International Journal of Medical Science and Clinical Invention 7(8): 4951-4959, 2020
DOI:10.18535/ijmsci/v7i09.03 ICV: 77.2
e-ISSN:2348-991X, p-ISSN: 2454-9576

Case Report,

Logistic Regression Analysis To Determine Cardiovascular Diseases Risk Factors A Hospital-Based Case-Control Study, 2019.

Ebtehag Mustafa Mohammed[1], Ahmed Alnory[2]

1 Faculty of economic and rural development, University of Gezira, Sudan

2 Professor of statistics and epidemiology, University of Gezira, Sudan

*Corresponding author E-mail: ebtehagmustafa22@gmail.com

Abstract:

Cardiovascular disease (CVD) is one of the preventable causes of death in the world since the majority of its risk factors are controllable. A hospital-based case-control study was conducted and carried out in Gezira state (Wad Medani heart Centre). Data collection lasted for almost one year (From 2018 to 2019), via structured face-to-face interviews of 800 patients. Several risk factors for cardiovascular diseases, both modifiable and no modifiable were measured on 750 patients and 50 individuals as control. Summary descriptive statistics analysis was performed for some modifiable factors where information from the survey is viable. A Binary logistic regression model was adopted to examine the multiplicative effects of pathological, socio-economic and demographic variables as explanatory variables. The study showed that the most important causes leading to cardiovascular diseases are smoking with (9.84) times odds, followed by patients suffering from renal diseases with (6.19) times odds, in addition to drinking alcoholic with (3.62) times odds. The study recommended nonsmoking, avoiding alcohol and dealing well with chronic diseases.

Key words: logistic regression, cardiovascular disease (CVD), risk factors, chronic disease

Introduction:

Cardiovascular diseases:

Cardiovascular diseases are considered multifactorial conditions that especially affect the essential components of the circulatory system of the human body such as heart, blood vessel and blood itself (Mendis S et al, 2010). Cardiovascular diseases (CVD) are becoming a leading cause of morbidity, mortality, and disability in the world. It is becoming a large global burden. In 2010, the American Heart Association (AHA) defined national goals for ideal cardiovascular health that included meeting lifestyle-related recommendations for physical activity and dietary behaviors, nonsmoking, and a body mass index (BMI) less than 25kg/m^2.(Lloyd-Jones DM,, et al. 2010). The use of logistic regression modeling has exploded during the past decade. From its original acceptance in epidemiologic research, the method is now commonly employed in many fields including but not nearly limited to biomedical research, business and finance, criminology, ecology, engineering, health policy, linguistics and wildlife biology. Logistic regression sometimes called the logistic model or logit model, analyses the relationship between multiple independent variables and a categorical dependent variable. It estimates the probability of occurrence of an event by fitting data to a logistic curve. There are two models of logistic regression, binary logistic regression and multinomial logistic regression. Binary logistic regression is typically used when the dependent variable is dichotomous and the independent variables are categorical. When the dependent variable is not dichotomous and is comprised of more than two categories, a multinomial logistic regression can be employed (Park, Hyeoun-Ae, 2013).

As an illustrative example, consider how coronary heart disease (CHD) can be predicted by the level of serum cholesterol. The probability of CHD increases with the serum cholesterol level.

However, the relationship between CHD and serum cholesterol is nonlinear and the probability of CHD changes very little at the low or high extremes of serum cholesterol. This pattern is typical because probabilities cannot lie outside the range from 0 to 1. The relationship can be described as an 'S'-shaped curve. The logistic model is popular because the logistic function, on which the logistic regression model is based, provides estimates in the range 0 to 1 and appealing S-shaped description of the combined effect of several risk factors on the risk for an event (Kleinbaum & Klein, 2010). There are two primary reasons for choosing the logistic distribution. First, from a mathematical point of view, it is an extremely flexible and easily used function. Second, its model parameters provide the basis for clinically meaningful estimates of effect (David Hosmer, et al, 2013).

The logistic curve:

Logistic regression is a method for fitting a regression curve, $y = f(x)$, when y consists of binary coded (0, 1- -failure, success) data. When the response is binary (dichotomous) variable and x is numerical, logistic regression fits a logistic curve to the relationship between x and y. Logistic curve is an S-shaped or sigmoid curve, often used to model population growth (Eberhard & Breiwick, 2012).

