FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 8 No. 1, February, 2024, pp 195 - 200
DOI: https://doi.org/10.33003/fjs-2024-0801-2263

PREVALENCE AND RISK FACTORS OF DIABETES MELLITUS AMONG WOMEN USING THE MULTINOMIAL LOGISTIC REGRESSION MODEL

[1]Olayeye, Teniola O., *[1]Bodunwa, Oluwatoyin K. and [2]Adewole, Ayoade I.

[1]Department of Statistics, Federal University of Technology Akure . Ondo State Nigeria
[2]Department of Mathematics, Tai Solarin University of Education Ijagun Ogun State Nigeria.

*Corresponding authors' email: okbodunwa@futa.edu.ng

ABSTRACT

Diabetes mellitus (DM) is a prolonged disease with debilitating effect on man. This includes many health problems because the disease is a risk factor for a number of complications. This study employs a multinomial logistic regression model to explore the prevalence of diabetes and identify contributing factors. Analyzing a diverse range of variables, the study aims to provide in-depth insights into the complex relationships influencing diabetes occurrence. The findings indicated that poor health status contributed more, among other factors, in terms of influencing diabetes. This could be as a result of having other health challenges. Also, women with stroke, high blood pressure, high cholesterol and heart disease were at greater risk of having diabetes compared to those not having. Women who were active had lower risk of having diabetes compared to those who were inactive as physical activities help control bodyweight through increased fat metabolism. Increasing age is often accompanied by a progressive decline in most physiological functions, resulting in increased susceptibility to disease. It was observed in this research that DM was more prevalent in elderly women than women of younger age.

Keywords: Diabetes mellitus, Risk Factors, Prevalence, Multinomial Logistic

INTRODUCTION

Diabetes mellitus (DM) is one of the most common human diseases and has become a significant public health concern worldwide. There were approximately 450 million people diagnosed with diabetes that resulted in around 1.37 million deaths globally in 2017 (Cho *et al.,* 2017). DM is defined as "a metabolic disorder characterized by hyperglycemia resulting from either the deficiency in insulin secretion or the action of insulin." (Davies et al., 2003, Magurová et al., 2012 and Xinjie Yu et al., 2020).

DM can be of three major types, based on etiology and clinical features. These are type1 DM (T1DM), type2 DM (T2DM), and gestational DM (GDM). In T1DM, there is absolute insulin deficiency due to the destruction of β cells in the pancreas by a cellular mediated autoimmune process. In T2DM, there is insulin resistance and relative insulin deficiency. GDM is any degree of glucose intolerance that is recognized during pregnancy. Pre-diabetes, also known as borderline diabetes or intermediate hyperglycemia, is observed in individuals with an FPG level from 100 mg/dl to < 126 mg/dl and an OGTT > 200 mg/dl (Lee *et al.,* 2014). There is an increasing prevalence of DM worldwide. Globally, the prevalence of diabetes is on the rise with an estimated 387 million diabetics; and it is estimated that by 2035, 592 million people will have diabetes (Abraham, 2013). As many as 27 million American women have prediabetes. Which can be reversed when healthy changes are made, such as doing some type of physical activity on most days, to lower the risk of getting diabetes and return to normal blood sugar levels. Also, losing 7% of body weight can lower the risk for type 2 diabetes by more than half. Prediction models can screen pre diabetes or people with an increased risk of developing DM to help decide the best clinical management for patients. The logistic regression models a relationship between the categorical response variable and covariates. Logistic regression can be extended to handle multiple responses (i.e. taking r>2 categories). Multinomial logistic regression is used when the dependent variable in question is nominal and for which there are more than two categories. The study aimed at examining the prevalence of Diabetes mellitus among women using the multinomial logistic regression model. Seyyed,*et al.* (2021) Conducted a study using the logistic regression model which was aimed at estimating the probability of experiencing diabetic foot ulcers at least for one time from the time of diagnosing DM up to the considered age. The covariates were sex, age, body mass index (BMI), fasting blood sugar (FBS), hemoglobin A1C (HbA1C), low-density lipoprotein (LDL), high-density lipoprotein (HDL), triglyceride (TG), insulin dependency, and statin use. Accordingly, age, body mass index (BMI), fasting blood glucose (FBS), and insulin dependency were positive predictors while high-density lipoprotein (HDL) was a negative predictor. This model seemed to be accurate enough, and since the *P* values of the final model were low, this model seemed to be repeatable for prospective use. (Dhakal, 2021) used multinomial logistic regression and a classification tree from the Pima Indian dataset to identify the important factors for type 2 diabetes. Variables were selected based on the goodness of fit test and model selection criteria such as *AIC*, *BIC*, and Mallows' Cp. The decision trees were plotted, and the prediction accuracy and cross-validation error rate were calculated for the purpose of validation. Variables selected from the logistic regression and decision tree are very similar, suggesting that the variable identified helps predict diabetes and may be used as a decision tool. Ifechukwude *et al.* (2015) conducted a study that was designed to estimate the burden of pre-diabetes and diabetes in Florida and to investigate their predictors in this population and to assess how the associations between the outcomes (pre-diabetes and diabetes). Some differences were observed in the degree of association for some of the predictor variables depending on the diabetes status.

