# MODELLING THE SURVIVAL RATE FOR CARDIOVASCULAR DISEASE PATIENTS: A CASE STUDY OF METHODIST HOSPITAL, UGBOJU, BENUE STATE

**ELLA, BENEDICT ANTHONY**

**M. TECH/OR/07/0166**

**SEPTEMBER, 2012.**

# MODELLING THE SURVIVAL RATE FOR CARDIOVASCULAR DISEASE PATIENTS: A CASE STUDY OF METHODIST HOSPITAL, UGBOJU, BENUE STATE

**BY**

**ELLA, BENEDICT ANTHONY**

**(M. TECH/OR/07/0166)**

**A PROJECT REPORT SUBMITTED TO**

**THE DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH, SCHOOL OF PURE AND APPLIED SCIENCE, MODIBBO ADAMA UNIVERSITY OF TECHNOLOGY, YOLA**

**IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF TECH DEGREE IN OPERATIONS RESEARCH**

**SEPTEMBER, 2012.**

## DECLARATION

I hereby decree that this project report was written by me and it is a record of my own research work. It has not been presented before in any previous application for a higher degree. All references cited have been duly acknowledge.

## DEDICATION

This project is dedicated to God Almighty for His inspiration, guidance and protection throughout the period of my study.

## APPROVAL PAGE

This project report entitled “Modelling the Survival Rate for Cardivascular Disease Patients A Case Study of Methodist Hospital, Ugboju, Benue State” meets the regulations governing the award of master of Technology of the Modibbo Adama University of Technology, Yola and is approved for its contribution to knowledge and literary presentation.

_____ _____

Dr. A. Okolo Date

Supervisor

_____ _____

Dr. I.J. Dike Date

Internal Examiner

_____ _____

Prof. Date

External Examiner

_____ _____

Dr. A. Okolo Date

Head of Department

_____ _____

Prof. R.M. Odekunle Date

Dean, Postgraduate School

## ACKNOWLEDGEMENT

Thanks be to God Almighty for His guidance and protection. I wish to express my profound gratitude to my parents Mr and Mrs. Ella Bernard Onoja and my wife Mrs. Anthony Victoria, my sons Adakole and Onoja for their encouragement and support during the period of my study and my concerned sisters and brothers, nephews and nieces, Esther Bernard, Onyeche Bernard, Mrs Bala Omeche, Alice Owoicho, Charles Odeh, Oche Bernard, Gabriel Bernard, Onoja Bernard and Onyemowo Bernard.

I am particularly grateful to my able supervisor, Head of Department, Dr. A. Okolo for his patience, encouragement, supervision and all the necessary guidance he provided throughout the period of my study. And a special thank goes to all my friends: Ochekwu Agbenyi, Damian Omale, Emma Bello, Adah Adaka, Inalegwu Agbenyi, Nelson onum, John Ejaidu, Oga Bello Ikili and Adaedoh Audu.

My specially thanks also goes to Dr. S.S. Abdukadir and Dr. Mua'su Hamadiko for all the necessary assistance support and encouragement they gave me during the period of my study. And to all the lecturers in the Department of Statistics and Operation Research I say thank you for making the study a success.

I would also like to express my gratitude to Naval Capt. Bala Musa for his assistance. A special thanks goes to Mr. Raphael Odoh, Director of Metro

International School, Kubwa, Abuja for his countless favour during this project work.

I would not forget to express my deep appreciation to my colleagues Mr. Nyor Ngutor, Mr. Bala Aji, Mr. Ayuba Ibrahim, Mr. Hassan Umar, Mr. Auwal Sale and Mrs. Penninah Abraham for their tolerance and exchange of ideas during the course of the programme.

Finally, I wish to express my profound gratitude for help and encouragement I received from individual whose names cannot all be mention here.