Some related works:

Ezzat etal, 2017 of Illness **investigated** Predictors Perception in Patients Undergoing Coronary Artery Bypass Surgery. The purpose of this study was to determine the predictors of illness perception in patients undergoing coronary artery bypass surgery (CABG). **In a** cross-sectional study, 217 patients who had CABG for over 6 months were selected by systematic random sampling method and by using a questionnaire consisting of two sections including socio-demographic and illness perception instrument. Data were analyzed using descriptive statistics and inferential statistics (Kruskal-Wallis and Mann-Whitney test). Also they used multiple logistic regressions for detecting predictor factors. **Results:** The mean age of samples was 58.70 ± 9.46. The majority of them were male (61.3%), married (83.4%), housewife (34.1%), the education level of majority was high school diploma (44.2%). The majority had desirable illness perception (93.5%). The mean score of total illness perception was 83.53 ± 5.96 and mean scores of its domain including cause, treatment and control, consequence and timeline were 32.36 ± 3.6,28.07 ± 3.68,17.54 ± 2.26 and 5.57 ± 2.07,respectively. Multiple logistic regressions results showed about cause domain only monthly income ($P<0.034$, OR =0.290) and about treatment and control only age ($P< 0.014$, OR=2.217) were predictor factors. **Conclusion:** Fortunately in this study the majority of samples had desirable illness perception. Illness perception can affect the patients' adaptation with their illness and their treatment adherence after CABG. Thus, interventions to promote understanding of the disease appear to be necessary.

Mukund etal, 2017 studied Risk factors for Complex and Severe Coronary Artery Disease in Type 2 Diabetes Mellitus. They looked at the possibility of hyperinsulinemia being a marker for severe and complex coronary artery disease in type 2 diabetes mellitus, to select patients who would benefit from aggressive treatment. **Methods:** A cross sectional study of 290 type 2 diabetic patients, who underwent coronary angiogram for the evaluation of clinically suspected CAD at a tertiary care hospital were recruited. Biochemical and anthropometric parameters were analyzed. Insulin resistance was measured by homeostasis model assessment method. Angiographic ally measured syntax score of more than 22 is considered to be severe and complex CAD. Receiver operating curve characteristic was performed to find out the optimal cut-off value for insulin resistance and fasting insulin. Predictors of syntax score greater than 22 were identified by multiple logistic regression analysis. **Results:** An insulin level > 20 μIU/ml (OR: 6.86, 95% CI: 2.25-20.88) emerged as an independent predictor of severe and complex CAD. The optimal cut-off of insulin for predicting severe CAD was 20 with sensitivity and specificity of 80% (95% CI: 0.68 - 0.88) and 79% (95% CI: 0.73 - 0.83) respectively. **Conclusion:** Mobile Left Atrial Mass – Clot or Left Atrial Myxoma.

Ehab A. M. Frah, Mohammed A. M. Naeem,2015, Investigation Risk Factors of Cardiovascular Disease in Khartoum State, A case-control study was conducted in the Ahmed Gasim Hospital- Cardiac Surgery and renal Transplantation Center targeting the cardiovascular disease and were included patients who came for treatment or follow-up during 2015. A total of 162 patients with cardiovascular Disease (CVD) in Ahmed Gasim Hospital- Cardiac Surgery and renal

Transplantation Center and 162 control who are not patients or have any history of cardiovascular diseases. They were interviewed using purposively designed questionnaires. Logistic regression has been used for modeling the probability that a Cardiovascular Disease (CVD) could be developed as a function of risk factors. **Results**: This study shows that high bold pressure, family history and lack of physical activity are the main risk factors of cardiovascular diseases in Khartoum states, Sudan. **Conclusion:** Hypertension, family history of CVD and life style were found to be main risk factors of CVD.

Conceptual framework:

The suggested model deals with both modifiable and non-modifiable factors without very much details of the specific influence of each of them. Heart disease is a pathological intuitional disorder that is closely related to people dietary and habitual behavior and lifestyle. For this reason, it is governed by a diverse infrastructure. People who contract heart disease and those who do not are two very different clusters; for this reason it was relatively easy to conceptualize a relevant framework for analyzing heart disease risks. A starting point is to determine factors that are pathologically intuitional resulting from individual life long dietary behavior and habitual behavior or genetic (non- modifiable). This is based on the premise that most individuals begin life in a physical healthy state, barring congenital or genetic abnormalities. Further, these biological processes often work simultaneously and many times synergistically. Of these result in factors that affect the heard (**blood pressure, diabetic, cholesterol.** But there are some factors that are modifiable (residence, state, occupation, education level, state and marital status, etc…).

