Modeling Diabetes Risk Factors (A Case Study of Focus Medical Centre in Kiambu, Kenya 2016)

Thomas Mageto, Efron Njuguna, Dolleen Osundwa

Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Email address:
tttmageto@gmail.com (T. Mageto)

Abstract: This study sought to model risk factors of diabetes (A case study of Focus Medical Center in Kiambu, Kenya) for the year 2016. We considered sample of size 181 patients and carried descriptive statistics, bivariate analysis, Chi-Square test and Hosmer and Lemeshow test. The independence test between response variable (diabetes) and predictor variables (age, obesity, alcohol, smoking and hypertension) was carried. The variables age, obesity, alcohol and hypertension were found to be statistically significant at $\alpha = 0.05$ level of significant. A multiple logistic regression model was fitted and the fitted regression model indicated that the predictor variables age, obesity and alcohol were statistically significant. The results of the odds ratios show that age, obesity and alcohol consumption are positively associated with diabetes. The fitted reduced multiple logistic regression model was subjected to an overall goodness-of-fit test and results indicate that there is no significant difference between the observed and predicted probability. Based on the results of this study, we recommend that special attention should be given to individuals advanced in age, consume alcohol or who are obese for screening as there is a high possibility of testing positive for diabetes for health care givers to monitor and manage the condition. Further, healthy lifestyles should be promoted among the general population and in particular, the diabetic patients to increase the chance of properly managing the condition. A further study ought to be conducted to assess treatment interventions of diabetes to ascertain the effectiveness and recommend the best medication for patients suffering from diabetes.

Keywords: Diabetes, Multiple Logistic Regression, Risk Factors, Odd Ratio, Sample Size, Hosmer and Lemeshow Test

1. Introduction

1.1. Background of Study

Diabetes is a chronic disease that occurs either when the body cannot use the insulin it produces effectively or when the insulin production from the pancreas is not enough. It is characterized by the elevation of blood sugar which overtime lead to serious damage to the nerves, heart, eyes, kidneys and the blood vessels. In the past, people had a belief that diabetes was a disease for the rich but according to the first World Health Organization global report on the disease, the disease which was once a predominant problem in the rich countries has become one that excessively affects the poorer countries [1]. Over the past decade, the prevalence of diabetes in the low and middle income countries has risen faster than in the high income countries. The fast increase in obesity and diabetes cases in the low and middle income countries can be explained by the very rapid urbanization which has led to change in lifestyle in these countries.

There are two main types of diabetes; type 1 and type 2 diabetes, type 1 diabetes formerly known as insulin-dependent is where the insulin producing cells are attacked and destroyed by the immune system of the body while type 2 diabetes formerly known as non-insulin dependent, that is, the cells of the body do not react to insulin or insulin production by the pancreas is not enough. In the world, type 2 diabetes is far more common than type 1 diabetes accounting for around 90% of all diabetic cases in the world. According to International Diabetes Federation (IDF), some of the risk factors of diabetes include; family history of diabetes, age, high blood pressure, physical inactivity, being overweight, impaired glucose tolerance and unhealthy diets. Diabetes (type 1 and type 2) have no cure but can be prevented if diagnosed at its early stages. If the disease is not diagnosed on its onset, one may develop very serious

complications such as stroke, heart problems, kidney disease, cardiovascular diseases and eye problems. Diabetes management is mainly pharmacological, that is, administration of insulin and oral hypoglycemic agents and non pharmacological, that is reducing lifestyle risk factors. Nowadays, diabetes is managed as a cardiovascular risk factor and the emphasis is on the control of the present risk factors be it biological or behavioral. The lifestyle (behavioral) risk factors are important due to a number of reasons. First, they play important role in development and management of biological factors such as high blood pressure and obesity. Secondly, lifestyle risk factors give a cost effective way of preventing and controlling diabetes and other non-communicable diseases.

Diabetes is a huge burden in the world, for example, the total estimated cost of diagnosed diabetes in United States of America in 2012 was 245 billion which included 176 billion in direct medical costs and 69 billion in reduced productivity [2]. The expenditures for people with diabetes were 2.3 times higher than what their expenditures would be if they didn't have diabetes, it is with this in mind that we wish to study on causes and management of the disease.