MATERIAL AND METHODS

Table 1: Dataset Description

Feature	Explanation	Measurement	Range
Age	Age of respondent	Years	[18, .84]
Diabetes	If a respondent does not have diabetes, have pre diabetes or have diabetes	Categorical	[0,1,2]
Fruits	If a respondent takes fruits or not	Boolean	[0,1]
Veggies	If a respondent takes veggies or not	Boolean	[0,1]
Stroke	If a respondent has stroke or not	Boolean	[0,1]
High blood pressure	If a respondent has high blood pressure or not	Boolean	[0,1]
High cholesterol	If a respondent has high cholesterol or not	Boolean	[0,1]
Body Mass Index (BMI)	Respondent's BMI	Kilogram square (kgm^2)	[12, 96]
Physical activity	If a respondent is physically active or not	Boolean	[0,1]
Heart disease	If a respondent has heart disease or not	Boolean	[0,1]
General health	General health condition of respondents	Categorical	[1,2,….5]
Difficulty in walking	If a respondent has difficulty in walking or not	Boolean	[0,1]
Healthcare coverage	If a respondent has healthcare coverage or not	Boolean	[0,1]

Table 1 gives the description of the secondary data used for this study. Also, considering women of reproductive age, (15-49) would be difficult because the data consists of women from 18 years and above which means we can only consider the ages 18-49 when comparing the prevalence of diabetes in the reproductive age to other ages. The dataset contains originally 253,680 survey responses to the CDC's BRFSS2015. The Behavioral Risk Factor Surveillance System (BRFSS) is a health-related telephone survey that is collected annually by the CDC. Each year, the survey collects responses from over 400,000 Americans on health-related risk behaviors, chronic health conditions, and the use of preventive services. It has been conducted every year since 1984 (https://www.kaggle.com/code/alexteboul/diabetes-health-indicators-dataset-notebook.) The target variable Diabetes_012 has three features classes.0 is for no diabetes, 1 is for prediabetes, and 2 is for diabetes. This dataset has originally 21 features variables in which some were binary indicating the presence or absence of a particular condition. Since this research is about the prevalence of diabetes among women, the data was cleaned, hereby reducing the number of observations to 141,974 (Involving only women) with 13 variables. The dataset contains information on various clinical, body and lifestyle patterns of the patients, such as age, blood pressure, cholesterol, fruits consumption, veggies consumption, difficulty in walking, general health, stroke, heart disease, health care coverage, physical activity, body mass Index and the outcome variable, diabetes. STATA 14.0 software was used for the analyses.

The multinomial logistic function

Considering a logistic function for a binary response, Let the function be called $f(z)$ given by

$$f(z) = \frac{1}{1+e^{-z}} \quad (1)$$

where z varies from -∞ to +∞..

To obtain the logistic model from the logistic function, *z is written* as the linear sum

$$z = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \quad (2)$$

Where the x's are independent variables of interest and α and the βi 's are constant terms representing unknown parameters.

Substituting equation 1 into 2 to produce equation (3) below;

$$f(z) = \frac{1}{1+e^{-(\alpha+\sum \beta_i x_i)}} \quad (3)$$

probability statement is denoted as p(x) where x is a notation for the collection of variables x_1 through x_t

Thus, the logistic model may be written as

$$p(x) = \frac{1}{1+e^{-(\alpha+\sum \beta_i x_i)}} \quad (4)$$

However, since the above logistic model is non-linear, the logit transformation would be used to make it linear, this is given by

$$\text{Logit } p(x) = ln\left[\frac{p(x)}{1-p(x)}\right] \quad (5)$$

Where $p(x) = \frac{1}{1+e^{-(\alpha+\sum \beta_i x_i)}}$ (6)

This transformation allows the computation of logit p(x), for an individual with independent variables given by x.