## ABSTRACT

Cardiovascular diseases have over the years been a major cause of deaths annually in developing countries. Several drugs at different dose levels are in use to manage the various cases, without much success. This study focuses on modeling the survival rate for cardiovascular disease patients. Here data was obtained from record cards of Methodist Hospital on five cardiovascular related diseases. Data was collected on drug level, age and blood cholesterol level for renovascular, sleep apnea, thyroid, primary aldosteronism and Cushing syndrome diseases. A logistic regression analysis was used to develop models to assess the survival rate of patients, a modified Pearson goodness-of-fit test was adopted to ascertain the adequacy of the fitted models and each disease was discussed separately. The study revealed that intensive reduction and monitoring of blood cholesterol level would increase survival of patients suffering from these diseases. The fits were also shown to be adequate.


## TABLE OF CONTENTS

**CHAPTER THREE: RESEARCH METHODOLOGY**

## LIST OF TABLE

## CHAPTER ONE

# INTRODUCTION

## 1.1 BACKGROUND OF THE STUDY

Hypertension is a common health problem in developed countries and a major risk factor for cardiovascular disease (Castelli, 1984). Cardiovascular disease is responsible for about 23,800 deaths annually in the United Kingdom. This account for 36% of total mortality there (W.H.O., 2002). Its prevalence is probably on the increase in developing countries where adoption of western lifestyles and the stress of urbanization both of which are expected to increase the mortality associated with unhealthy lifestyle are not on the decline (Castelli, 1984). Genetic and environmental factors are reported to play a key role on hypertension. High blood pressure in adults has a high impact on the economy and on the quality of life of individuals with implications on resource expenditures.

High blood pressure is common worldwide and is regarded as a major public health problem (Murray & Lopez, 1997; Wolf *et al.,* 2003). It is the most common cardiovascular disease in black Africa and a major cause of morbidity and mortality among Nigerians (Balogun & Ladipo, 1988; Lawal & Falase, 1998; Cooper & Rotimi, 1997).

In blacks, hypertension prevalence studies show 24% in St Lucia, 26.9% in Jamaica and 27.9% in Barbados (Freeman *et al.* 1996). Prevalence study in Ghana shows 5% for rural dwellers and 13% for urban dwellers. Another study carried out among the Ashantis in Ghana living in semi-urban and rural villages found a prevalence of 28.7%. (Freeman *et al*., 1996).

While some studies have been carried out in some parts of Nigeria (Akinkugbe & Ojo, 1968; Nwankwo *et al*, 1990; Akinkugbe, 1996; Kaufman *et al*. 1999), reports indicate that incidence of hypertension in adults living in the Niger Delta region is on the increase. This study was

carried out to assess the prevalence and treatment of cases of hypertension in Ugboju, Benue State of Nigeria.

## 1.2 STATEMENT OF THE PROBLEM

Hypertension is of serious concern to the world as it leads to undue strain in the heart, causes heart attack, heart failure and eventually death of patients. As such, there is the urgent need to minimize if not completely eradicate health problems associated with hypertension.

It is therefore of interest to undertake this studies that could give the information that will assist in reducing the risk of having somebody dead while on treatment for hypertension or its related disease.

## 1.3. OBJECTIVES OF THE STUDY

The objectives of this study are as follows;

i. To obtain an appropriate model for the various diseases associated with hypertension.
ii. To determine the contributions of the predictor variables of the various disease to survival rate.

## 1.4 SIGNIFICANCE OF THE STUDY

The study is significant as it should assist to reduce mortality rate associated with hypertension or its related ailment.

## 1.5 SCOPE OF THE STUDY

This study covers only high blood pressure patients of Methodist hospital Ugboju, Benue State, Nigeria.

## 1.6 LIMITATION OF THE STUDY

A secondary source of data collection was employed in the study. Such data may not be properly recorded which might limit the study. This investigation is limited to hypertension as recorded in our case study. An expansion of investigation was not possible due to inadequate finance and time constraint.

# CHAPTER TWO

# LITERATURE REVIEW

## 2.1 HYPERTENSION

Hypertension is the medical term for high blood pressure. Blood pressure refers to the pressure that blood applies to the inner walls of the arteries. Arteries carry blood from the heart to other organs and parts of the body. Rose (2007) argued that "an individual's blood pressure is defined by two measurement; systolic and diastolic pressure. Systolic pressure is the pressure in the arteries produced when the heart contract (at the time of heart beat) and diastolic pressure refers to the pressure in the arteries during relaxation of the heart between heart beats".

