit Model in SAS Institute Inc, Cary, NC.

Kim, Y., Suh, Y.K., and Choi, H. (2004). BMI and metabolic disorder in South Korean adults: 1998 Korean National Health and Nutrition Survey. **Obesity Research**. 12(3), 445-453

King, R., and Henry, K. (1997). The Burden of Mortality Attributable to Diabete. *Diabetes Atlas* 2000. 10, 20-23.

King, R., Donald, B. K. and Wiley, B. (1998). Global Prevalence of Diabetes. *Diabetes Atlas*2000. 11, 25-28.

Kleinbaum D. and Klein M, (1994), Logistic Regression-A self-Learning Text, Second Edition, Springer, New York.

Kumar, P., Clark, M . (2002). Clinical medicine, fifth edition, Saunders.

Long, J. S. (1997). Regression Model for Categorical and Limited Dependent variables. Advanced Quantitative Techniques in the Social Sciences. Sage Publications.

Mahadeven, S. (1998). Logistic Regression and General Linear Model. University of Massachusette. Available on: mahadeva@cs.unass.edu (Assessed on 11 November 2010)

Melidonis, A., Tournis, S., Kompoti, M. G., Lentzas, J., Roussou, V., and Iraklianou, S.(2006)Increasing prevalence of diabetes mellitus in rural Greek population. Available onhttp://rrh.deakin.edu.au. (Assessed on March 3, 2010)

Mensah, E., Gyemfi, C. and Hamidu, A.(2006) Characteristic of Hypertensive and Diabetic Patient. Upublished HND Project Report. Department of Statistics, Kumasi Polytechnic.

Micah, F., and Agyenim, B.(2003). Hypertension and Renal Failure in Kumasi at Komfo

AnokyeTeaching Hospital. *Cardio Journal*, 6, 45-51.

Microsoft Encarta.(2009).

Milton, J. S. and Arnold, J. C. (1995). Introduction to Probability and Statistics, Third Edition. McGraw-Hill Inc, New York.

Mohan, V., Sandeep, S., Deepa, R., & Varghese, C.(2007). Epidemiology if Type 2 Diabetes: Indian scenario. Indian J Med Res 125(3) 217- 320.

Montonen, J., Knekt, P., Jarvinen, R., and Reunanen, A.(2004) Dietery antioxidant intake and Risk of type 2 diabetes. *Diabetes Care*. 27(2) 363-366.

Neil H. T. (2002), Applied Multivariate Analysis, Springer-Verlay, New York Inc.

Pickup, J.C. and Williams, G. (2003). Textbook of diabetes, Oxford: Blackwell Science

Rosemary, N., Thompson, R. P., and Anna, B. (2006). Women's Health Study to Assess Blood Pressure in urban Women in Accra. *American Heart Association*. 40, 30-34

Rush, S. (2001). Logistic Regression: The Standard Method of Analysis in Medical Research. Available on: http.www.med.logistic.sav.com (Assessed on 15 May 2011)

Sharma S. (1996). Applied Multivariate Techniques, John Wiley and Sons Inc. New York.

Storving, M., Donald, R., and Ricky, C. G. (2003). Prevalence of Obesity and Obesity as risk Factor. The Fact Sheet. 30, 20-22

Unger, R. H., Foster, D.W. (1998) Diabetes mellitus. In: Wilson JD, Foster DW, Kronenberg HM, Larsen, P. R, eds. *Williams Textbook of Endocrinology*. Philadelphia, Pa: WB Saunders Co; 1998:973–1059.

Vasisht, A. K. (2008). Logit and Probit Analysis. I.A.S.R.I., Library Avenue, New Delhi-110-012. Available on: http:/amitvasisht@lasri.res.in (assessed on 20 January 2011)

Walpole, R. E., Myers, R. H. and Ye, K. (2007). Probability and Statistics for Engineers and

Scientist, Eight Edition. Pearson Education Inc, Upper Saddle River, New York.

Wild, S., Roglic, G., Green, A., Sicree, R., and King, H. (2004). Global Prevalence of diabetes. *Diabetes Care*, 27(5).1047-1053

William M., Dennis D. W. and Richard L. S., (1990). Mathematical Statistics with Application, Fourth Edition. PWS-KENT Publishing Company, USA.

World Health Organization (2008). Diabetes, Media centre. Available on: http://www.who.int/mediacentre/factsheets/fs312/en/ (Accessed on November 3, 2010)

World Health Organization. (2006). Obesity and overweight. WHO Media centre. available on: http://www.who.int/mediacentre/factsheets/fs311/en/index.html (Accessed on 11 November 2010)

World Health Organization. (2007). Cardiovascular diseases. Available on: http://www.who.int/mediacentre/factsheets/fs317/en/print.html. (Accessed on 20 November 2010).

World Health Organization (2011). Management of diabetes and associated cardiovascular risk factors in seven countries: A comparison of data from national health examination surveys. Bulletin of the WHO 2011; 89:172-183.

APPENDICES

APPENDIX I
Questionnaire

Code No ……..

NAME:…… … ID/M:……

RISK FACTORS

1. Gender: Male [ ] Female [ ]

2. Age: ……

3. Place of residence: …… Urban[ ] Peri urban [ ] Rural [ ]

4. Marital Status: Single [ ] Married [ ] Divorced [ ] Widow/er [ ]

5. Level of Education: Not at all [ ] Basic [ ] Secondary [ ] Tertiary [ ]

6. Occupation: Unemployed [ ] Formal sector [ ] Trading [ ] Farming [ ] Retired [ ] other (specify) ……

7. Weight: ……kg

8. Height: ……..m

9. Body Mass Index (BMI) ……kg/m^2

LIFESTYLES

10. Do you smoke currently? Yes [ ] How many sticks/day ……....

No [ ]

11. Have you ever smoked? Yes [ ] How many sticks/day ……..

No [ ]

12. Did you use to take alcoholic drinks? Yes [ ] How many bottles/day …….. type……..

No [ ]