1.2. Problem Statement

The goal of treatment of diabetes is to keep the levels of glucose as near normal as possible while also avoiding chronic and acute complications. Diabetes management does not only depend on drug therapy but also on diet, physical exercise and other factors. The main channels of achieving the goals of diabetes treatment are education to patients and full participation in control of diabetes. Failure of patients adhering to the recommended regimens leads to complications of diabetes and their associated societal, economic and individual costs. According to an International Diabetes Federation report, 450 million people in the world in 2015 had diabetes and predict that by 2040, an estimated 642 million people will have diabetes, that is, 54.7 % increase [3]. In 2014, the number of adults (20-79) years living with diabetes was 422 million from 108 million in 1980. It is evident that the number people with diabetes are increasing every year, from 2014 to 2015, there was 6.6% increase in diabetic cases. Most of these people live in the developing countries. Worldwide, diabetes causes 3.8 million deaths per year which is more than AIDS and malaria combined. In 2012, 80% of deaths caused by diabetes took place in the low and middle income countries. According to WHO, the dramatic increase in diabetic cases is as a result of obesity and being overweight [1]. Due to the increase in diabetic cases, people have been urged to for early screening for early detection of the disease so that its management can start when the disease is in its early stages.

Due to rapid growth of urbanization and increase in sedentary lifestyles in Kenya which are known to predispose people to diabetes, the number of people with diabetes has being increasing annually. The increase in diabetic cases in Kenya can also be explained by the low diabetes awareness and limited access to quality healthcare. In 2015, there were 478 thousand cases. The prevalence of diabetes in adults 20-79 years was 2.2% and the number of deaths due to diabetes was 8722. Due to this, the Kenyan Government in its National Medium Term Plan (2014 – 2018) and National Health Strategic Plan have identified preventing and controlling of non-communicable diseases as a priority. The aim of the plans is to halt the rising trend of non-communicable diseases. This study helped determine the significance of the risk factors to diabetes by using data of patients attending Focus Medical Centre in 2016. The factors considered in the study are age, obesity, hypertension, smoking status, alcohol consumption and gender.

1.3. Objectives

General objective
To model the risk factors associated with diabetes (A case study of Focus Medical Centre in Kiambu, 2016)
Specific objectives
(i) To determine the risk factors that are associated with diabetes
(ii) To fit the multivariate logistic regression model
(iii) To determine the predictive strength of fitted multivariate logistic regression model.

1.4. Hypothesis

To achieve the objectives of this study, the following hypothesis was tested;
H_0: There is no relationship between having diabetes and the risk factors
H_a: There is a relationship between having diabetes and the risk factors

1.5. Justification

In Kenya, there are fewer studies being done on non-communicable diseases compared to communicable diseases. However some studies have been done in Kiambu but none has focused on modeling the risk factors of diabetes. This study was therefore necessary. In this study, secondary data on diabetes which is readily available at Focus Medical Centre was used. The results of this study are important because of the following reasons;
(i) They would give information as to whether the factors; age, obesity, hypertension, smoking, alcohol consumption and gender are really the risk factors of being diagnosed with diabetes
(ii) The findings would help people to know which factors contribute more to diabetes. With this information, people without diabetes would know the estimated risk and take preventive measures to prevent getting diabetes.
(iii) The results of this study would benefit the people managing diabetic patients as they design education materials and strategies to better achieve their objectives. It would help them know which sub groups of their patients to give specialized attention.

2. Literature Review

Management of diabetes is affected by alcohol in different ways, that is, chronic ingestion of alcohol (45g/day) leads to deterioration in short and long term metabolism of glucose in persons with type 2 diabetes [4]. The effects caused by excessive consumption of alcohol can be reversed if the patient abstains from alcohol for three days [5]. In adults who have diabetes, there is an association between consumption of alcohol (5-15g/day) and reducing risk of coronary heart disease, this is perhaps due to the associated increase in High Density Lipoprotein (HDL) cholesterol. In Kenya, this translates to not more than one 330mls Tusker drink which has the lowest alcohol content. A study [6] type 2 diabetic patients from ethnic minority groups found that consumption of alcohol within a period of 30 days was related to poor oral medications, appointment keeping, adherence to diet and self monitoring of blood glucose. Another study which examined use of alcohol in 154 older men who were diabetic, found out that greater use of alcohol was related to poor adherence to insulin injections [7].

The relationship between blood pressure and alcohol consumption appears to be J or U-shaped and light to moderate alcohol amounts do not lead to rise in blood pressure. However, there is a strong relationship between excessive alcohol consumption (30-60g/day) and elevation of blood pressure in both men and women. An increment by 10g of alcohol consumption above 30 g/day leads to an increase to diastolic blood pressure by 1mmHg and systolic blood pressure by 1-2 mmHg. Alcohol is a risk factor of hypertension; it also interferes with anti hypersensitive therapy and may also be a risk factor of stroke. The Dietary Guidelines for Americans has recommended that adult men should have two drinks per day and adult women should have one drink per day. The recommended units are a maximum of 14 units for adult women per week and 21 units for adult men per week. Locally, this is approximately one Tusker drink per day for adult men [3]. It has been observed that the risk of getting diabetes is higher for people who consume alcohol compared to people who do not consume alcohol. The findings were that subjects who consume alcohol are 2.282 times more likely to get diabetes compared to subjects who do not consume alcohol.