Untreated hypertension increase the strain on the heart and arteries, eventually causing organ damage. Hypertension increase the risk of heart failure, heart attack (myocardial fraction), and stroke.

The standard definition of high blood pressure is determined by the Joint National Committee (J.N.C) on evaluation and diagnosis of High blood pressure (Chobanian & Hill, 2007). A person is considered to have high blood pressure after three to six elevated blood pressure measurements over several months (Chobanian & Hill, 2007). These definitions apply to adult who are healthy and not using medication for high blood pressure.

Person (s) with systolic blood pressure 120mmHg and diastolic blood pressure of 80mmHg is said to have normal blood pressure. Bakris (2007) claimed that patients in the pre hypertension range are at increased risk of progressing to hypertension and developing cardiovascular complications.

## 2.2 RISK FACTOR

Hypertension is a common health problem. In the United States, approximately 32 percent of African-Americans and 23 percent of white people and Mexican-Americans have hypertension (Bakris, 2007).

Hypertension is more common as people grow older. Post (2005) claimed that among people over age 60, hypertension occurs in 65 percent of white men, 80 percent of African-American women, and 65 percent of white women.

Unfortunately, many people's blood pressure is not well controlled.

According to Post (2005) hypertension is in good control in only 25 percent of African-Americans and whites, and 14 percent of Mexican-Americans.

## 2.3 DETECTION AND EVALUATION OF HIGH BLOOD PRESSURE

For more than three decades, the National Heart, Lung, and Blood institute (NHLBI) has administered the National High Blood Pressure Education program (NHBPEP) coordinating committee, a coalition of 39 major professionals, public, and voluntary organizations and several federal agencies. One of their important functions is to issue guidelines and advisories designed to increase awareness, prevention, treatments and control of high blood pressure (hypertension).

Chobanian and Hill (2007) observed that "Individuals with a systolic blood pressure of 120 to 139mmHg or a diastolic blood pressure of 80 to 89mmHg should be considered as pre-hypertension and require health-promoting life style modifications to prevent cardiovascular diseases". Black et.al (2008) claimed that the report of the Joint National Committee on

Prevention and Treatment of High Blood Pressure provides a guideline for prevention and management of the problem.

The main thrust of the report highlighted was that;

i. In persons older than 50 years, systolic blood pressure (BP) of more than 140mmHg is a much more important cardiovascular disease risk factor than diastolic blood pressure.
ii. The risk of cardiovascular disease, beginning at 115/75mmHg doubles with each increment of 20/10mmHg; individual who are normotensive at 55 years of age have a 90% life time risk for developing high blood pressure.

## 2.4 CARDIOVASCULAR DISEASE RISK

High blood pressure affects approximately 50 million individuals in the United State of America and approximately 1 billion individuals worldwide. Vassan et.al (2002) claimed that as the population ages, the prevalence of high blood pressure will increase even further unless broad and effective planning and management of patients are implemented. Recent data from the Framingham Heart Study also suggested that individuals who are normtensive at 55 years of age have 90% life time risk for developing high blood pressure. Whelton et.al (2002) argued that the relationship between high blood pressure and risk of cardiovascular disease is continuous, consistent, and independent of other risk factors. The higher the blood pressure the greater the chance of heart failure (HF), stroke, and kidney disease. Similarly, Vollmen, *et al* (2007) adduced that for individuals aged 40 to 70 years, each increment of 20mmHg in systolic blood pressure

or 10mmHg in diastolic blood pressure doubles the risk of cardiovascular disease across the entire BP range 115/75 to 185/115mmHg.

However, any fallout from this relationship signals the needs for increased education of health care professionals and the public to decrease high blood pressure level and to prevent the development of high blood pressure in the general population. Izzo *et al* (2006) observed that when a doctor fail to prescribe life style modifications, adequate anti high blood pressure drug doses, or appropriate drug combinations, inadequate BP control may result.