By substituting Equation 5 into Equation 4, we obtain

$$ln\left[\frac{p(x)}{1-p(x)}\right] = ln\left[\frac{\frac{1}{1+e^{-(\alpha+\sum \beta_i x_i)}}}{\frac{e^{-(\alpha+\sum \beta_i x_i)}}{1+e^{-(\alpha+\sum \beta_i x_i)}}}\right] \quad (7)$$

$$= \ln\left[e^{(\alpha+\sum \beta_i x_i)}\right] \quad (8)$$

$$= \alpha + \sum \beta_i x_i \quad (9)$$

$$\text{Logit p(x)} = \alpha + \sum \beta_i x_i \quad (10)$$

$$\therefore \quad \text{Logit p(x)} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \quad (11)$$

Thus, the logit of p(x) simplifies the linear sum which is the logit transformation for a binary response.

The multinomial logistic function is an extension of the logit function. Following the work of Hosmer and Lemeshow (2013), multinomial logistic model in the form of the logit equation is expressed as the natural logarithm of the odds.

$$g_1(x) = ln\left\{\frac{P(Y=1|X)}{P(Y=0|X)}\right\} = \beta_{10} + \beta_{11}\chi_1 + \beta_{12}\chi_{2=X'\beta_1} \quad (12)$$

$$g_2(x) = ln\left\{\frac{P(Y=2|X)}{P(Y=0|X)}\right\} = \beta_{20} + \beta_{21}\chi_1 + \beta_{22}\chi_{2=X'\beta_2} \quad (13)$$

Where $g_j(x)$ is the logit function for response level j,

$$\pi_j(\mathrm{x}) = P(Y{=}j|\mathrm{x}) = \frac{e^{g_j(x)}}{\sum_{k=0}^{2} e^{g_k(x)}} \quad (14)$$

j= 0,1,2, with $\pi_j(\mathrm{x})$ being the probability of level j of the response variable

In multinomial regression, a set of co efficient are estimated for each category of the dependent variable. Since the estimated model is over-determined, the coefficients are scaled according to one of the categories. By default, the most frequent category is selected as the base category and their exponents are called Relative Risk Ratios. Fewer assumptions include: independence of observations, categories of the outcome variable must be mutually exclusive and exhaustive, no multicollinearity between independent variables, linear relationship between continuous variables and the logit transformation of the outcome variable.

RESULTS AND DISCUSSION

It was observed that out of 141,974 responses recorded, 120,959 (85.2%) were non diabetic, 2,604(1.83%) were at the risk of having diabetes that is, prediabetes and 18,411(12.97%) were diabetic. Thus, there were more non diabetic women compared to those that were diabetic and having prediabetes. It can be seen from the Table 2 below that diabetes is prevalent in the age group 65-69.

Table 2: Prevalence of Predictors and their Bivariate Relationship with the Polytomous Diabetes Status Variable among Women