In epidemiological studies on the incidence of diabetes the studies have included smoking as a possible factor which increases the risk of getting diabetes but most studies have only included it as a variable among many more which may have an association with diabetes. A past study found out that in both males and females, there was an increase in the rate of diabetes as smoking increased. In group of men and women who smoked at least 2 packs every day, the rates were 45% and 74% higher than in men and women who had never smoked [8]. Smoking cessation is encouraged to reduce the risk of diabetes but there is little information on how effective cessation may be, however, it is necessary that frequency of advice and counseling on cessation be increased for diabetic patients because of the strong evidence of risk of diabetes in smokers as well as increased mortality and complications.

Hypertension is a condition in which there is persistently raised pressure in the blood vessels. In many low and middle income countries, hypertension has been referred to as a silent killer due to its high prevalence and being under-diagnosed frequently. Hypertension is the most common cardiovascular problem in Sub-Sahara Africa. If untreated, it can cause serious complications such as stroke, heart disease and diabetes. Many people with blood pressure also have diabetes, in 2014, 22% of adults had high blood pressure [9]. According to studies published between 1980 and 2003, prevalence of hypertension in India was 3.4% in men and 6.8% in women, in Poland, 68.9% in men and 72.5% in women. In Kenya, prevalence of hypertension is growing due to many risk factors such as poor diets, tobacco use, stress, obesity and decreased physical activity, prevalence in men and women ranges from 12.3-22.8% in urban areas and 7.4-21.4% in rural areas.

There appear to be a correlation between the increasing incidence of hypertension with the aging of the population and the rising number of people who are overweight and obese [9]. High blood pressure, that is, more than 140/90 mmHg has been associated with increased risk of type 2 diabetes. Trends show that there are regional variations which are associated with change of lifestyle and rate of urbanization. Change of lifestyle and rate of urbanization have led to emergence of cardiovascular diseases and diabetes and a decline in infectious diseases. Recent studies have shown that differences in prevalence of hypertension are not only among countries but also between ethnic groups. For instance, the prevalence among Blacks in US is higher than in Mexican-Americans and Whites. The association between diabetes and hypertension has also been shown in a study which found out that among all patients with diabetes, over 70% had hypertension. The frequency of coexistence of diabetes and hypertension is high. When diabetes and hypertension coexist, the rate of occurrence of complications is much higher [10].

Cholesterol is a fat substance that is present in the cells and the blood naturally. There are two types of cholesterol: High Density Lipoprotein (HDL) and Low Density Lipoprotein (LDL). According to WHO, diet high in cholesterol, calories and fat increases a person's risk of diabetes. A poor diet also leads to obesity and other health problems. If a person has low levels of HDL, there is higher risk of type 2 diabetes. According to [1] triglycerides are a type of fat in the blood and people with high levels are at a high risk of type 2 diabetes. Healthy eating and physical activity can help in improving a person's cholesterol levels. For diabetic patients, the main target should be to lower LDL cholesterol to 2.0 mmol/l or lower. High cholesterol levels leads to narrowing and blocking of arteries, diabetics are more prone to having high cholesterol levels which contribute to cardiovascular diseases. Increase in levels of cholesterol is an emerging burden of the risk factors of blood pressure and non communicable diseases and is now a major

problem in the public health sector.

A study estimating the prevalence of diabetes in a population associated with occupation is likely to suggest a type of work with more mental work than physical work. The people who are most affected are software engineers, doctors and businessmen. Studies have also shown that incidence of diabetes has affected women who belong to the upper stratum and house wives. However, associating risk of diabetes and occupation is difficult since it will depend more on the individual's attitude, physical exercise and food intake. A study on the incidence of type 2 diabetes among occupational classes in Sweden (35 year follow up cohort study) showed that low occupational class was significant for being diagnosed with diabetes. The risk remained higher in the low occupational class even after adjusting for well known diabetic risk factors. There was greater incidence among skilled and non-manual employees but this was not significant. The findings of this study were consistent with other studies. A study in Japan aimed to determine the relationship that exists between occupation and type 2 diabetes. The subjects in the study were classified into five categories namely; managers, laborers, clerical, technical and transport workers. After adjusting for age, family history and BMI, the relative risk of transport workers compared to laborers was 3.95 which suggested that development of type 2 diabetes is associated with occupation [12]. A study in Iran found no significant association but there were significant differences in the prevalence of diabetes between the retired people and people in other occupational categories. The difference may be accounted for by the fact that in Iran, the pension for retired people is low [13].