13. Do you take alcoholic drinks currently? Yes [ ] How many bottles/day……. type……..
No [ ]

14. Level of salt usage: Not at all [ ] As contained in the food [ ] At table [ ]

15. How often do you exercise? Regular (30 min/day) [ ] sometimes [ ] Not at all [ ]

CLINICAL INDICATORS AND HISTORY

16. Blood Pressure – Systolic(BP): ……..mmHg Diastolic(BP): ……mmHg

17. Blood Glucose Level (BGL): ……..mmol

18. When did you become Hypertensive: ……..

19. How long have you been Hypertensive: ……

20. Family Clinical History: Diabetes [ ] Hypertension [ ] Diabetes and Hypertension [ ] other (specify) ……

21. Relationship with the subject …….

22. Treatment type (tick as many as applicable):

For Diabetes: Diet [ ] Oral [ ] Insulin [ ]

For Hypertension-How many dose of: ACFI……, DRB…….

23. Level of cholesterol: ……mmol/L or ……mg/dl

HDL [ ] LDL [ ] Triglyceride [ ]

APPENDIX II

THE DATA

Serial No	Diagnose	Gender	Age	BMI	SBP	DBP	BGL	Residence	Education	Occupation	Smoking	Drinking	Exercise	Salt
1	1	1	64	29	160	90	10.6	1	2	3	0	0	3	2
2	1	1	60	29	190	140	11.5	1	1	1	0	0	3	1
3	0	1	48	27	119	86	7.2	1	4	4	0	0	2	3
4	1	1	63	28	120	70	7.3	1	2	3	0	0	3	2
5	1	1	75	25	150	80	11.9	2	1	1	0	0	3	2
6	1	0	58	28	140	80	6.3	2	3	2	0	0	2	2
7	0	0	63	25	140	70	7.1	1	3	4	0	1	3	2
8	1	1	63	24	140	70	6.4	3	1	3	0	0	3	2
9	1	1	60	32	160	70	10	1	2	3	0	0	3	3
10	1	1	57	26	140	90	9.4	1	1	1	0	0	3	2
11	1	1	74	26	170	100	8.9	3	1	4	0	0	3	2
12	1	1	47	25	160	80	14	2	1	4	0	0	3	1
13	1	1	70	32	150	80	6.2	1	1	3	0	0	3	2
14	1	1	55	27	120	70	10.5	1	2	3	0	0	3	1
15	1	0	55	23	160	70	8.8	2	1	4	0	0	3	2
16	1	1	45	22	150	70	6.3	3	1	4	0	0	3	2
17	1	1	67	32	150	70	6.1	1	2	3	0	0	3	1
18	1	1	59	30	140	80	5.4	2	4	2	0	0	2	3
19	1	1	60	30	130	80	18.6	1	1	4	0	0	3	2
20	0	1	41	22	100	60	9.7	3	2	4	0	0	2	3
21	0	1	40	22	130	70	22.6	1	2	3	0	0	2	2
22	1	1	60	27	140	90	18.8	1	4	2	0	0	2	2
23	1	0	50	29	160	70	6.2	1	2	3	0	1	2	2
24	1	1	82	26	160	100	10.1	1	1	1	0	0	3	2
25	0	1	50	29	120	70	12.3	1	2	3	0	0	3	1
26	1	1	69	31	140	90	14.5	1	2	1	0	0	3	2
27	0	0	52	20	120	70	6.1	2	2	2	0	1	3	2
28	0	1	65	25	120	80	9.7	2	1	1	0	0	3	2
29	1	1	68	28	150	70	6.7	1	2	3	0	0	3	2
30	0	1	45	21	120	80	9.4	1	2	1	0	0	2	3

31	1	1	72	27	150	100	9.7	1	1	1	0	0	3	2
32	0	1	40	22	120	80	14.6	1	2	3	0	0	3	1
33	1	0	60	34	150	70	13.2	2	2	4	0	1	2	2
34	1	1	76	25	130	70	4.6	1	1	1	0	0	3	2
35	0	1	54	25	110	70	7.5	1	4	2	0	0	2	1
36	1	0	69	26	150	70	5.9	1	3	5	0	1	3	3
37	1	1	57	30	140	80	24.7	3	1	4	0	0	3	2
38	1	1	84	26	160	100	5.6	1	2	3	0	0	3	1
39	1	0	79	25	140	80	6.4	1	1	3	0	0	2	2
40	0	0	62	19	110	70	13.7	2	2	4	0	0	3	2
41	0	1	56	22	100	70	24.2	3	1	3	0	0	3	3
42	1	1	53	30	180	100	12.4	1	2	3	0	1	3	2
43	1	1	50	32	130	70	6.1	2	2	3	0	0	3	2
44	0	0	40	22	110	70	10.1	1	3	3	0	1	2	2
45	1	1	72	28	130	70	21.2	1	4	4	0	1	3	2
46	1	1	53	29	160	90	2.6	1	2	3	0	0	3	2
47	1	1	63	31	130	70	4.3	2	3	5	0	0	3	2
48	0	0	54	23	180	80	10.5	3	2	4	0	1	3	2
49	1	0	60	32	150	90	5.4	3	1	3	0	0	3	2
50	0	1	44	24	130	80	6.5	1	2	3	0	0	2	2
51	1	1	50	26	140	70	7.6	1	2	4	0	1	3	2
52	0	1	52	26	120	70	7.8	1	1	1	0	0	3	2
53	1	0	51	22	160	70	14	1	4	2	0	1	2	2
54	1	1	63	32	150	90	8.3	2	2	1	0	0	3	2
55	1	1	40	32	140	90	8.3	2	2	3	0	0	3	2
5	0	1	52	30	120	70	18.9	1	3	2	0	1	3	2
57	1	1	62	32	130	80	4.3	1	1	4	0	1	3	2
58	1	0	64	25	145	90	4.9	2	4	5	0	1	3	2
59	1	1	44	23	160	90	5.8	2	2	3	0	0	3	2
60	0	1	50	24	110	70	9	1	2	3	0	1	2	2
61	0	0	60	22	110	80	22	1	2	3	1	1	1	2
62	1	1	62	27	150	90	16.6	1	1	3	0	0	3	2
63	1	0	76	29	140	90	10.6	1	2	1	0	1	3	1
64	0	0	40	17	110	70	18.5	1	2	1	0	1	1	2
65	1	1	60	25	130	80	12.5	1	1	1	0	0	3	2
66	1	1	61	30	130	80	4.8	3	2	3	0	0	3	2
67	0	0	42	22	130	100	12.5	1	3	3	0	0	3	2
68	1	1	51	23	140	90	11.2	1	2	2	0	0	2	2
69	1	1	48	30	150	100	8.2	3	3	2	0	0	1	2
70	1	0	71	25	140	70	22.4	1	1	1	0	0	3	2
71	1	1	61	28	170	80	9.1	3	2	3	0	0	2	2