Variables	Categories	No Prediabetes/ diabetes (%) n=120,959	Prediabetes (%) n =2,604	Diabetes (%) n = 18,411	Total (100%) n =141,974
Age	18-24	2,684 (97.78)	16(0.58)	45(1.640)	2,745
	25-29	3,865 (96.84)	37(0.93)	89(2.23)	3,991
	30-34	5,819(95.99)	52(0.86)	191(3.15)	6,062
	35-39	7,279 (94.23)	80(1.04)	366(4.740	7,725
	40-44	8,464(92.64)	96(1.05)	576(6.30)	9,136
	45-49	9,841(90.05)	185(1.69)	906(8.25)	10,928
	50-54	12,929(87.33)	234(1.58)	1,642(11.09)	14,805
	55-59	14,887(85.22)	327(1.870	2,225(12.91)	17,469
	60-64	14,918(81.65)	385(2.11)	2,968(16.240	18,271
	65-69	14,120(79.58)	373(2.10)	3,250(18.32)	17,743
	70-74	10,282(78.14)	324(2.46)	2,553(19.40)	13,159
	75-79	7,343(77.97)	241(2.56)	1,834(19.470	9,418
	80+	8,528(81.05)	254(2.41)	1,740(16.54)	10,552
Fruits	No	38,526(83.04)	975(2.10)	6,895(14.86)	46,396
	Yes	82,433(86.25)	1,629(1.70)	11,516(12.05)	95,578
Veggies	No	18,738(79.46)	577(2.45)	4,268(18.10)	23,583
	Yes	102,221(86.34)	2,027(1.71)	14,143(11.95)	118,391
Stroke	No	117,155(85.96)	2,453(1.80)	16,680(12.24)	136,000
	Yes	3,804(66.90)	151(2.66)	1,7319(30.44)	5,686
High blood pressure	No	78,822(93.48)	993(1.18)	4,506(5.34)	84,321
	Yes	42,137(73.09)	1,611(2.79)	13,905(24.12)	57,653
High cholesterol	No	76,814(91.77)	998(1.19)	5,890(7.040	83,702
	Yes	44,145(75.76)	1,606(2.76)	12,521(21.49)	58,272
Physical activity	No	27,892(76.80)	867(2.39)	7,561(20.82)	36,320
	Yes	93,067(88.09)	1,737(1.64)	10,850(10.27)	105,654
Heart disease	No	114,409(86.83)	2,307(1.75)	15,053(11.42)	131,769
	Yes	6,550(64.18)	297(2.91)	3,358(32.91)	10,205
General health	Excellent	24,954(97.65)	152(0.59)	449(1.76)	25,555
	Very good	46,075(92.63)	648(1.30)	3,017(6.07)	49,740
	Good	33.,24(81.03)	983(2.38)	6,865(16.59)	41,372
	Fair	12,164(66.42)	621(3.39)	5,529(30.19)	18,314
	Poor	4,242(60.66)	200(2.86)	2,551(36.48)	6,993
Difficulty in walking a	No	102,490(89.29)	1,789(1.56)	10,500(9.15)	114,779
	Yes	18,469(67.91)	815(3.00)	7,911(29.09)	27,195
Healthcare coverage	No	5,503(85.69)	139(2.16)	780(12.15)	6,422
	Yes	115,456(85.17)	2,465(1.82)	17,631(13.01)	135,552
BMI	Underweight	2,324(94.43)	24(0.98)	113(4.59)	2,461
	Healthy weight	42,440(94.42)	435(0.97)	2,072(4.61)	44,947

Over weight	41,247(87.35)	800(1.69)	5,172(10.95)	47,219
Obese	34,948(73.81)	1,345(2.84)	11,054(23.35)	47,345

The decision on which factor contributes to diabetes is known by the coefficients that are significant to the model. This is arrived at by comparing the p–value with the significance level (5%). Significance is established if p–value ≤ 0.05. The significance level used for the purpose of analysis was α = 0.05.

Note that, for a dichotomous variable, "No" was used as the reference(ref) category (All interpretation is in relation to this) while other reference categories were stated in Table 3.

Table 3: Multinomial Logistic Regression showing the risks factors and Likelihood of Diabetes and Prediabetes