As one gets older, the risk for type 2 diabetes, stroke and heart diseases goes up. This is especially after 45 years. This is because one tends to exercise less, gain weight and lose muscle mass. Recent studies on type 2 diabetes have shown that there is a dramatic increase of the disease among adolescents, children and younger adults. Studies have also shown that white people aged 40 and above are at an increased risk of getting diabetes. People of Black African, African Caribbean, Chinese and South Asian descent are at a higher risk of having type 2 diabetes at an earlier age. In the U.S, type 2 diabetes accounts for 95% of the 25.8 million diabetic cases (8.3% of the total U.S population) [14]. The increasing cases of diabetes in the U.S are as a result of rising age of population. The elderly and African-Americans are affected by diabetes disproportionately. 10.2% of whites over 20 years have diabetes compared to 18.7% African-Americans in the same age group. 26.7% Americans who are aged 65 and above have diabetes compared to 11.3% adults who are over 20 years. A study was conducted on 3,778 males and females in Iran where it was found out that ageing had a considerable association with diabetes odd ratio 1.052 [13].

There is gender related differences in people's lifestyle which may lead to differences in risk of getting diabetes and also differences in its prevalence in both men and women [15]. In the Sub-Saharan countries, it is more likely that female will be overweight or obese than male, this makes female to have higher prevalence of diabetes, compared with men, women in Uganda [16] were found to have higher prevalence of diabetes. However, in Nigeria [17] and Ghana [18], women were found to have lower prevalence of diabetes compared to men in the same areas. The results of a present meta-analysis showed that men living in low income countries in the Sub-Saharan region were more likely to be diagnosed with diabetes than women in those countries [19]. Also in a recent meta-analysis, the level of income of the country of residence appeared to also contribute in this heterogeneity between prevalence of diabetes and gender. Female of socioeconomic status in United States of America [20] appear to be at higher risk than male. In another meta-analysis, it was found that the incidence of type 2 diabetes in female with low socioeconomic status was higher than in male [21]. Impaired glucose tolerance is also a risk factor of diabetes, past researches have shown that impaired glucose tolerance is more common in female than in male. Female have a small muscle mass compared to men and therefore the muscle available for uptake of the glucose load (75g) used in the OGTT [22]. Female also have high levels of progesterone and estrogen which reduces the sensitivity of insulin in the whole body. Impaired glucose tolerance has also been associated with unhealthy diet and physical inactivity [23]. In Sub-Saharan countries, female are likely to be more physically inactive compared to male [24]. Another study was conducted on 3,778 males and females in Iran found no significant association between diabetes and gender [13]. Although there was no significant association, the prevalence of diabetes was higher in female compared to male, this may be explained by the fact that Iranian female spend more time at home than male and are engaged in activities which are less physical.

An individual is more likely to develop type 2 diabetes and gestational diabetes if they are obese or overweight with a Body Mass Index (BMI) of 30 or more. The more fatty tissue one has, the more resistant the cells become to insulin. Fat around one's abdomen increases one's risk because the fat releases chemicals that can upset the metabolic and cardiovascular systems of the body. This increases the risk of developing stroke, coronary heart disease and some type of cancer. In the U.S, type 2 diabetes accounts for 95% of the 25.8 million diabetic cases (8.3% of the total U.S population) [14]. More than 90% of type 2 diabetes patients are obese or overweight and increasing cases of diabetes in the U.S are as a result of high obesity rates. More than 35.7% of adults in the U.S are obese and this rate has nearly tripled from 1960-2010. Children and adolescents are being increasingly diagnosed with type 2 diabetes. However, it is difficult to calculate the prevalence of the condition because it can go undiagnosed for long time periods [25]. The rise of diabetes in children has been credited to the increasing rates of obesity in children. In a study [26] the prevalence of obesity in 2000 and 2001 was 19.8% and 20.9% respectively. In the same period, the prevalence of diabetes increased from 7.3% to 7.9%. In this study, a significant association was found

between diabetes and obesity. Obese subjects were 7.37 times more likely to become diabetic compared to adults whose weights were normal. Female with waist 80 cm or more are at a higher risk to develop type 2 diabetes. Studies have also shown that Asian male with waist 89 cm are more and White or Black male with waist 94 cm or more are at increased risk of diabetes. Regular exercising and reducing body weight by 5% could reduce the risk by more than 50%.

Studies have shown among affected people, at least one parent has diabetes. Some genes contribute 40-50% of the inheritable risk of type 1 diabetes. If a person is from a family where a sibling or parent has type 1 diabetes, the person's risk of developing the condition increases greatly. It is known that type 2 diabetes is inheritable. If a parent or sibling has diabetes, a person's risk of having the condition increases. Studies on families have revealed that a person from a family with type 2 diabetes history is 3 times more likely to have the condition compared to a person from a family with no type 2 diabetes history. Children born to individuals with type 2 diabetes were more likely to have high blood pressure and to be obese [27]. A person's risk of having gestational diabetes if she has pre-diabetes which is a precursor of type 2 diabetes is high. The risk also increases if a sibling or parent has type 2 diabetes. A person is also at a greater risk of developing gestational diabetes if she delivers a baby that higher mass, if the person had gestational diabetes during a previous pregnancy or had an unexplained still birth [28].