72	0	1	44	32	130	80	14.8	2	3	3	0	0	3	2
73	0	1	62	19	114	66	15.5	1	2	4	0	0	2	2
74	1	0	71	27	140	80	4	1	1	3	0	0	3	2
75	0	1	42	26	120	80	14.7	3	2	3	0	0	1	1
76	0	0	54	23	110	70	10.1	1	4	2	0	0	2	2
77	1	1	67	27	130	80	3.8	3	2	1	0	0	2	2
78	1	1	56	33	140	70	11.9	3	2	4	0	1	3	2
79	1	1	65	27	190	90	5.6	1	1	3	0	0	2	2
80	1	0	66	26	140	80	8.8	1	4	5	0	0	1	2
81	1	1	60	26	150	80	13.9	1	1	1	0	0	2	2
82	1	0	57	27	150	80	8.9	1	4	2	0	0	2	2
83	1	1	63	26	140	70	10	1	2	3	0	0	3	2
84	1	1	51	30	140	90	12.5	1	2	3	0	1	3	1
85	0	1	51	24	120	80	8.3	1	2	3	0	0	2	2
86	1	0	57	24	130	70	5.1	2	2	2	0	0	2	2
87	1	1	55	27	140	70	16.9	1	2	3	0	1	3	2
88	0	1	56	18	120	80	11	2	1	4	0	0	3	3
89	0	1	45	23	100	70	22	1	2	3	0	1	3	2
90	1	1	52	32	140	100	13.3	1	2	3	0	1	3	2
91	1	1	47	29	190	100	11.1	1	2	4	0	0	3	2
92	0	1	58	32	130	80	15.6	1	4	2	0	0	2	2
93	0	1	40	24	110	70	11.6	1	2	3	0	0	2	2
94	0	0	48	35	130	80	12.2	2	4	2	0	0	2	2
95	0	1	51	19	110	70	27.5	1	2	3	0	0	3	2
96	1	1	50	33	140	90	10.6	1	4	2	0	0	2	2
97	1	1	50	26	120	90	13.6	3	2	4	0	0	3	3
98	0	1	56	26	120	70	11.2	1	2	3	0	0	2	2
99	1	1	70	28	120	80	13.1	1	1	5	0	0	3	1
100	0	1	54	22	120	80	6.1	1	1	3	0	0	3	1
101	0	1	58	26	110	70	7.5	1	2	3	0	0	3	2
102	0	0	90	21	130	80	4.5	3	1	4	0	0	2	2
103	1	1	48	26	150	100	10.4	3	2	4	0	0	2	2
104	1	1	67	23	140	100	17.6	1	1	1	0	0	3	2
105	0	1	44	23	120	90	18.5	1	1	3	0	0	2	2
106	0	1	43	28	120	90	5.3	1	2	3	0	0	2	3
107	1	0	55	24	140	90	16.7	1	2	2	0	0	2	2
108	1	1	63	35	140	90	8.1	1	2	4	0	0	2	3
109	0	1	56	34	110	70	8.7	2	2	1	0	0	3	3
110	0	1	55	21	130	90	5.8	1	2	3	0	0	3	2
111	0	1	45	16	110	70	9.7	1	2	3	0	0	3	2
112	0	0	50	20	130	95	13.2	2	4	2	0	1	1	2

113	1	1	48	28	140	90	14	1	2	3	0	1	2	3
114	1	1	62	25	130	70	7.4	1	1	1	0	0	3	2
115	1	1	60	27	140	70	15.9	1	1	3	0	0	3	2
116	0	1	53	31	120	70	13.3	1	2	3	0	1	3	2
117	1	1	52	27	170	100	4.9	1	1	3	0	0	3	3
118	0	1	46	30	130	70	18.4	2	2	2	0	0	3	3
119	1	1	58	27	160	100	13	2	4	2	0	0	2	2
120	0	1	49	19	130	70	18.2	3	1	3	0	0	3	2
121	1	0	78	28	135	100	7.9	2	3	5	0	1	3	2
122	0	0	56	24	120	70	21.9	1	2	3	0	0	3	3
123	0	0	43	26	110	70	11.7	1	3	3	0	0	2	2
124	1	1	76	25	150	100	6.5	1	2	4	0	0	3	1
125	1	0	74	27	140	90	5.2	1	2	1	0	0	3	1
126	1	1	58	24	140	70	8	3	1	4	0	0	2	3
127	0	1	68	26	120	70	13.2	1	1	3	0	0	3	2
128	1	1	57	26	180	70	10.8	2	2	3	0	0	3	2
129	1	1	45	33	195	99	10.2	3	2	3	0	1	2	3
130	1	1	63	34	150	80	7.2	1	2	3	0	0	3	2
131	1	0	70	28	130	80	14.7	1	3	5	0	0	2	2
132	1	1	82	28	150	80	4.5	1	1	4	0	0	3	1
133	0	0	78	21	120	70	12.8	2	4	5	0	0	2	2
134	1	1	59	30	170	90	12	1	2	3	0	0	2	3
135	0	1	63	18	110	70	5.5	2	2	4	0	0	3	3
136	1	1	42	28	130	90	17.7	1	2	3	0	0	2	2
137	0	0	40	20	120	70	21.3	1	1	3	0	0	2	3
138	0	1	54	30	110	70	11.9	3	1	3	0	0	3	3
139	1	0	63	25	140	90	10	3	4	5	0	0	2	2
140	0	0	46	21	120	70	17.8	1	4	2	0	1	2	2
141	1	1	46	24	140	90	10	1	4	2	0	0	2	2
142	1	0	64	24	130	70	16.7	1	3	5	0	1	2	2
143	1	1	64	24	140	90	13.3	2	4	2	0	0	2	2
144	0	1	52	22	110	80	12.2	1	2	3	0	0	3	3
145	0	1	41	24	120	70	5.6	3	2	4	0	0	3	3
146	0	0	44	28	120	90	8.3	1	2	3	0	1	2	2
147	1	1	62	19	120	70	7	1	2	3	0	0	3	2
148	1	0	47	24	170	120	9.5	1	2	3	0	1	2	3
149	0	1	58	24	112	78	16.3	1	2	3	0	1	3	2
150	1	0	78	19	150	70	7.2	1	4	5	0	0	3	2
151	0	0	42	16	100	60	20.1	1	2	3	0	0	2	2
152	0	0	42	23	120	80	21.2	1	3	2	0	0	1	3
153	1	1	70	25	120	80	5.3	2	2	1	0	0	3	1