Characteristic	Categories	Prediabetes			Diabetes		
		Coefficients (95% CI)	P-value	RRR	Coefficients (95% CI)	P- value	RRR
High blood pressure	Yes	0.422(0.332, 0.511)	0.000***	1.537	0.796(0.756,0.837)	0.000***	2.235
	No						
High cholesterol	Yes	0.581(0.497,0.666)	0.000***	1.789	0.617(0.580, 0.654)	0.000***	1.856
	No						
Stroke	Yes	-0.027(-0.202, 0.146)	0.754	0.976	0.179(0.111,0.247)	0.000***	1.201
	No						
Heart disease	Yes	0.033(-0.972, 0.167)	0.605	1.042	0.267(0.214, 0.319)	0.000***	1.314
	No						
Physical activity	Yes	0.052(-0.037,0.141)	0.253	1.035	-0.0671(-0.105,-0.028)	0.001***	0.920
	No						
Fruits	Yes	-0.0407(0.126, 0.044)	0.350	0.947	0.038(0.0002,0.076)	0.049***	1.025
	No						
Veggies	Yes	-0.136(0.237, -0.036)	0.007***	0.847	-0.093(-0.137,-0.489)	0.000***	0.881
	No						
Health care coverage	Yes	-0.212(0.392, -0.031)	0.021***	0.748	-0.036(-0.124,0.051)	0.421	0.907
	No						
General health	Excellent (ref)						
	Very good	0.498(0.319, 0.677)	0.000***	1.651	0.796(0.694, 0.899)	0.000***	2.232
	Good	0.969(0.792, 1.146)	0.000***	2.701	1.565(1.465, 1.665)	0.000***	4.914
	Fair	1.334(1.143, 1.526)	0.000***	4.061	2.077(1.972, 2.181)	0.000***	8.448
	Poor	1.195(0.960, 1.429)	0.000***	3.604	2.219(2.104, 2.335)	0.000***	9.885
Difficulty in walking	Yes	0.0479(-0.052, 0.148)	0.348	1.060	0.141(0.992, 0.184)	0.000***	1.162
	No						
Age	18-24(ref)						
	25-29	0.366(-0.224, 0.956)	0.224	1.401	0.132(-0.236, 0.501)	0.482	1.101
	30-34	0.224(-0.340, 0.788)	0.436	1.224	0.348(0.013,0.683)	0.041***	1.374
	35-39	0.362(-0.178, 0.903)	0.189	1.400	0.649(0.330,0.969)	0.000***	1.847
	40-44	0.303(-0.229, 0.837)	0.265	1.318	0.783(0.470,1.097)	0.000***	2.108
	45-49	0.719(0.203, 1.234)	0.006***	1.993	0.924(0.614,1.233)	0.000***	2.432
	50-54	0.592 (0.0803, 1.104)	0.023***	1.773	1.109(0.803,1.415)	0.000***	2.955
	55-59	0.710(0.202, 1.218)	0.006***	1.991	1.170(0.865, 1.475)	0.000***	3.140
	60-64	0.832(0.325, 1.339)	0.001***	2.235	1.388(1.084, 1.692)	0.000***	3.882
	65-69	0.823 (0.315, 1.331)	0.001***	2.241	1.491(1.187,1.796)	0.000***	4.341
	70-74	0.955(0.445,1.465)	0.000***	2.593	1.496(1.190,1.801)	0.000***	4.419
	75-79	0.958(0.443, 1.473)	0.000***	2.636	1.432(1.125, 1.739)	0.000***	4.198
	80+	0.911(0.396,1.427)	0.001***	2.528	1.267(0.960,1.575)	0.000***	3.567
Constant		-7.953(-10.072,-5.83)	0.000***	0.0009	-7.277(-7.918,-6.637)	0.000***	0.0005

It can be noted from Table 3 that all the following factors were significantly associated with diabetes; high blood pressure (pvalue<0.05), high cholesterol (pvalue<0.05), stroke(pvalue<0.05), heart diasease (pvalue<0.05), physical activity (pvalue<0.05), fruits consumption (pvalue<0.05), veggies consumption (pvalue<0.05), general health status (pvalue<0.05), difficulty in walking (pvalue<0.05), age group 30-80+ (pvalue<0.05), and BMI(pvalue<0.05). Therefore, we reject the null hypothesis (H_0) and conclude that there is enough evidence to show that these variables (predictors) are each not equal to zero at 95% confidence interval. This mean that these factors contribute to diabetes. The relative risk ratio (RRR) for "High blood pressure" was 2.325 times that of low blood pressure. This means that women with high blood pressure were at greater risk of having diabetes and were at lower risk of not having diabetes which is similar to previous studies (Latt, 2019). The RRR for the "poor heath" dummy variable indicated that the risk of women with a "poor health" status (relative to the risk of not having diabetes) was 9.885 times of women having an "Excellent health". This is said to be the most contributing factor with the greatest RRR.

We can infer that women in the age category 70-74 were 4.419 times at risk of having diabetes than those in the age category 18-24. This implies that older women had greater risk of having diabetes than younger women.

Also, the RRR of physical activity and veggies consumption were below 1, (0.920) and (0.881) respectively. This means that women who were physically active were at lower risk of having diabetes compared to those who were physically inactive. Similarly, women who consumed veggies were at lower risk of having diabetes compared to those who did not consume veggies. Obese women were at 5.303 times at risk of having diabetes compared to those underweight. This means that women having a body mass index (BMI) of 30 and above had greater risk of having diabetes than those with BMI less than 18. Women with healthcare coverage such as health insurance were at lower risk of having diabetes (RRR<1) than those with no healthcare coverage.