3. Methodology

3.1. Sampling Procedure and Sample Size

Sampling is the process of selecting a number of individuals from a population such that the selected sample is representative of the entire population. The sampling frame for the study was all individuals who visited the centre for diabetes check-up in 2016. The size of sample was determined by the Cochran formula [29] given as follows:

$$n_0 = \frac{z_{\alpha/2}^2 * p * q}{e^2} = \frac{1.96^2 * 0.5 * 0.5}{0.05^2} = 385$$

Where n_0 is the sample size, $z_{\alpha/2}$ is the standard normal distribution value at α level of significance, e is the level of precision and p is estimated proportion of the attribute that is present in the population under study and q is 1-p. Because the population was small, the sample size was reduced slightly using the correction formula:

$$n = \frac{n_0}{1 - \frac{n_0 - 1}{N}} = \frac{385}{1 - \frac{385 - 1}{339}} = 181$$

Where n is the sample size and N is the population size [30]

The data available was further divided into two strata of clinics and stratified sampling technique was used to determine the number of individuals to be selected from each stratum. The sample size for each stratum was obtained using the following formula, $n_i = \frac{N_i}{N}$, such that $N_1 = 218$ and $N_2 = 121$, that is, in the first clinic we selected sample of size $n_1 = 116$ patients and in the second clinic sample of size $n_2 = 65$ was selected using simple random sampling procedure such that each sample has an equal probability of selection.

3.2. Proposed Model and Inclusion Population

The study population was all patients attending Focus Medical Centre for diabetes check up. Depending on diabetes control, the patients attend the clinic on regular appointments. These patients are those who have been referred to the centre from other health facilities. The data available on diabetes in the clinic include: whether or not a person has diabetes, age, weight, gender, height, blood pressure and smoking status and alcohol consumption. Secondary data for the year 2016 were used. A multiple logistic regression model was used to model the risk factors (age, obesity, smoking status, alcohol consumption, gender and hypertension) of diabetes. The model is considered because the response variable (whether a person has diabetes or not) is dichotomous and the independent variables are a combination of continuous and categorical variables.

The multiple logistic regression model was used to analyze risk factors associated with diabetes. In this multiple logistic regression analysis, the odds of an outcome occurring or not is examined and by use of the natural log of the odds of an outcome as the dependent variable, the relationships can be made linear and treated like multiple linear regression. Consider response variable of binary Y and multiple predictor variables denoted by X_1, X_2… X_k. If π represents the probability, P (Y=1) that is presence of outcome and $1-\pi$ to represent probability, P (Y=0), multiple regression model can be written as:

$$\pi = P(Y \mid X_1, X_2, ..., X_k) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k}}$$

Where β_0, β_1… β_k are the regression coefficients.

The multiple logistic regression is rewritten such that the outcome is the expected log of odds that the outcome is actually present. $\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k$ where right hand side of the equation looks like the equation of a multiple regression model. However, in multiple logistic regression, the technique for estimating the regression coefficients is different from that of estimating the coefficients of multiple linear regression. The coefficients in logistic regression, for example β_i indicate the change in expected log odds relative to a one unit change in Xi, holding all other predictors constant. The antilog of the estimated regression coefficient, that is, e^{β_i} produces an odds ratio.

In order to determine whether a variable need to be included in the multiple logistic model, we use the chi-square test by testing the following hypothesis;

H_0: The diabetes status variable and predictor variable are independent

H_a: There is an association between one's diabetes status and the predictor variable

The test statistic is given as $X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$

Where O_i the observed value and E_i is the expected value, the null hypothesis is rejected if the p-value is less than 0.05 at $\alpha = 0.05$ level of significance.

3.3. Description of Study Variables

The variables and coding scheme is as described in Table 1.

Table 1. Study variables and coding scheme.