154	1	1	50	23	140	80	6.8	1	1	1	0	0	3	3
155	1	0	49	27	120	80	6.3	1	3	2	0	0	2	2
156	1	0	60	23	140	90	16	2	2	2	0	0	2	2
157	0	1	74	14	110	60	25.1	1	1	3	0	0	3	2
158	1	0	69	30	180	100	17.6	1	1	3	0	0	3	2
159	0	1	42	21	110	70	18.3	1	2	3	0	0	3	2
160	1	1	76	26	140	90	5.6	1	2	3	0	0	3	1
161	1	1	66	35	140	90	7.6	2	2	4	0	1	3	2
162	1	1	46	24	150	90	8.4	1	4	2	0	0	3	2
163	0	1	44	27	120	70	16.4	3	2	4	0	0	2	2
164	1	1	59	29	140	90	9.2	1	4	2	0	1	3	2
155	1	1	66	27	140	100	5.4	1	1	3	0	0	3	3
156	1	0	73	22	140	80	7.4	3	1	4	1	0	3	3
157	1	0	52	25	140	80	4	1	2	1	0	1	3	2
158	1	1	78	27	145	80	4.7	2	1	4	0	0	3	2
159	1	1	73	31	150	90	12.4	1	3	3	0	0	3	3
170	1	1	70	22	120	80	12.7	2	1	1	0	0	3	3
171	1	1	67	20	180	90	11.1	2	1	1	0	0	3	2
172	1	0	59	31	170	90	12.8	1	4	2	0	0	2	2
173	1	0	44	29	140	80	5.2	1	4	2	0	1	1	2
174	1	1	57	23	140	70	5.8	3	2	3	0	1	3	1
175	0	0	43	24	110	70	18.5	1	2	3	1	0	3	2
176	0	0	46	22	120	80	17.2	1	4	2	0	1	2	3
177	0	0	44	22	120	80	17.7	1	2	4	0	1	2	2
178	1	0	73	26	140	80	8.2	1	4	5	0	1	3	1
179	0	0	47	21	110	70	21	1	3	3	0	1	2	2
180	1	0	56	26	160	90	7.9	1	2	3	0	1	2	2
181	0	1	55	29	120	70	11.8	1	1	1	0	1	3	2
182	0	1	42	26	120	80	21.7	1	2	3	0	1	2	2
183	1	1	54	29	170	100	16	2	2	3	0	0	3	2
184	0	1	60	24	130	90	12.5	1	1	3	0	1	3	2
185	1	1	43	35	140	90	15.4	1	2	3	0	0	2	2
186	1	0	48	21	180	100	12	1	2	2	0	0	3	2
817	1	1	62	26	150	100	10.2	1	1	1	0	0	3	2
188	1	1	55	27	160	80	6.9	1	2	1	0	0	3	2
189	0	1	51	19	130	70	16.7	3	2	4	0	0	2	2
190	1	1	70	27	140	80	5.5	1	1	3	0	1	2	2
191	1	1	74	23	130	70	7.8	2	1	1	0	1	3	2
192	1	1	53	25	150	80	20.4	1	4	2	0	1	3	2
193	1	1	52	28	140	80	6.9	1	4	2	0	1	3	1
194	0	0	44	21	120	70	12.5	1	2	3	0	0	2	2