Methodology:

A hospital-based case-control study was conducted and carried out in Gezira state (Wad Medani Heart Centre), Data collection lasted for almost one year (From 2018 To 2019), via structured face-to-face interviews of 800 patients in Wad Medani heart hospital, Several risk factors for cardiovascular diseases were measured on 750 patients and 50 individuals control. Binary logistic regression model was adopted to examine the multiplicative effects of Pathological, socio-economic and demographic variables as explanatory variables. This hospital-based case-control study involved 750 CVD cases and 50 controls all recruited from Wad Medani Heart Centre in Sudan. The research study used simple random sampling equation for provisional sample size n*using the following formula:

$n^* = (t^2 \times p \times q)/d^2$ ………(1)

Where:

n^*: required sample size

p: anticipated population proportion taken as 50% because it gives the maximum possible sample size.

t = Confidence level, taken as 95%.

d = Absolute precision required on either side of anticipated proportion taken as 5 %, then:-

$n^* = (2^2 \times 50 \times 50)/25 = 400$.   …….(2)

However, in actual filed work this simple reckoning does not work because the population is not homogenous and we have to change the sample design from simple random sampling to multi stage cluster random sampling .For this reason we have multiply n by 2 (design effect).and the final sample becomes:

$n = 400 * 2 = 800$ ……. (3)

Results and discussions:

Distribution of respondents by age and sex

As presented in table (1) below the study sample compassed 440 males representing 55% and 360 females representing 45%. At the very young age group and the very old age group, there are more male patients than females 56% compared to 44% and 60% compared to 40% respectively. Female patients are concentrated in the three middle age groups 25-34, 35-44 and 45-54 (35.5 % males compared to 64.5% females). This result suggests that CVDs affect women at their mid ages more frequently than men while there is a higher probability that males contract CVDs at very early and very late ages. This result is supported by the fact that in our sample the youngest patient is a male 12 years of age while the youngest female patient is aged 16. Also, the oldest patient is a male 92 years of age while the oldest female patient is aged 88. The mean age for male patients is 60 years compared to 56.6 years for females giving probability ranges (57.1, 61.1) and (55.4, 57.7) respectively. Both distributions are skewed to the left as the means are lower than the mode and both with negative kurtosis. All these features are clearly reflected in the table and figure below.

Table (1): Age and sex distribution of the study sample

Age Group		Sex of respondents		Total
		Male	Female	
15-24	N	19	15	34
	%	55.9%	44.1%	100.0%
25-34	N	18	30	48
	%	37.5%	62.5%	100.0%
35-44	N	30	41	71
	%	42.3%	57.7%	100.0%
45-54	N	65	66	131
	%	49.6%	50.4%	100.0%
55-64	N	119	78	197
	%	60.4%	39.6%	100.0%
65-74	N	122	85	207
	%	58.9%	41.1%	100.0%
74>	N	67	45	112
	%	59.8%	40.2%	100.0%
Total	**N**	**440**	**360**	**800**
	%	**55.0%**	**45.0%**	**100.0%**
Minimum		12	16	12
Maximum		92	88	92
Mean		60.0	56.6	57.6
Standard Error		0.576	0.586	0.562
P-value = 0.017				

The conclusion that can be drawn from studying CVD patients by age and sex is that variation in cvds prevalence by age does exist between males and females. It is clear that only at the very broad middle age range are there more females than males with cvds. All results are significant at 5% level of significance, P value=0.017.

Distribution of passive smoking of respondents by locality:

Cardiovascular morbidity and mortality as a result of inhaled tobacco products continue to be a global healthcare crisis, particularly in low- and middle-income nations lacking the infrastructure to develop and implement effective public health policies limiting tobacco use. The pathophysiological effects of tobacco predispose both active tobacco users and passive smokers to the formation of atherosclerosis or narrowing of the arteries, leading to various types of cvds such as ischemic heart disease, cerebrovascular disease, peripheral artery disease, and aortic aneurysm. All tobacco products are inherently harmful, including smokeless tobacco, which contains over 2000 chemical compounds, including nicotine (United States Department of Health and Human Services, 2010). Table (2) bellow shows that the majority of respondents were passive smokers (79.4%), most of them from Greater Wad Medani; on the other hand, a small number of respondents were exposed to passive smoking in Al Kamleen locality. Value of the level of significance (0.03) in the table also shows the relationship between passive smoking and CVD.