Variable	Description	Code Number
Diabetes	Dependent variable	0 - non diabetic, 1 - diabetic
Age	Continuous variable and persons aged 20 years and above.	Continuous variable
Gender	Gender of patients in hospital in prescribed period.	0 – male, 1 - female
Obesity	Index, BMI = $\frac{\text{Mass in Kilograms}}{(\text{Height in Metres})^2}$	0 - non-obese, 1 - obese
Hypertension	A hypertensive status, that is, either systolic blood pressure is $\geq$ 140mmHg or diastolic blood pressure is $\geq$ 90mmHg.	0 -non hypertension, 1 -hypertension
Smoking	Smoking status.	0 - non-smokers, 1 - smokers
Alcohol	Consumption of alcohol status	0 - non-alcoholics, 1 - alcoholics

3.4. Ethical Considerations

(i) Focus Medical Centre Ethical and Research Committee approved the study
(ii) The Director and the diabetes clinic managers gave authority to conduct the study

4. Results and Discussion

This section presents the research findings obtained by carrying out descriptive analysis on the risk factors associated with diabetes. In addition results of the multiple regression to examine the association between the dependant variable and the independent variables are discusses as well.

4.1. Descriptive Statistics

The collected data was recorded in a contingency table as show in Table 2.

Table 2. Descriptive statistic.

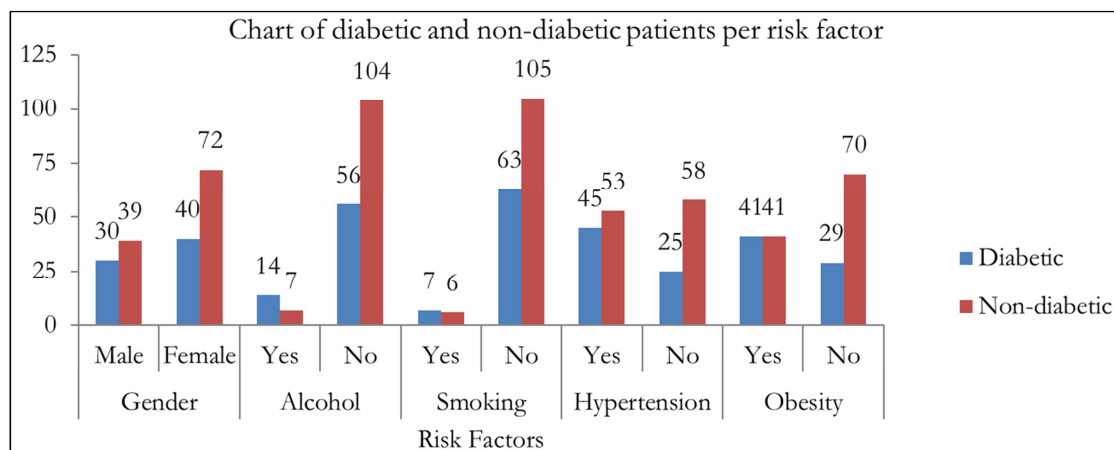|		Gender		Alcohol		Smoking		Hypertension		Obesity	
		Male	Female	Yes	No	Yes	No	Yes	No	Yes	No
Diabetic	Yes	30	40	14	56	7	63	45	25	41	29
	No	39	72	7	104	6	105	53	58	41	70

Figure 1. Chart of risk factors of diabetes and their frequencies.

4.2. Bivariate Analysis

In this section the association of every risk factor was tested with diabetes using Chi-square test and results recorded in Table 3. Since age is continuous variable, simple logistic regression model was fit to determine its association with Diabetes.

***Table 3.** Chi-square analysis for the different risk factors.*

Variable	p-value	Description
Age	0.0103	Significant
Gender	0.3764	Not significant
Obesity	0.0101	Significant
Smoking	0.0786	Not significant
Alcohol	0.0104	Significant
Hypertension	0.0432	Significant

The p-value of age 0.0103 is less than 0.05, age is therefore significant, hence there is an association between age and diabetes. The risk factors obesity, alcohol and hypertension are also significantly associated with diabetes since the respective p-values are less than 0.05.

4.3. Multiple Logistic Regression Model

A multiple logistic regression model was considered in which the risk factors age, obesity, hypertension and alcohol were included in the model after the chi-square analysis indicated that the risk factors are significant at α =0.05 level of significant. The analysis results of generalized linear model are recorded as shown in Table 4.

***Table 4.** Multiple logistic regression.*

Variable	Estimate	Std. Error	P-value
Intercept	-2.68853	0.72232	0.00019
Age	0.03134	0.01427	0.02800
Obesity	0.69474	0.33517	0.038192
Alcohol	1.35717	0.52341	0.009517
Hypertension	0.48316	0.33414	0.148179

The multiple logistic regression model fitted is given as;

$$P(Y=1|X_1,X_2,\ldots,X_k) = \frac{\exp(-2.6885 + 0.0313\text{Age} + 0.6947\text{Obesity} + 1.3572\text{Alcohol} + 0.4832\text{Hyper})}{1 + \exp(-2.6885 + 0.0313\text{Age} + 0.6947\text{Obesity} + 1.3572\text{Alcohol} + 0.4832\text{Hyper})}$$

However, it is observed that the risk factors age, obesity and alcohol are significant at α =0.05 level of significance, whereas hypertension is not significant. We consider the model in which only significant variables are included, that is, hypertension variable is removed from the model to get the best fit model. The analysis of the model gives the results recoded in Table 5 that we consider an improvement of the previous model in which all the variables are statistically significant.