195	1	1	58	28	150	80	8	1	1	4	0	0	3	2
196	0	0	44	21	110	70	5.8	1	4	2	0	0	3	2
197	0	1	56	24	120	70	21.9	1	1	1	0	1	3	3
198	1	1	61	24	140	80	5.2	3	1	3	0	1	3	2
199	1	1	49	28	150	90	16.3	1	2	3	0	0	3	2
200	1	1	68	26	120	70	13.2	1	2	3	0	0	3	3
201	0	0	49	21	120	70	7.4	1	2	3	0	0	2	2
202	0	1	46	26	100	70	13.3	1	2	1	0	0	2	2
203	1	0	50	27	140	90	5.3	2	4	2	0	1	2	2
204	1	1	72	27	140	80	7.2	2	1	1	0	1	3	2
205	1	1	76	17	140	90	17	1	3	1	0	0	3	1
206	1	1	46	32	180	100	5.1	1	2	3	0	0	3	2
207	0	0	55	18	120	70	7.2	2	1	3	0	1	2	2
208	0	0	45	30	120	80	12.6	2	1	3	0	1	3	2
209	1	1	59	25	140	70	15.1	1	4	2	0	1	2	2
210	0	0	46	21	120	80	9.5	1	2	3	0	1	2	2
211	1	1	55	30	130	80	5.5	1	1	3	0	0	3	2
212	1	1	56	30	140	90	10.2	1	2	3	0	1	3	2
213	1	1	60	23	150	100	10.5	1	2	4	0	0	3	2
214	1	0	50	26	140	80	10.6	1	3	3	0	0	2	2
215	0	0	54	18	130	70	27.4	1	2	3	0	1	3	3
216	1	1	67	30	150	70	9.7	1	1	3	0	1	3	2
217	1	0	47	30	200	100	6.9	1	4	2	0	0	2	3
218	1	1	84	23	130	70	13.2	3	1	1	0	0	3	2
219	0	0	45	21	110	70	8.5	1	2	3	0	0	2	2
220	0	1	41	27	130	100	16.4	1	2	3	0	0	2	2
221	1	1	60	26	150	80	12.4	1	2	3	0	0	3	3
222	0	1	52	23	120	80	7.7	1	2	3	0	0	2	2
223	1	0	64	28	170	100	10.3	2	2	3	1	0	2	2
224	1	1	50	33	150	110	7.3	1	2	3	0	1	3	3
225	1	0	58	25	140	90	13.4	1	3	2	0	1	3	2
226	1	0	60	22	140	90	5.3	3	3	1	0	0	3	2
227	1	1	45	28	140	100	11.7	1	3	3	0	1	2	2
228	1	0	59	23	140	70	8.8	1	1	3	1	0	3	3
229	1	1	48	20	160	100	11	1	2	3	0	0	2	2
230	1	1	70	18	120	80	8.9	1	1	5	0	1	3	2
231	0	1	50	20	110	60	9.1	1	2	5	0	0	2	2
232	1	0	78	16	150	70	4.3	1	1	4	0	0	3	1
233	1	1	57	26	140	70	19.9	1	2	3	0	1	3	2
234	0	1	52	19	100	60	20.8	1	2	2	0	0	2	2
235	1	1	67	38	180	90	7.9	2	2	1	0	0	3	3

236	0	1	45	19	120	70	21.8	1	2	3	0	1	3	3
237	0	1	56	27	120	80	6.8	1	1	3	0	0	3	2
238	1	1	80	18	110	70	5.6	3	1	1	0	0	3	3
239	1	1	60	16	170	100	9.3	1	2	4	0	0	2	2
240	1	1	56	30	210	100	8.8	2	1	1	0	0	3	3
241	1	1	74	24	130	70	5.1	1	2	1	0	1	3	1
242	1	1	71	21	120	70	5.5	1	2	4	0	0	3	2
243	0	1	50	23	100	70	19	2	2	3	0	1	2	2
244	0	1	45	15	130	70	5	1	1	1	0	0	2	2
245	0	1	80	16	130	70	5	1	1	1	0	0	2	2
246	1	1	42	30	140	100	8	1	4	2	0	1	2	2
247	0	1	51	20	120	70	16.2	3	2	4	0	1	2	2
248	1	1	60	28	170	110	4.8	1	1	3	0	0	3	2
249	0	0	48	26	135	85	5.7	3	3	2	0	1	1	3
250	0	0	49	20	110	70	20	1	2	3	0	1	2	2
251	0	0	45	25	120	70	8.3	2	4	2	0	1	2	2
252	0	0	55	25	130	70	10.3	1	2	3	0	0	2	2
253	0	1	70	20	120	70	7.6	2	1	4	0	0	3	1
254	0	0	52	20	140	90	9	2	2	3	1	0	3	3
255	1	1	53	32	140	90	8.8	1	2	1	0	1	3	2
256	1	1	50	19	140	80	21	1	2	3	0	1	3	2
257	1	1	45	20	140	90	15.1	1	2	4	0	0	3	2
258	1	1	60	25	130	70	27	1	2	3	0	1	2	2
259	1	1	55	20	150	70	15.7	1	2	3	0	0	3	2
260	1	1	57	27	150	70	9.8	1	4	5	0	1	2	2
261	1	1	70	24	170	90	12.9	1	1	1	0	0	3	2
262	1	1	65	25	150	90	5.6	1	4	5	0	0	3	2
263	1	1	57	26	170	90	8.9	1	2	3	0	0	3	2
264	0	0	62	20	110	70	8.2	1	2	3	0	1	3	2
265	1	1	72	26	140	70	5.9	1	1	1	0	0	3	2
266	1	0	58	26	190	100	4.1	1	3	2	0	1	3	2
267	1	0	59	24	140	90	19.8	1	2	3	0	1	2	2
268	1	1	77	21	150	80	13	1	1	1	0	0	3	1
269	1	1	55	26	170	110	5.6	1	1	1	0	0	2	2
270	1	1	70	30	150	90	14.1	1	1	1	0	1	2	1
271	0	1	40	18	110	70	12	1	2	3	0	1	1	2
272	0	1	53	32	110	70	13.8	1	2	3	0	1	2	2
273	1	0	55	24	170	100	6.1	1	2	3	0	1	1	2
274	1	1	62	31	150	70	7	1	2	1	0	0	2	1
275	0	1	44	26	110	80	12	1	2	1	0	0	2	2
276	0	1	59	24	120	70	10.1	1	3	1	0	1	2	2