## 2.5 BENEFIT OF LOWERING BLOOD PRESSURE

A clinical study by Vassan *et al* (2002) revealed that anti hypertensive therapy is associated with 40% to 30% mean reduction in stroke incidence; 28% to 25% in myocardial infection; and more than 50% in heart failure.

It is estimated that in patients with stage 1 hypertension (systolic BP, 140-159mmHg and/or diastolic BP, 90-99mmHg), and additional cardiovascular risk factor, achieving a sustained 12mmHg decrease in systolic Bp for 10 years will prevent one death for every eleven patients treated. However, it was observed by Sheps and Roccella (2004) that in the presence of cardiovascular disease of target organ damage, one patient would require blood pressure reduction to prevent death.

## 2.6 AMBULATORY BLOOD PRESSURE MONITORING

Ambulatory blood pressure monitoring, or measuring blood pressure at regular intervals throughout the day, is very important in managing patients with treatment-resistant hypertension.

According to Chabanian and Hill (2007), continuously measuring blood pressure may help predict heart disease and related death among individuals with treatment-resistant hypertension. In a related study carried out by Gil and Salles (2007) they argued that blood pressure reading taken in a medical clinic do not appear to predict future heart risk.

A clinic study performed by Gil and Salles (2007) suggested that ambulatory blood pressure monitoring should be performed throughout the whole day, with separate analysis of day time and night periods, because it seems that night time blood pressure are better cardiovascular risk factors than day time blood pressure However, in most individuals blood pressure decreases by 10% to 20% during the night; those in whom such decreases are not present are at increased risk for cardiovascular events.

## 2.7 PATIENTS EVALUATION

Evaluation of patients with documented hypertension has three objectives;

i. To assess lifestyle and identify other cardiovascular risk factors or concomitant disorder that may affect prognosis and treatments;
ii. To reveal identifiable cause of high blood pressure;
iii. To assess the pressure or absence of target organ damage and cardiovascular diseases. The data needed are acquired through medical history, physical examination, routine laboratory tests and other diagnosis procedures.

## 2.8 MONITORING BLOOD PRESSURE

Once anti-hypertensive drug therapy is limited, most patients should return for follow-up and adjustment of medications at approximately monthly interval until the blood pressure goal is reached. Hill and Miller *(1996)* argued that more frequent visits would be necessary for patient with complicated Comorbid conditions. After blood pressure is at goals and stable, follow-up visits can usually be at three to six month intervals.

Comorbidities, such as heart failure, associated diseases, such as diabetes, and the need for laboratory test influence the frequency of visits. Other cardiovascular risk factors should be treated to their respective goals, and tobacco avoidance should be promoted vigorously.

## 2.9 IMPROVING HYPERTENSION CONTROL

Behavioural models suggest that the most effective therapy prescribed by the most careful clinical will control hypertension only if the patient is motivated to take the prescribed medication and to established and maintain a health promoting lifestyle. Tepper (2001) observed that "motivation improves when patients have positive experience and trust in their clinicians". However, patients' attitudes are greatly influenced by cultural differences, beliefs, and previous experience with the health care system.

This attitude must be understood if the clinician is to build trust and increase communication with patients and families.

## 2.10 PUBLIC HEALTH CHALLENGES AND COMMUNITY PROGRAM

Public health approached, such as reducing calories, saturated salt in processed foods and increasing community and school opportunities for

physical activity, can achieve a downward shift in the distribution of a population's blood pressure, thus potentially reducing morbidity, mortality, and life time risk of an individual becoming hypertensive. In a study by Barlow and Dietz (1998), they reported that about one hundred and twenty two million adult Americans are over weighed or obese, which contributes to the risk in blood pressure and related conditions. When public health intervention strategies address the diversity of their services, the likelihood of their acceptance by the community increase. This public health approaches can provide an attractive opportunity to interrupt and prevent the community costly cycle of managing hypertension and its complications.

## 2.11 THE EFFECT OF SMOKING ON BLOOD PRESSURE

The chronic epidemiological effects of cigarette smoke on the incidence and level of hypertension and in conjunction with hypertension as an addition risk factor; cardiovascular disease is a thing of serious health concern.