***Table 5.** Modified multiple logistic regression.*

Variable	Estimate	Std. Error	P-value	Odds Ratio	Lower Limit	Upper Limit
Intercept	-2.4214	0.68819	0.00043	0.0888	0.0230	0.3421
Age	0.0304	0.01469	0.03178	1.0308	1.0026	1.0598
Obesity	0.7656	0.33018	0.02041	2.1503	1.1257	4.1073
Alcohol	1.4578	0.51418	0.00458	4.2964	1.5683	11.7701

The final model is then given as;

$$P(Y=1|X_1,X_2,\ldots,X_k) = \frac{\exp(-2.4214 + 0.0304\text{Age} + 0.7656\text{Obesity} + 1.4578\text{Alcohol})}{1 + \exp(-2.4214 + 0.0304\text{Age} + 0.7656\text{Obesity} + 1.4578\text{Alcohol})}$$

In interpreting the model coefficients, the reference group for obesity was taken as non-obese individuals, odds ratio for obesity is 2.1503, which means that a person who is obese is 2.1503 times likely to be diabetic compared to a person who is not obese. The odds ratio for individuals who consume alcohol is 4.2964, which means that an alcoholic is 4.2964 times likely to be diabetic compared to a non-alcoholic. The odds ratio for age is 1.0308 which means that as age of an individual increases by a unit, one is 1.0308 times likely to be diabetic. The intercept -2.4214 has log odd of 0.0888, that is, log odd of individual to become diabetic when the explanatory variables are at reference level.

4.4. Goodness of Fit for the Model

The fitted model was used to predict diabetes status and the results are shown in the Table 6 below.

***Table 6.** Goodness-of-fit model.*

		Predicted Diabetes	
		0	1
Observed Diabetes	0	96	15
	1	42	28

The results from the Table 6 indicate the following about the fitted multivariate logistic model;

(i) In a total of 111 patients who do not have diabetes, 96 (86.5%) are correctly predicted and 15 (13.5%) are incorrectly predicted by the model.
(ii) In a total of the 70 patients who have diabetes, 28 (40.0%) are correctly predicted and 42 (60.0%) are incorrectly predicted.
(iii) A total of 57 (31.5%) patients are misclassified.
(iv) In overall, when considering all sampled 181 patients, 124 (68.5%) are correctly predicted which reflects the overall explanatory strength of the model.

4.5. Hosmer and Lemeshow Goodness-of-Fit Test

The Hosmer and Lemeshow goodness-of-fit test is used to assess whether there is evidence of poor fit in model. The test simply compares the expected and observed defined by the predicted probability of the outcome in R and results of the test are recorded in Table 7.

***Table 7.** Hosmer and Lemeshow Goodness-of-fit of the overall model.*

Chi-square	p-value
7.3119	0.5033

The test statistic of 7.3119 with p-value of 0.5033 indicate

that we fail to reject the null hypothesis and conclude that there is no difference between the observed and predicted values, that is, there is no evidence of poor model fit.

5. Conclusion

The objective of this study was to model the risk factors of diabetes and the relationship of these risk factors with diabetes. The study results indicate that, as person ages, he or she is more susceptible to diabetes. We also observe that an obese person is more susceptible to diabetes and persons who consume alcohol are also more susceptible to diabetes compared to persons who does not consume alcohol.

Recommendations

(i) Special attention should be paid to individual who are aged, consume alcohol, obese or those exposed to at least one of these risk factors. The medical personnel in diabetic clinics should keep detailed records of the history of the patients to help them manage and overcome this problem.

(ii) Healthy lifestyles ought to be promoted among the general population and in particular, the diabetic patients, for example, people should be encouraged to involve themselves in regular exercises and consume healthy diet to prevent health problems such as hypertension, diabetes and other cardiovascular diseases which may result from being overweight.

(iii) Studies should be conducted to assess the effectiveness of different treatment interventions of managing diabetes.

References

[1]   World Health Organiation, (2016). Global Report on Diabetes.

[2]   American Diabetes Association, (2013). Economic Costs of Diabetes in the U.S. in 2012 Association, American Diabetes, Diabetes care, 36 (4).

[3]   International Diabetes Federation (IDF). (2015).

[4]   Ben, R., et al. (1991). Effects of chronic alcohol intake on carbohydrates and lipid matabolisim in subjetcs with type II (non-insulin-dependent) diabetes. The American Journal of Medicine: 90 (1), 70-76.