Table (2): Distribution of respondents by locality and smoking:

Locality		Passive smoking		
		Yes	No	Total
Greater Wad Medani	n	211	44	255
	%	82.7%	17.3%	100.0%
East Gezira	n	98	34	132
	%	74.2%	25.8%	100%
South Gezira	n	153	41	194
	%	78.9%	21.1%	100%
Al-Managil	n	56	15	71
	%	78.9%	21.1%	100.0%
Hasaheisa	n	62	17	79
	%	78.5%	21.5%	100.0%
Al Kamleen	n	12	5	17
	%	70.6%	29.4%	100.0%
Umm Al-Qura	n	43	9	52
	%	82.7%	17.3%	100.0%
Overall (n=800)	n	635	165	800
	%	79.4%	20.6%	100.0%
P value = 0.03				

Body Max Index:

Body Mass Index (BMI), formally called Quetelet index, is a measure indicating nutritional status in adults. It is defined as a person's weight in kilograms divided by the square of the person's height in meters (kg $/m^2$). Figure 2 shows the classification that was used to determine the BMI.

Table (3): BMI classification:

BMI	Nutritional status
Below 18.5	Underweight
18.5-24.9	Normal weight
25.0-29.9	Pre-obesity
30.0-34.9	Obesity class 1
35.0-39.9	Obesity class 2
Above 40	Obesity class 3

BMI was developed as a risk indicator of diseases , some common conditions related to overweight and obesity include : premature death, cardiovascular disease , high blood pressure, some cancers and diabetes , once again , figure (1) below shows the state of BMI of the study sample , it shows that a large number of them suffer from obesity and abnormal BMI in general , this , in turn , confirms the relationship of the BMI to CVD.

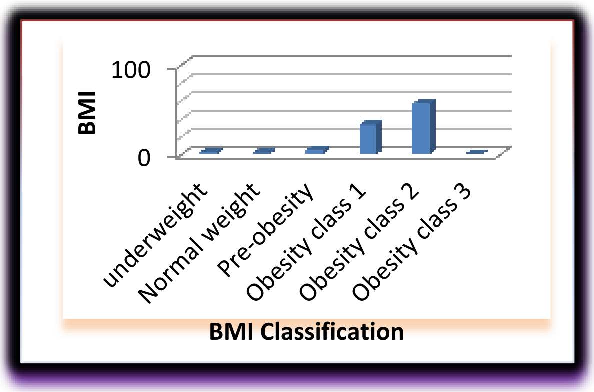

Fig. (1): Distribution of respondents by BMI

Alcohol and other drugs:

According to the Dietary Guidelines Advisory Committee (US)(United States Department of Health and Human Services ,2015), if alcohol is consumed, it should be consumed in moderation (≤1 and 2 drinks/day for women and men, respectively) and only by adults of legal drinking age. However, alcohol consumption guidelines vary substantially across the globe: low-risk guidelines range from 10–42 g/day or 98–140 g/week for women and 10–56 g/day or 150–280 g/week for men (Kalinowski A., Humphreys K, 2016). In 2016, 32.5% (25% women and 39% men) were current drinkers, and the mean amount of alcohol consumed was 0.73 standard drinks daily for females and 1.7 standard drinks daily for males (Griswold M.G., Fullman N., 2016) The respondents in this study agreed in the attitudinal scaling that alcohol is risk factor if consumption is not moderated. 89% of them agree that drinking alcohol and taking drugs are risk factor for CVD. This result was highly significant under chi-square test.However, only about 1.5% of the sample members drink alcohol or use other drugs.

Renal and other diseases:

Recent studies point out to the secondary role of chronic kidney disease in determining cardiovascular disease. Chronic kidney diseases in chronic pathology (such as diabetes, obesity, thyroid and respiratory diseases) are some of the diseases that should be linked with CVD some way or another, but our data file does not include any information on that respect.

Multivariate Analysis:

Deamination of the categorical dependent variable to examine the multiplicative effect of pathological, socio-economic and demographic and behavioral variables as explanatory on heart disease status as the dependent variable. Patients were separated into two categories: (1): Cases are labeled 1. (2) Controls are labeled (2). Table (4) below shows the result of the encoding for the dependent variable.