277	1	1	71	24	150	90	3.6	1	2	1	0	1	2	2
278	0	1	47	26	100	60	13.7	2	2	3	0	1	1	2
279	0	0	73	22	130	80	5.4	1	2	3	1	1	2	2
280	0	1	55	30	110	70	6.2	1	1	3	0	0	2	2
281	1	0	53	24	160	80	4	1	2	2	1	1	1	2
282	0	1	40	21	110	80	11.4	1	2	1	0	0	2	2
283	1	1	66	23	160	100	6.1	1	1	1	0	0	2	2
284	1	1	70	20	120	80	8	1	1	1	0	1	3	2
285	1	1	85	23	158	73	7.7	1	1	3	0	0	3	2
286	1	1	48	25	140	90	9.8	1	2	3	0	0	2	2
287	1	0	63	26	140	70	6.9	3	2	2	1	1	2	2
288	1	1	72	29	160	80	8.6	2	1	4	0	0	3	2
289	1	0	59	24	160	90	10.6	1	2	3	0	1	2	2
290	0	1	63	18	100	70	11	3	3	4	1	1	2	2
291	1	1	64	23	170	100	19	3	2	3	0	1	2	1
292	0	1	70	19	120	80	8	1	1	3	0	1	2	2
293	1	1	63	22	160	100	13.6	2	4	3	0	1	1	2
294	1	1	50	27	140	80	13.6	1	2	3	0	1	1	2
295	0	1	41	26	110	90	4	1	2	2	0	0	2	3
296	0	0	47	22	130	80	7	1	4	2	0	1	2	2
297	0	1	58	33	120	70	13.7	1	1	3	0	1	2	2
298	1	0	54	26	160	70	7.2	1	2	4	0	0	2	3
299	1	1	51	34	130	70	20	1	2	3	0	0	2	2
300	1	1	40	27	150	80	8	1	3	3	0	0	2	2
301	1	1	79	28	160	80	7.8	1	1	1	0	0	3	2
302	1	0	60	25	130	80	6.1	1	2	3	0	0	2	2
303	1	1	87	25	140	90	10.2	1	1	3	0	0	2	2
304	1	1	64	26	150	90	7.2	1	2	5	0	0	2	2
305	1	1	58	26	160	80	7.5	1	2	3	0	0	3	2
306	0	1	53	24	130	70	4.8	1	2	3	0	0	2	2
307	0	0	54	26	110	70	10.4	1	2	2	0	0	2	2
308	1	1	71	27	140	60	7.2	2	1	1	0	0	3	2
309	1	0	41	32	140	190	3.7	1	2	3	0	1	3	2
310	1	0	57	26	140	90	5.8	1	2	5	0	0	2	2

APPENDIX III

DESCRIPTIVE STATISTICS

Descriptive Statistics of Diabetes with hypertension

	N	Minimum	Maximum	Mean	Std. Deviation
height	198	1.34	1.87	1.6182	.07751
weight	198	42.00	103.00	70.2407	11.94415
SBP	198	110.00	210.00	147.5152	16.85131
DBP	198	60.00	190.00	85.1616	14.11246
BGL	198	2.60	61.00	10.1237	5.82761
AGE	198	40.00	87.00	60.7980	10.21493
BMI	198	16.00	38.00	26.5551	3.79097
Valid N (listwise)	198				

Descriptive Statistics of Diabetes without hypertension

	N	Minimum	Maximum	Mean	Std. Deviation
height	112	1.40	1.87	1.6371	.09179
weight	112	6.68	97.70	62.2753	13.06857
SBP	112	100.00	180.00	118.3036	10.94774
DBP	112	60.00	100.00	74.2857	8.01398
BGL	112	4.00	32.50	12.8330	5.92803
AGE	112	40.00	90.00	51.4196	9.41404
BMI	112	14.00	35.00	23.2687	4.23707
Valid N (listwise)	112				

APPENDIX IV

LOGISTIC OUTPUT

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	310	100.0
	Missing Cases	0	.0
	Total	310	100.0
	Unselected Cases	0	.0
	Total	310	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
Diabetes without Hypertension	0
Diabetes with Hypertension	1

Block 0: Beginning Block

Classification Table[a,b]

Observed			Predicted		
			HTD		
			Diabetes without Hypertension	Diabetes with Hypertension	Percentage Correct
Step 0	HTD	Diabetes without Hypertension	0	112	.0
		Diabetes with Hypertension	0	198	100.0

Overall Percentage			63.9

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	.570	.118	23.223	1	.000	1.768

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Age	53.168	1	.000
		BMI	42.794	1	.000
		Occupation	.323	1	.570
		BGL	23.111	1	.000
		Exercise	16.713	1	.000
		Drinking	1.631	1	.202
		Gender	4.035	1	.045
		Residence	.006	1	.940
		Education	.379	1	.538
		Overall Statistics	102.953	9	.000

Block 1: Method = Backward Stepwise (Wald)

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	120.878	9	.000
	Block	120.878	9	.000
	Model	120.878	9	.000
Step 2[a]	Step	-.005	1	.945
	Block	120.873	8	.000
	Model	120.873	8	.000
Step 3[a]	Step	-.120	1	.729
	Block	120.752	7	.000
	Model	120.752	7	.000
Step 4[a]	Step	-.203	1	.653
	Block	120.550	6	.000
	Model	120.550	6	.000
Step 5[a]	Step	-.463	1	.496
	Block	120.087	5	.000
	Model	120.087	5	.000
Step 6[a]	Step	-1.859	1	.173
	Block	118.228	4	.000
	Model	118.228	4	.000

Step 7[a]	Step	-1.719	1	.190
	Block	116.509	3	.000
	Model	116.509	3	.000

a. A negative Chi-squares value indicates that the Chi-squares value has decreased from the previous step.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	284.700[a]	.323	.442
2	284.704[a]	.323	.442
3	284.825[a]	.323	.442
4	285.028[a]	.322	.442
5	285.490[a]	.321	.440
6	287.349[a]	.317	.435
7	289.068[a]	.313	.429

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	8.879	8	.353
2	8.841	8	.356
3	4.360	8	.823
4	6.388	8	.604

5	11.179	8	.192
6	7.844	8	.449
7	10.947	8	.205

Contingency Table for Hosmer and Lemeshow Test

		HTD = Diabetes without Hypertension		HTD = Diabetes with Hypertension		
		Observed	Expected	Observed	Expected	Total
Step 1	1	29	27.679	2	3.321	31
	2	24	23.197	7	7.803	31
	3	16	18.312	15	12.688	31
	4	15	13.639	16	17.361	31
	5	5	10.035	26	20.965	31
	6	8	7.410	23	23.590	31
	7	7	5.324	24	25.676	31
	8	5	3.556	26	27.444	31
	9	1	1.997	30	29.003	31
	10	2	.851	29	30.149	31
Step 2	1	29	27.680	2	3.320	31
	2	24	23.204	7	7.796	31
	3	16	18.306	15	12.694	31