However, stronger associations were observed for each of the predictors of diabetes compared to pre-diabetes. For instance, being 30years or older increased the risk of diabetes but not prediabetes, stroke, fruits consumption, heart diseases and physical activity had significant associations with diabetes but not pre-diabetes.

By comparing the prevalence in the reproductive age group, which is age group 18-49(according to the data used) to other age groups. It was observed that diabetes was more prevalent in other age groups compared to the reproductive age group. This suggests that, diabetes was more prevalent in women of older age than women of reproductive age.

Table 4: Prevalence of diabetes between women of reproductive age and other age categories

Age Categories	Non-diabetic (%) (n=120,959)	Prediabetes (%) (n =2,604)	Diabetes (%) (n=18,411)	Total (100%) (n=141,97)
18-49	37,952 (93.51)	466 (1.15)	2,169 (5.34)	40,587
50-54	12,929 (87.33)	234 (1.58)	1,642 (11.09)	14,805
55-59	14,887 (85.22)	327 (1.87)	2,255 (12.91)	17,469
60-64	14,918 (81.65)	385 (2.11)	2,968 (16.24)	18,271
65-69	14,210 (79.58)	373 (2.10)	3,250 (18.32)	17,743
70-74	10,282 (78.14)	324 (2.46)	2,553 (19.40)	13,159
75-79	7,343 (77.97)	241 (2.56)	1,834 (19.47)	9,418
80+	8,528 (85.20)	254 (2.41)	1,740 (16.54)	10,522

From Table 3 above, the total number of women who were diabetic was about 18,411. In which 2,169 (11.8%) women belong to the age group 18-49 while 16,242(88.2%) belong to age group 50 and above. This implies that the prevalence of diabetes was high in older women (Fig. 1).

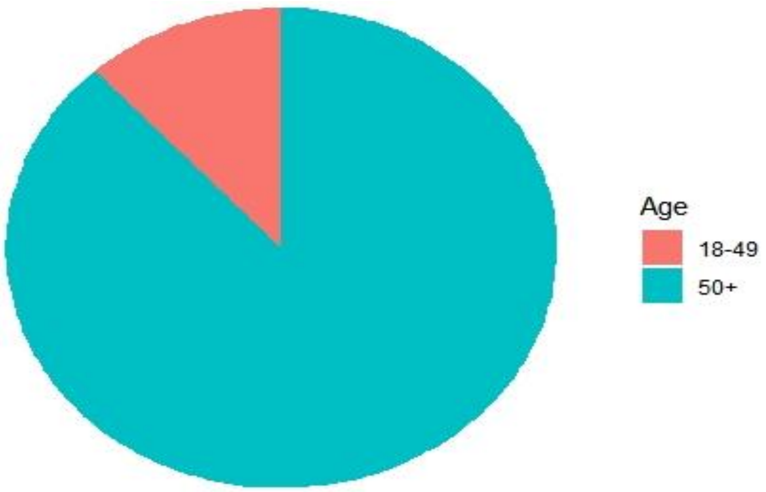

Figure 1: Pie chart comparing the prevalence of diabetes between women of reproductive age and other age categories.

Health status was the most contributing factor to diabetes. It is therefore necessary to estimate the probability of having diabetes according to the health status of the women. It was noticed that, women with “Poor” health status had greater probability of having diabetes than those who had “Excellent” health status.

Table 5: Probability of Diabetes according to health status

General health status	Probability	Std. error	Z	P value	95% CI	
Excellent	0.0375	0.0017	22.02	0.000	0.0341	0.0408
Very good	0.0762	0.0013	58.20	0.000	0.0736	0.0787
Good	0.142	0.00151	93.27	0.000	0.1392	0.1451
Fair	0.206	0.00263	78.51	0.000	0.2016	0.2119
Poor	0.230	0.0044	51.24	0.000	0.2213	0.2389

It can be seen from Table 5 that health status was significantly associated with diabetes as women with the" Excellent" health status had lower probability of having diabetes (0.0375) and those with "Very good" health status had a probability of 0.0762, women with "good" health status had a probability of 0.142, women whose health status can be regarded as "fair" had a probability of 0.206 while those in the "poor" health category had the greatest probability of having diabetes (0.230) .