[5]   Feingold, K. and Siperrstein M. D. (1983). Normalization of fasting blood glucose levels in insulin requiring diabetes the role of ethanol absentention. Diabetes care, 6 (2), 186-188.

[6]   Johnson, K. H., M. Bazargan and E. G. Bing, (2000). Alcohol Consumption and compliance among inner-city minority patients with type 2 diabetes mellitus. Archives of Familty Medicine: 964-970.

[7]   Cox, W. M., et al (1996). Diabetic patients' alcohol use and qaulity life: relationships with prescribed treatment compliance among older males. Acoholism, Clinical and Experimental Research 20 (2), 327-31.

[8]   Will, C. Julie, et al. (2001). Cigarette smoking and diabetes mellitus: evidence of a positive association from a large prospective cohort study. International Journal of Epidemiology vol. 30 (3) 540-546.

[9]   Opie, I. H. and Y. K. Seedat. (2005). Hypertension in Sub-saharan African Population. Circulation 112 (23), 3562-3568.

[10] Mancia, G. (2005). The association of hypertension and diabetes: prevalence, cardiovascular risk and protection by blood pressure reduction. Acta Diabetol 42, 17-25.

[11] Maty, S. C., et al. (2005). Education, Income, Occupation, and the 34-year Incidence (1965-1999) of type 2 diabetes in Alameda County Study. International Journal Epidemiology 34 (6), 1274-81.

[12] Yuko, Morikawa, et al. (1997). Ten-year follow-up study on the relation between the development of non-insulin-dependent diabetes mellitus and occupation. Ameican Journal of Industrial Medicine 31, 80-84.

[13] Azimi-Nezhad, M., et al (2008). Prevalence of type 2 diabetes mellitus in Iran and its relationship with gender, urbanization, education, marital status and occupation. Singapore Medicine Journal 49 (7), 571-6.

[14] Fonseca, V. A., et al (2012). National Institute of Diabetes and Digestive Kidney Diseases (NIDDK).

[15] Belue, R., et al (2009). An overview of cardiovascular risk factor burden in sub-saharan African countries: a social-cultural perspective. Globalization and Health, 5 (1), 10.

[16] Lasky, D., et al (2002). Obesity and Gender differences in the risk of type 2 diabetes mellitus in Uganda. Nutrition (burbank, Los Angeles County, Califonia), 18 (5), 417-21.

[17] Ejim, E. C., et al (2013). Cardiovascular risk factors in middle-aged and elderly residents in South-East Nigeria: the influence of urbanization. Journal of the National Association Medicine of Resident Doctors of Nigeria, 22 (4), 286-91.

[18] Amoah, A. G., et al (2000). A national diabetes care and education programme: the Ghan model. Diabetic Research and Clinical Practice, 49 (2-3), 149-157.

[19] Kavanagh, A., et al (2010). Socialeconomic position, gender, health behaviors and biomarkers of cardiovascular disease and diabetes. Social Science and Medicine, 71 (6), 1150-60.

[20] Robbins, J. M., et al (2001). Social economic status and type 2 diabetes in African American and non-Hispanic white women and men: Evidence fron the third National Health and Nutrition Examination Survey. American Journal of Public Health, 91 (1), 76-83.

[21] Agardh, E., et al (2011). Type 2 diabetes incidence and social-economic position: a systematic review and meta-analysis. International Journal of Epidemiology 1-15.

[22] Faerch, K., et al (2009). Predictions of future fasting and 2-h post-OGTT plasma glucose levels in middle-aged men and women. Diabetic Medicine, 26 (4), 377-83.

[23] Assah, F. K., et al (2011). Urbanization, physical activity, and metabolic health in sub-saharan Africa. Diabetes Care 34, 491-496.

[24] Guthold, R., et al (2008). Worldwide Variability in Physical Inactivity. A 51-County Survey. American Journal of Preventive Medicine 34 (6), 486-494.

[25] Estimation of diabetes and its burden in the Epidemiologic estimation methods. National Diabetes Statistics Report. 2014.

[26] Mokdad, A. H., et al (2003). Prevalence of obesity, diabetes, and obesity-related health risk factors. Journal of American Medical Association 289 (1), 76-9.

[27] Florez, J. C., J. (2003). Hirschhorn and D. Altshuler. The inherited basis of diabetes mellitus: implications for the

generic analysis of complex traits. Annual review of genomics and genetics, 4, 257-91.

[28] International Diabetes Federation (IDF). 2013.

[29] Cochran, W. G. Sampling Techniques, (2007). New York: John Wiley & Sons.

[30] Glenn, D. Israel, (1992). Determining Sample Size. Fact Sheet PEOD-6. Florida: Florida Cooperative Extension Service.