The acute and transient effect of smoking in man is to increase heart rate and blood pressure. These effects are thought to be due primarily to the action of nicotine releasing cathechlamines. Beaumont and Buxtorf (1993) reported that paroxysmal arteries hypertension is a reaction to cigarette smoking in which, under clinical diagnostic testing, high nicotine cigarette induce a rise in blood pressure of about 50mmHg systolic and 20mmHg diastolic over about 20 minutes. The reaction was accompanied by headache, palpitations, and sweating. However, in the presence of hypertension as a risk factor for coronary heart disease, smoking acts synergistically to increase the effective risk by joining the risks attributable to hypertension and to smoking alone.

## 2.12 LOGISTIC REGRESSION MODEL

Logistic regression is a form of regression analysis that is specifically tailored to situation in which the dependent variable is binary (or dichotomous) (Anthony, 1990). For example, among a sample of people under investigation, a researcher might be interested in what factors are associated with the likelihood of someone being employed rather than unemployed receiving university education or not receiving university education, live/die, has disease/doesn't have disease.

In recent years, multinomial logistics is increasingly common, involving analysis in which the possible casual effects of independent variables on categories are accessed via comparison of series of dichotomous outcomes (Anthony, 1990).

According to Anthony (1990) "the result of logistic regression models can be expressed in the form of odd ratios, telling us much change there is in the probability of being employed, receiving university education (or whatever), given a unit change in any other given variable – but holding all other variable in the analysis constants".

Most published account of research using this particular techniques report three statistics for the model. The first of these is the beta parameter estimate (or standardize regression coefficient), which is-crudely speaking- a measure of the size of the effect than an independent variable have been taken into account (Nelson, 1984).

According to Aldridge (1984) "the standard error provides us with means of judging the accuracy of our predictions about the effect in

question". One rule of thumb is that beta should be at least twice the size of the standard error (Aldridge, 1984).

### 2.12.1 LOGISTIC REGRESSION MODEL FOR BINARY RESPONSE VARIABLE

A binary response variable y, denote its two categories by 1 and 0, the generic term success and failure are these two outcome (Liao, 1994). Menard (1994) argued that "regression model for binary response variable described the population proportions and the population proportion of success represents the probability P(y=1) for a randomly selected subject". This probability varies according to the explanatory variables.

Models for binary data ordinarily assume a binomial distribution for the response variable (Liao, 1994). This is natural for binary outcomes. The models are special cases for generalized linear models.

### 2.12.2 LINEAR PROBABILITY MODEL

For a simple explanatory variable, the simple model

$$\text{Logit}\,[P(y=1)] = \alpha + \beta x$$

Implies that the probability of success is a linear function of x (Liao, 1994). This is called the linear probability model

Liao (1994) claimed that this model is often inappropriate, meaning that its probabilities fall below 0 and above 1 for sufficiently small or large x values, whereas probability must fall between 0 and 1. The model may be valid over a restricted range of x values, but it is rarely adequate when the model has several predictors (Collect, 1991).

### 2.12.3 MULTIPLE LOGISTIC REGRESSIONS

An explanation of logic regression begins with an explanation of the logistic function:

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

The "input" is z and the "output" is f(z). The logistic functions is useful because it can take as an input any value from negative infinity, where as the output is confined to value values between 0 and 1. The variable z represents the exposure to some set of risk factors, while f(z) represents the probability of a particular outcome given that set of risk factors. The variable z is a measure of the total contribution of all the risk factors used in the model and is known as the logit also called the log odds. The logistic model formula computes the probability of the selected responses as a function of the predictor variables.

The variable z is usually defined as;

$$Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k$$

Where $\beta_0$ is called the intercept. The intercept is the values of z when the value of all risk factors is zero (i.e. the value of z in someone with no factors). Each of the regression coefficients describes the size of the contribution of that risk factor. A positive regression coefficient means that risk factor increases the probability of the outcome, while a negative regression coefficient means that risk factors decreases the probability of that outcome; a large regression coefficient means that the risk factor strongly influences the probability of that outcome; while a non-zero

regression coefficient means that risk factor has little influence on the probability of that outcome (Hilbe and Joseph, 2009).