CONCLUSION

This study provided evidence on the prevalence and predictors which influence diabetes in women using the BRFSS 2015 data. The model indicated that poor health status contributed more, among other factors, in terms of influencing diabetes. This could be as a result of having other health challenges as it is seen from the studies, women with stroke, high cholesterol, heart disease were at greater risk of having diabetes compared to those not having. Having high blood pressure approximately doubles the risk of diabetes compared to having low blood pressure. Also, women who were active had lower risk of having diabetes compared to those who were inactive as physical activities help control bodyweight through increased fat metabolism. On the contrary, physical activity was not significantly associated with pre-diabetes. It was also observed that obese women were at higher risk of having diabetes which could be due to the association between diabetes and weight. This seems to suggest that increased BMI increases the severity of the disease. which could be as a result of low or no physical activity. Also, diet plays a vital role in having diabetes as eating healthy such as veggies consumption reduces the risk of having diabetes. Fruits do not necessary reduce the risk as consuming more than the recommended daily allowance of fruit may add too much sugar to the diet. With regards to healthcare coverage, it was found that diabetes was significantly associated with having healthcare coverage. This suggests that women having healthcare coverage / insurance were at lesser risk of having diabetes compared to those not having. This could be as a result of regular checkups of general health. Increasing age is often accompanied by a progressive decline in most physiological functions, resulting in increased susceptibility to disease. Certain diseases are more prevalent in the elderly than in younger adults. Diabetes is not an exception; older women are more likely to have this disease compared to women of younger age groups.

REFERENCES

AbrahamTM, F. (2013). *Implications of rising prediabetes prevalence.* Diabetes Care. 36 (8) 2139- 2141.

Cho N.H, Shaw. J.E, Karurangs S., Huang Y., Fernandes R.,Ohiraggae A.W., and Malanda B. (2017). *Global estimates of Diabetes prevalence for 2017 and projections for 2045.*, Diabetes Research and Clinical practice.138, 271-281.

Davies R, Roderick.P and Rolfery J. (2003). The evaluation of disease prevention and treatment using simulation models. European Journal of Operational Research, 150, 53-66.

Dhakal, C. K. and Joshi R.D (2021). Predicting type2 diabetes using the logistic regression *and* Machine learning Approaches, *Int. J. Environ. Res. Public health*. 18 (14), 7346. Doi: https://doi.org/10.3390/ijerph18147346.

Hosmer Jr. D.W., Lemeshow. S. and Studivint R.X. (2013). Applied *Logistic Regression. 3rd Edition, John Wiley & Sons, Hoboken, NJ. https://doi.org/10.1002/9781118548367* .

Ifechukwude Obiamaka Okwechime, S. R. (2015). *Prevalence and Predictors of Pre-Diabetes and Diabetes among Adults 18 Years or Older in Florida:* A multinomial Logistic Modelling Approach. *PLos One 10 (12): e0145781. Doi: https://doi.org/10.1371/journal.pone,0145781* .

Lee, A (2014) Preventive Medine Race, regionality and pre-daibetes in the Reasons for Geographical and Racial deiffeneces in stroke REGARDS study

Latt, K.-K. Z. (2019). Diabetes Metabolic Syndrome and Obesity: Targets and Therapy, 291-298.

Magurová D, M.ajernfkova L., Sergej H., Hakan T. and Kerim G.. (2012). Knowledge of diabetes in patients with Type 2 diabetes on insulin therapy from Eastern Slovakia. Diabetologia Croatica 41(3) . 95-102.

Seyyed Amir Ysin Ahmadi, Razieh Shirzadegan, Nazanin Mousavi, Ermia Farokhi, Maryam Soleimaninejad and Mehrzard Jafarzadeh. (2021). Designing a logistic Regression model to predict diabetic foot ulcer in diabetic patients: High - Density Lipoprotein (HDL) Cholestrol was the Negative Predictor. *Journal of Diabetes Research*.1-6. 5521493 doi: https://doi.org/10.1155/2021/5521493 .

Xinjie Yu, Fang D., DA Lin, Hai Li, Jian Zhang, Q. Wang, X. S.Wang etal (2020). Prevalence of Diabetes, Prediabetes, and Associated Factors in an Adult Chinese population: Baseline of a Prediabetes Cohort Study. International journal of Endocrinology. Article ID 8892176, 1-8.