	4	15	13.646	16	17.354	31
	5	5	10.031	26	20.969	31
	6	8	7.393	23	23.607	31
	7	7	5.324	24	25.676	31
	8	5	3.567	26	27.433	31
	9	1	1.996	30	29.004	31
	10	2	.854	29	30.146	31
Step 3	1	29	27.672	2	3.328	31
	2	23	23.248	8	7.752	31
	3	17	18.263	14	12.737	31
	4	13	13.650	18	17.350	31
	5	7	9.920	24	21.080	31
	6	7	7.417	24	23.583	31
	7	8	5.400	23	25.600	31
	8	5	3.566	26	27.434	31
	9	2	2.006	29	28.994	31
	10	1	.857	30	30.143	31
Step 4	1	29	27.680	2	3.320	31
	2	23	23.198	8	7.802	31
	3	17	18.263	14	12.737	31
	4	13	13.601	18	17.399	31
	5	7	9.982	24	21.018	31

	6	7	7.435	24	23.565	31
	7	8	5.370	23	25.630	31
	8	5	3.622	26	27.378	31
	9	1	1.980	30	29.020	31
	10	2	.869	29	30.131	31
Step 5	1	28	27.648	3	3.352	31
	2	24	23.207	7	7.793	31
	3	17	18.250	14	12.750	31
	4	15	13.483	16	17.517	31
	5	6	10.055	25	20.945	31
	6	6	7.442	25	23.558	31
	7	7	5.440	24	25.560	31
	8	7	3.620	24	27.380	31
	9	0	1.981	31	29.019	31
	10	2	.873	29	30.127	31
Step 6	1	29	27.553	2	3.447	31
	2	23	22.974	8	8.026	31
	3	18	18.192	13	12.808	31
	4	14	13.653	17	17.347	31
	5	6	10.137	25	20.863	31
	6	5	7.502	26	23.498	31
	7	9	5.334	22	25.666	31

	8	5	3.761	26	27.239	31
	9	2	2.032	29	28.968	31
	10	1	.864	30	30.136	31
Step 7	1	29	27.571	2	3.429	31
	2	22	22.942	9	8.058	31
	3	21	18.009	10	12.991	31
	4	11	13.240	20	17.760	31
	5	10	10.257	21	20.743	31
	6	2	7.650	29	23.350	31
	7	8	5.615	23	25.385	31
	8	6	3.783	25	27.217	31
	9	2	2.043	29	28.957	31
	10	1	.890	30	30.110	31

Classification Table[a]

			Predicted		
			HTD		
	Observed		Diabetes without Hypertension	Diabetes with Hypertension	Percentage Correct
Step 1	HTD	Diabetes without Hypertension	71	41	63.4
		Diabetes with Hypertension	26	172	86.9
		Overall Percentage			78.4

Step 2	HTD	Diabetes without Hypertension	71	41	63.4
		Diabetes with Hypertension	26	172	86.9
		Overall Percentage			78.4
Step 3	HTD	Diabetes without Hypertension	71	41	63.4
		Diabetes with Hypertension	26	172	86.9
		Overall Percentage			78.4
Step 4	HTD	Diabetes without Hypertension	70	42	62.5
		Diabetes with Hypertension	26	172	86.9
		Overall Percentage			78.1
Step 5	HTD	Diabetes without Hypertension	71	41	63.4
		Diabetes with Hypertension	25	173	87.4
		Overall Percentage			78.7
Step 6	HTD	Diabetes without Hypertension	70	42	62.5
		Diabetes with Hypertension	26	172	86.9
		Overall Percentage			78.1
Step 7	HTD	Diabetes without Hypertension	72	40	64.3
		Diabetes with Hypertension	21	177	89.4
		Overall Percentage			80.3

a. The cut value is .500

Variables in the Equation

							95% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1[a]	Age	.106	.018	36.491	1	.000	1.112	1.074	1.150
	BMI	.226	.040	31.507	1	.000	1.254	1.159	1.357
	Occupation	.010	.147	.005	1	.945	1.010	.757	1.349
	BGL	-.067	.029	5.263	1	.022	.935	.883	.990
	Exercise	.412	.274	2.269	1	.132	1.510	.883	2.583
	Drinking	-.109	.315	.120	1	.729	.897	.484	1.662
	Gender	.216	.343	.399	1	.528	1.242	.634	2.431
	Residence	-.099	.213	.214	1	.644	.906	.597	1.376
	Education	.264	.183	2.098	1	.147	1.303	.911	1.863
	Constant	-11.821	1.865	40.172	1	.000	.001		
Step 2[a]	Age	.106	.017	36.604	1	.000	1.111	1.074	1.150
	BMI	.226	.040	31.585	1	.000	1.254	1.159	1.356
	BGL	-.067	.029	5.261	1	.022	.935	.884	.990
	Exercise	.413	.274	2.276	1	.131	1.511	.884	2.583
	Drinking	-.109	.315	.121	1	.728	.896	.484	1.661
	Gender	.215	.342	.395	1	.529	1.240	.634	2.426