Table 4: Dependent Variable Encoding for multiple effects

Original Value	Internal Value
Heart patient	1
Non - heart patient	0

Binary logistic regression analysis for the respondent's characteristics

Table (5) below shows the Omnibus Test of Model Coefficients used to check that the new model (with explanatory variables included) is an improvement over the baseline model (does not include explanatory variables). It uses Chi-square tests to see if there is a significant difference between the Log-likelihoods of the baseline and the new model. The statistics for the Step, Model and Block are the same because we have not used stepwise logistic regression or blocking. The statistic shown in this table is the most reliable test of model fit for SPSS statistics binary logistic regression because it aggregates the observations into groups of similar cases.

Table 5: Tests of Model Coefficients for respondent's characteristics

		Chi-square	Df	Sig
Step 1	Step	4.374	2	.112
	Block	4.374	2	.112
	Model	4.374	2	.112

The model summary in the table (6) below provides the -2 Log-likelihood (-2LL), Cox & Snell R Square and Nagelkerke R Square values for the full model. The -2 LL value for the model (369.692). Cox and Snell's R –Square attempt to imitate R- squared based on likelihood and (usually less than 1). Here it is indicating that (5%) of the variation in the dependent variable is explained by the logistic model. Nagelkerke R Square is a more reliable measure of a relationship. It is indicating that (15%) of the variation in the dependent variable is explained by the logistic model.

Table 6: Model Summary for respondent's characteristics

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	369.692	.005	.015

Table (7) below shows the Hosmer and Lemeshow test. H- L test of the goodness of fit which suggests that the model is a good fit to the data. From the table Chi-square has 7 degrees of freedom, a value of 9.917 and probability of P = .0.1391 which is greater than 0.05 suggesting that the model was fit to the data well.

Table 7: Hosmer and Lemeshow Test for respondent's characteristics

Step	Chi-square	Df	Sig.
1	9.917	7	.193

Table (8) below shows the results of the contingency Hosmer and Lemeshow Test. The test divides subjects (patients with or without heart disease) into nine groups. For each of these groups, we then obtain the predicted group memberships and the actual group memberships. This result in a 2 x 9 contingency table, as shown (observed and expected). A chi-square statistic is computed comparing the observed frequencies with those expected under the linear model. As can be seen, they were slightly different from observed and expected values. But in general, the expected is within range of observed which once again reflects that the binary logistic fits the data reasonably.

Table 8: Contingency Table for Hosmer and Lemeshow Test for respondent's characteristics

		Non- heart patient		Heart patient		Total
		Observed	Expected	Observed	Expected	
Step1	1	9	8.229	70	70.771	79
	2	2	6.452	84	79.548	86
	3	7	6.248	83	83.752	90
	4	6	6.589	96	95.411	102
	5	5	4.548	70	70.452	75
	6	11	5.779	91	96.221	102
	7	4	3.922	71	71.078	75
	8	4	3.898	77	77.102	81
	9	2	4.336	108	105.664	110

The classification table shows the practical results of using the logistic regression model. Table (9) shows that the cases where the observed values of the dependent variable (heart disease status) were 1 or 0 respectively have been correctly predicted. From the table, the columns are the two predicted values of the dependent, while the rows are the two observed (actual) values of the dependent. In a perfect model, all cases will be on the diagonal and the overall per cent correct will be (100%). In this study, overall (93.8%) were correctly classified (750+0)/800=0.938), while 50 cases were classified incorrectly.

Table 9: Classification table Test for respondent's characteristics

Observed		Predicted		
		Non-heart patient	Heart patient	Percentage correct
Pathological case	Non-heart patient	0	50	.0
	Heart patient	0	750	100.0
Overall Percentage				93.8

Logistic regression parameterization respondent's characteristics

The parameter estimate coefficient B in the table (10) summarizes the effect of each predictor. The ratio of the coefficient to its standard error squared equals the Wald statistic, the standard interpretation of the binary logistic regression is that for a unit change in the predictor variable, the logistic regression of outcome is expected to change by its respective parameter estimate which is log- odd units .from the table, .locality had the highest odds ratio of .977. The odds ratio for age suggests that it is (.87) times more likely to get heart disease as the person gets older. However, we generally note that each of the two variables has p-value less than .05. This means that they have little effect on the quality and accuracy of the model as a whole.