Logistic regression is a useful way of describing the relationship between one or more risk factors (e.g. age, sex, etc.) and an outcome, expressed as a probability, that has only two possible values, such as death ("dead" or "not dead").

### 2.12.4 INTERPRETING THE LOGISTIC REGRESSION MODEL

Each of the regression coefficients describes the size of the contribution of the risk factor, a positive regression coefficients means that the risk factor increases the probability of the outcome, while a negative regression coefficient means that risk factor deceases the probability of that outcome; a large regression coefficients means that risk factor strongly influences the probability of that outcome; while a near-zero regression coefficients means that risk factor has little influences on the probability of that outcome (Hilbe and Joseph, 2009). According to Hilbe and Joseph (2009) the intercept is the value of z (dependent variables) when the value of all risk factors is zero. Hilbe and Joseph (2009) argued that the main interpretation of logistic regression results is to fine the significant predicators of Y. However, other things can be done with the result.

## 2.13 REGRESSION ANALYSIS

Regression analysis includes techniques for modeling and analyzing several variables, when the focus is on the relationship between dependent variable and one or more independent variables. More specifically, regression analysis helps us understand how the typical value of the dependent variable changes when any one of the independent variables is

varied while the other independent variables are held fixed. Most commonly, it estimates the conditional expectation of the dependent variable given the independent variables-that is, the average value of the dependent variables when the independent variables are held fixed. In all cases, the estimation target is a function of the independent variable called the regression function.

A large body of techniques for carrying out regression analysis has been developed. Familiar methods such as linear regression and ordinary least square regression are parametric, in that the regression functions is defined in terms of a finite number of unknown parameters that are estimated from the data. Non-parametric regression refers to techniques that allow the regression function to lie in a specified set of function which may be infinite dimensional.

The performance of regression analysis in practice depends on the data generating process, and how it relates to regression approach being used. Since the true form of the data is not known, regression analysis depends to some extent on making assumptions about this process. These assumptions are sometimes (but not always) testable if a large amount of data is available. Regression models for prediction are often useful even when the assumptions are moderately violated, although they may not perform optimally.

### 2.13.1 UNDERLYING ASSUMPTION

- The sample must be representative of the population for the inference prediction

- The error is assumed to be a random variable with a mean of zero conditional on the explanatory variables.
- The variables are error free, if this is not so, modeling may be done using error – variables model techniques.
- The predictions must be linearly independent i.e it must be possible to express any predictor as a linear combination of the others.
- The errors are uncorrelated, that is, the variance-covariance matrix of the errors is diagonal and each non-zero element is the variance of the error.
- The variance of the error is constant across observations (Homoscendasticity). If not, weighted least squares or other methods might be used.

These assumptions imply that the parameters estimates will be unbiased, consistent and efficient in the class of linear unbiased estimators. Many of these assumptions may be relaxed in more advanced treatments.

### 2.13.2 REGRESSION DIAGNOSTICS

Once a regression model has been constructed, it may be important to confirm the goodness of fit of the model and the statistical significance of the estimated parameters. Commonly used checks of goodness of fit include the R-squares, analysis of the pattern of residual and hypothesis testing. Statistical significance can be checked by an F-test of the overall fit, followed by t-test of individual's parameters.

Interpretations of these diagnosis tests rely heavily on the model assumptions.

Although examination of the residual can be used to invalidate a model, the results of a t-test or F-test are sometimes more difficult to interpret if the model's assumptions are violated. For example, if the error term does not have a normal distribution, in small samples the estimated parameters will not follow normal distributions and complicate inference. With relatively large samples, however, central limit theorem can be invoked such that hypothesis testing may proceed asymptotic approximations.

# CHAPTER THREE

# RESEARCH METHODOLOGY

## 3.1 INTRODUCTION

Kinner and Taylor (1981) defined research methodology as the basic plan which guides data collection and its analysis. It is the framework which specifies the types of information to be collected, the sources of data and the data collection procedure. It shows the procedure and techniques adopted before dependable solutions to problems were obtained.