	Residence	-.096	.210	.209	1	.647	.908	.602	1.371
	Education	.264	.182	2.101	1	.147	1.303	.911	1.863
	Constant	-11.790	1.808	42.532	1	.000	.001		
Step 3[a]	Age	.105	.017	36.618	1	.000	1.111	1.074	1.150
	BMI	.225	.040	31.492	1	.000	1.253	1.158	1.355
	BGL	-.068	.029	5.649	1	.017	.934	.883	.988
	Exercise	.422	.273	2.400	1	.121	1.525	.894	2.602
	Gender	.234	.338	.480	1	.489	1.264	.652	2.450
	Residence	-.095	.210	.203	1	.652	.910	.603	1.373
	Education	.259	.181	2.037	1	.154	1.295	.908	1.848
	Constant	-11.805	1.806	42.702	1	.000	.000		
Step 4[a]	Age	.105	.017	36.483	1	.000	1.111	1.074	1.150
	BMI	.225	.040	31.513	1	.000	1.253	1.158	1.356
	BGL	-.068	.029	5.553	1	.018	.935	.883	.989
	Exercise	.425	.272	2.441	1	.118	1.530	.898	2.607
	Gender	.230	.338	.463	1	.496	1.259	.649	2.442
	Education	.264	.181	2.124	1	.145	1.302	.913	1.856
	Constant	-11.956	1.778	45.210	1	.000	.000		
Step 5[a]	Age	.105	.017	36.234	1	.000	1.111	1.073	1.149
	BMI	.230	.040	33.405	1	.000	1.258	1.164	1.360
	BGL	-.065	.028	5.297	1	.021	.937	.886	.990
	Exercise	.451	.269	2.819	1	.093	1.570	.927	2.660

	Education	.236	.175	1.811	1	.178	1.266	.898	1.785
	Constant	-11.925	1.779	44.952	1	.000	.000		
Step 6[a]	Age	.101	.017	35.230	1	.000	1.107	1.070	1.144
	BMI	.233	.040	34.677	1	.000	1.263	1.168	1.365
	BGL	-.066	.028	5.458	1	.019	.936	.886	.989
	Exercise	.330	.252	1.718	1	.190	1.391	.849	2.277
	Constant	-10.999	1.608	46.811	1	.000	.000		
Step 7[a]	Age	.107	.017	41.646	1	.000	1.113	1.077	1.150
	BMI	.236	.039	35.590	1	.000	1.266	1.171	1.367
	BGL	-.065	.028	5.443	1	.020	.937	.887	.990
	Constant	-10.581	1.567	45.620	1	.000	.001		

a. Variable(s) entered on step 1: Age, BMI, Occupation, BGL, Exercise, Drinking, Gender, Residence, Education.

Variables not in the Equation

			Score	df	Sig.
Step 2[a]	Variables	Occupation	.005	1	.945
		Overall Statistics	.005	1	.945
Step 3[b]	Variables	Occupation	.005	1	.941
		Drinking	.121	1	.728
		Overall Statistics	.125	2	.939
Step 4[c]	Variables	Occupation	.000	1	.995
		Drinking	.115	1	.735

		Residence	.203	1	.652
		Overall Statistics	.329	3	.954
Step 5[d]	Variables	Occupation	.001	1	.971
		Drinking	.196	1	.658
		Gender	.464	1	.496
		Residence	.187	1	.665
		Overall Statistics	.795	4	.939
Step 6[e]	Variables	Occupation	.000	1	.983
		Drinking	.072	1	.789
		Gender	.128	1	.721
		Residence	.280	1	.597
		Education	1.824	1	.177
		Overall Statistics	2.595	5	.762
Step 7[f]	Variables	Occupation	.000	1	1.000
		Exercise	1.730	1	.188
		Drinking	.224	1	.636
		Gender	.417	1	.518
		Residence	.286	1	.593
		Education	.728	1	.394
		Overall Statistics	4.336	6	.631

a. Variable(s) removed on step 2: Occupation.

b. Variable(s) removed on step 3: Drinking.

c. Variable(s) removed on step 4: Residence.

d. Variable(s) removed on step 5: Gender.

e. Variable(s) removed on step 6: Education.

f. Variable(s) removed on step 7: Exercise.

APPENDIX V

PROBIT OUTPUT

Data Information

		N of Cases
	Valid	310
Rejected	Missing	0
	Number of Responses > Number of Subjects	0
	Control Group	301

Convergence Information

	Number of Iterations	Optimal Solution Found
PROBIT	22	Yes

Parameter Estimates

	Parameter	Estimate	Std. Error	Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
PROBIT[a]	Age	.026	.005	4.816	.000	.016	.037
	BMI	.058	.013	4.344	.000	.032	.085
	Occupation	.001	.049	.021	.983	-.094	.096
	BGL	-.024	.011	-2.085	.037	-.046	-.001
	Drinking	-.009	.120	-.071	.943	-.243	.226

	Education	.077	.064	1.194	.233	-.049	.203
	Residence	-.020	.079	-.259	.796	-.176	.135
	Gender	.081	.132	.612	.541	-.178	.339
	Exercise	.111	.105	1.058	.290	-.095	.317
	Intercept	-3.758	.608	-6.184	.000	-4.366	-3.151

a. PROBIT model: PROBIT(p) = Intercept + BX

Covariances and Correlations of Parameter Estimates

		Age	BMI	Occupation	BGL	Drinking	Education
PROBIT	Age	.000	.098	.008	.185	.035	.180
	BMI	.000	.000	.009	.139	-.052	-.119
	Occupation	.000	.000	.002	-.029	-.021	-.092
	BGL	.000	.000	.000	.000	-.105	-.010
	Drinking	.000	.000	.000	.000	.014	-.100
	Education	.000	.000	.000	.000	.000	.004
	Residence	.000	.000	.000	.000	.000	.000
	Gender	.000	.000	.001	.000	.002	.002
	Exercise	.000	.000	.000	.000	.001	.002

Covariances (below) and Correlations (above).

Covariances and Correlations of Parameter Estimates

	Residence	Gender	Exercise

PROBIT	Age	-.023	.051	-.224
	BMI	.010	-.191	-.077
	Occupation	-.101	.088	-.047
	BGL	.075	-.124	.020
	Drinking	.040	.141	.054
	Education	.069	.218	.239
	Residence	.006	.006	.036
	Gender	.000	.017	-.175
	Exercise	.000	-.002	.011

Covariances (below) and Correlations (above).

Chi-Square Tests

		Chi-Square	df[a]	Sig.
PROBIT	Pearson Goodness-of-Fit Test	137.838	300	1.000

a. Statistics based on individual cases differ from statistics based on aggregated cases.