Table 10: Variables in the Equation table for respondent's characteristics

Variables	**B**	**S.E.**	**Wald**	**Df**	**Sig.**	**Exp(B)**	**%95 C.I.for EXP(B)**	
							Lower	upper
Locality	023	.015	2.540	1	.111	.977	.950	1.005
Age	.139	.099	1.982	1	.159	.870	.716	1.056
Constant	3.539	.538	43.346	1	.000	34.441		

Logistic regression parameterization for pathological factors

The parameter estimate coefficient B in a table (11) summarizes the effect of each predictor. the ratio of the coefficient to its standard error squared equals the wald statistic, the standard interpretation of the binary logistic regression is that for a unit change in the predictor variable, the logistic regression of outcome relative is expected to change by its respective parameter estimate which is log- odd units, given that the other variables in the model are held constant. From the table, the Wald statistics for all the independent variables confirms the significant effect on heart disease status. Looking first at the result of regression for renal disease, there is a highly significant overall effect (Wald = 5.345, DF=1 p=.021), "Based on our output we can see that all 3 explanatory variables are significant. Renal disease had the highest odds ratio of 6.185, the odds ratio of renal disease suggests that it is (6.185) times more likely to get heart disease if patients suffering from renal diseases.

Table 11: Variables in the Equation table for pathological factors

Variables	B	S.E.	Wald	Df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	upper
renal disease	1.822	.788	5.345	1	.021	6.185	1.320	28.992
causes of disease	.011	.003	11.634	1	.001	.989	.983	.995
Symptoms	.140	.017	65.707	1	.000	.869	.840	.899
Constant	3.128	1.647	3.605	1	.058	22.829		

Logistic regression parameterization for behavioral factors

The parameter estimate coefficient B in the table (12) summarizes the effect of each predictor. the ratio of the coefficient to its standard error squared equals the wald statistic, the standard interpretation of the binary logistic regression is that for a unit change in the predictor variable, the logistic regression of outcome relative is expected to change by its respective parameter estimate which is log- odd units, given that the other variables in the model are held constant. Based on our output we can see that 2 explanatory variables are significant. Drinking alcoholic had the highest odds ratio of 9.8 (p value>0.05) the odd ratio of drinking alcoholic suggests that it is (9.8) times more likely to get heart disease if the person drinking alcoholic, also smoker persons which are (3.62) times more likely to have heart disease than other.

Wald for variable

$$1 \text{ (smoking)} = \frac{(B_i\hat{})^2}{[SE(B_i\hat{})]^2} = \frac{(1.286)^2}{[0.402]^2} = 10.23$$

Table 12: Variables in the Equation table for behavioral factors

Variables	B	S.E.	Wald	Df	Sig.	Exp(B)	95% C.I .for EXP(B)	
							Lower	Upper
Smoking	1.286	.402	10.227	1	.001	3.620	1.645	7.964
negative smoking	.904	.493	3.368	1	.066	2.470	.940	6.486
drinking alcoholic	2.286	.629	13.225	1	.000	9.836	2.869	33.722
Constant	-5.240	1.518	11.917	1	.001	.005		

Logistic regression parameterization for Functional Factors

The parameter estimate coefficient B in the table (13) summarizes the effect of the predictor. The standard interpretation of the binary logistic regression is that for a unit change in the predictor variable, the logistic regression of outcome relative is expected to change by its respective parameter estimate which is log- odd units. From the table, the Wald statistics for the independent variable confirms the significant effect on heart disease status. Based on our output we can see that our explanatory variable is significant, the odds ratio of operated heart injury suggests that it is (.144) times more likely to get heart disease if a person had an operated heart injury.

Table 13: Variables in the Equation table for Functional Factors

Variables	B	S.E.	Wald	Df	Sig.	Exp(B) odd ratio	5% C.I.for EX	
							Lower	upper
operated heart injury	-1.936	.601	10.384	1	.001	.144	.044	.468
Constant	6.275	1.172	28.662	1	.000	531.261		

Logistic regression parameterization for nutritional Factors

The parameter estimate coefficient B in the table (14) summarizes the effect of the predictors. From the table, the p-value of the two variables is greater than 0.05, and therefore have less influence on the model.\

Table 14: Variables in the Equation table for nutritional Factors

Variables	B	S.E.	Wald	Df.	Sig.	Exp (B)	%95C.I.for EXP(B)	
							lower	upper
using oil cooking	22.84	19.97	1.309	1	.253	.000	.000	1193370
meals that has been eaten with oil cooking	.511	.466	1.204	1	.273	1.668	.669	4.157
Constant	25.19	19.43	1.680	1	.195	8692003		

3.6 Summary of all variables in the logistic regression model:

The table below shows the ranking of all variables in our logistic regression model according to their ODD ratios. The table shows that smoking has the highest effect on cardiovascular disease, followed by suffering from renal disease, while operated heart injury has the lowest effect.