According to Isaac and Michael (1981), the nature of the problem plays a major role in determining the approaches suitable. Kerlinger (1973) opined that a good research design will ensure that the information gathered is consistent with the study objectives, as this will enable the investigator to answer the research questions as objectively, reliably, accurately and economically as possible. See also Osuala (1985).

## 3.2 DATA COLLECTION PROCEDURE

The secondary method of data collection was used. Data was extracted from record cards of Methodist hospital, Ugboju, Benue State, Nigeria over the period 1994 – 2009, on patients who have diseases associated with hypertension.

Five (5) diseases were of interest namely:

i. Thyroid disease
ii. Sleep apnea
iii. Renovascular disease

iv. Cushing syndrome

v. Primary aldosteronism

For each disease, data on three risk factors were collected, namely:

i. Drug dosage for treatment ($x_1$)

ii. Age of patients ($x_2$)

iii. Blood cholesterol level ($x_3$).

## 3.3 PROCEDURE FOR DATA ANALYSIS

The multiple regression analysis is used for prediction of the survival given a set of risk factors.

Let $Y_i$ denote the survival given the $i^{th}$ set of risk factors.

Let $P_i$ denote the survival rate given the $i^{th}$ set of risk factors. Then

$$E(P_i) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots\ldots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots\ldots + \beta_n x_n}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.1)$$

Where $\beta_0$ is the intercept, and $\beta_1$, $\beta_2 \ldots \beta_n$ are the regression coefficients of $x_1$, $x_2$, … $x_n$ respectively.

To estimate these parameters, we used the logit curve, which is obtained by the transformation:

$$E(P_i') = E\left[\log_e\left(\frac{p_i}{1-p_i}\right)\right] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.2)$$

Set $Y_i = E(P_i')$, then

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_n x_{ni} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.3)$$

The above model is implemented in SPSS software.

### 3.3.1 INVESTIGATING ADEQUACY OF THE FITTED MODEL

The modified Pearson Chi-square test, $Q^2$, is used to test goodness-of-fit. It is given by

$$Q^2 = \frac{N-1}{N} \sum_{i=1}^{N} \frac{(P_i - \hat{P}_i)^2}{\hat{P}_i}, \quad \text{..... (3.4)}$$

Where N denotes the number of patients.

The fit is accepted if $Q^2 < \chi^2_{N-r}(\alpha)$, Where

$\alpha$ = level of significance, and

r = number of parameters estimated.

# CHAPTER FOUR

# DATA ANALYSIS AND RESULTS

## 4.1 DATA ANALYSIS

Logistic regression is used to develop different models that can predict the probability of survival of patients suffering from the various diseases considered.

## 4.2 RESULTS

### 4.2.1 THYROID DISEASE

**Table 4.1 Variables Entered for Thyroid Disease**

| Model | Variables Entered | Variables Removed | Method |
|---|---|---|---|
| 1 | $x_1$, $x_3$ | | Enter |

Log odds of the dependent variable: Y

**Tables 4.2 Model Summary for Thyroid Disease**

| Model | R | R square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .831 | .793 | .742 | .18715 |

**Tables 4.3 Analysis of Variance for Thyroid Disease**

| Model | Sum of squares | DF | Mean square | F | Sig |
|---|---|---|---|---|---|
| Regression | 1.546 | 3 | .515 | 17.758 | .000 |
| Residual | .348 | 12 | .029 | | |
| Total | 1.894 | 15 | | | |

a. Predictors: (constant), $x_3$, $x_1$
b. Log odds of the dependent variable: Y

**Table 4.4 Regression Coefficients for Thyroid Disease**

| Model | Regression coefficients | | T | Sig |
|---|---|---|---|---|
| | B | Std. Error | | |
| Constant | 2.170 | .377 | 5.756 | .000 |
| $x_1$ | .004 | .037 | .108 | .026 |
| $x_3$ | -.003 | .046 | -.065 | .008 |

Log odds of the dependent variable: Y

We can observe that all the P- values in table 4.4 are below the $X_{crit}$ of 0.05.The variable ($X_2$) was removed because it P-Value was higher than the $X_{crit}$ 0.05.