Table 15: Summary of all variables in the logistic regression model

Variable	B	ODD Ratio	influence
Smoking	1.286	9.84	10 to 1
renal disease	1.822	6.19	6 to 1
Drinking alcoholic	2.286	3.620	4 to 1
causes of disease	.011	.989	No specific effect
Symptoms	.140	.87	No specific effect
operated heart injury	-1.936	.601	No specific effect

Roc curve:

A measure of goodness-of-fit often used to evaluate the fit of a logistic regression model is based on the simultaneous measurement of sensitivity (true positive) and specificity (True negative) for all possible cutoff points. First, we calculate sensitivity and specificity pairs for each possible cutoff point and plot sensitivity on the y-axis by (1-specificity) on the x-axis. This curve is called the receiver operating characteristic (ROC) curve. The area under the ROC curve ranges from 0.5 and 1.0 with larger values indicative of better fit. Table (16) below shows that. The area under the curve is 0.985 with a 95% confidence interval (0.976, 0.995). Also, the area under the curve is significantly different from 0.5 since the p-value is 0.000 meaning that the logistic regression classifies the group significantly better than by chance.

Table 16: Area under the curve

Test result variables		Predicted probability	Asymptotic 95% confidence interval	
Area	std. Error	Asymptotic sig		
0.985	.004	.000	.976	.995

Conclusion:

This study was conducted on a group of patients in Medani heart Centre in Wad Medani, Sudan. The study investigated the factors affecting human health and thus leading to risks causing cardiovascular disease. The study showed that the most important cause leading to cardiovascular diseases is smoking with (9.84) times odds, followed by patients suffering from renal diseases with (6.19) times odds, in addition to drinking alcoholic with (3.62) times odds. The study recommended not smoking, avoiding alcohol and dealing well with chronic diseases.

References:

[1.] David Hosmer, Stanley Lemeshow, Rodney Sturdivant, 2013, Applied Logistic Regression-Wiley

[2.] Eberhardt, L. L., & Breiwick, J. M., 2012, Models for population growth curves. *ISRN Ecology, 2012, 1-7.* http://dx.doi.org/doi:10.5402/2012/ 815016

[3.] Ehab A. M. Frah, 2015, Investigation Risk Factors of Cardiovascular Disease in Khartoum State, Sudan: Case - Control Study, sudan

[4.] Ezzat etal, 2017, Predictors Perception in Patients Undergoing Coronary Artery Bypass Surgery. Epidemiology, physiology, and biochemistry. Circulation. 1991; 83: 1–12

[5.] Griswold M.G., Fullman N., Hawley C., Arian N., Zimsen S.R., Tymeson H.D. Alcohol use and burden for 195 countries and territories, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016-ClinicalKey. [(Accessed on 22 October 2019)]; Lancet. 2018 392:1015–1035. Doi: 10.1016/S0140-6736(18)31310-2.

[6.] Kleinbaum, D. G., & Klein, M., 2010, Logistic regression (statistics for biology and health) (3rd ed.), New York, NY: Springer-Verlag New York.

[7.] Kalinowski A., Humphreys K. Governmental standard drink definitions and low-risk alcohol consumption guidelines in 37 countries. Addiction. 2016; 111:1293–1298. doi:10.1111/add.13341. [PubMed] [CrossRef] [Google Scholar][7] Lloyd-Jones DM, Hong Y,

[8.] Labarthe D, et al., 2010, Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic Impact Goal through 2020 and beyond Circulation. Feb 2010;121(4):586-613

[9.] Mendes S, Puska P, Norrving B ,editors, 2011 , global Atlas on Cardiovascular Disease Prevention and control . World Health Organization, Geneva

[10.] Park, Hyeoun-Ae, 2013, an Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain, Seoul, Korea.

[11.] Mukund etal, 2017, Risk factors for Complex and Severe Coronary Artery Disease in Type 2 Diabetes Mellitus. Global Atlas on Cardiovascular Disease Prevention and control. World Health Organization, Geneva