The probability of survival from thyroid disease is given by the model

$\hat{p} = \frac{e^y}{1+ e^y}$ , where $y = 2.170 + 0.004x_1 – 0.003x_3$.

The modified Pearson Chi-square test statistic, $Q^2$, gives

$Q^2 = \frac{N-1}{N} \sum_{i=1}^{N} \frac{(P_i-\hat{P}_i)^2}{\hat{P}_i} = 0.6426,$

Where N = Number of patients = 16.

The fit is accepted since $Q^2 < \chi^2_{13}$ (0.05) = 22.36.

### 4.2.2 RENOVASCULRAR DISEASE

**Table 4.5 Variables Entered for Renovascular Disease**

| Model | Variables Entered | Variables Removed | method |
|---|---|---|---|
| 1 | $x_1, x_3$ | | Enter |

Log odds of the dependent Variables: Y

**Table 4.6 Model Summary for Renovascular Disease**

| Model | R | R square | Adjusted R square | Std Error of the Estimate |
|---|---|---|---|---|
| 1 | .818 | .764 | .721 | .10265 |

**Table 4.7 Analysis of variance for Renovascular Disease**

| Model | Sum of square | DF | Mean squares | F | Sig |
|---|---|---|---|---|---|
| Regression | .734 | 3 | .245 | 24.500 | .000 |
| Residual | .123 | 12 | .010 | | |
| Total | .857 | 15 | | | |

a. Predictors: (Constant) $x_1, x_3$

b. Log odds of the dependent Variable: Y

**Table 4.8 Regression coefficients for Renovascular Disease**

| Model | Regression Coefficients | | T | Sig |
|---|---|---|---|---|
| | B | Std Error | | |
| Constant | 1.203 | .289 | 4.163 | .000 |
| $x_1$ | .004 | .031 | .129 | .030 |
| $x_3$ | -.001 | .042 | -.024 | .017 |

Log odds of the dependent Variable: Y

We notice that all the p-values in table 4.8 are below the $\alpha_{crit}$ value of 0.05. The variable ($X_2$) was removed because it P-Value was higher than the $X_{crit}$ 0.05.

The probability of survival from renovascular disease is given by the model

$\hat{p} = \frac{e^y}{1+ e^y}$ , where $y = 1.203 + 0.004x_1 – 0.001x_3$

The modified Pearson Chi-square test statistic, gives

$Q^2 = 0.2563$. The fit is accepted since $Q^2 < \chi^2_{13}$ (0.05) = 22.36.

### 4.2.3 CUSHING SYNDROME

**Table 4.9 Variable Entered for Cushing Syndrome**

| Model | Variables Entered | Variables Removed | Method |
|---|---|---|---|
| 1 | $x_1, x_3$ | | Enter |

Log odds of the dependant Variable: Y

**Table 4.10 Model Summary for Cushing Syndrome**

| Model | R | R square | Adjusted R square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .902 | .853 | .802 | .11967 |

**Table 4.11 Analysis of variance for Cushing Syndrome**

| Model | Sum of squares | DF | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Regression | 1.327 | 3 | .442 | 19.217 | .000 |
| Residual | .279 | 12 | .023 | | |
| Total | 1.606 | 15 | | | |

a. Predictors (Constant ) $x_{1,} x_3$

b. Log odds of the dependent Variable: Y

**Table 4.12 Regression Coefficients for Cushing Syndrome**

| Model | Regression coefficient | | T | Sig |
|---|---|---|---|---|
| | B | Std Error | | |
| Constant | 1.120 | .426 | 2.629 | .000 |
| $x_1$ | .006 | .022 | .273 | .021 |
| $x_3$ | -.008 | .013 | .258 | .037 |

Log odds of the dependent variable: Y

We can notice from table 4.12 that all the P- values are less than the $\alpha_{crit}$ of 0.05. The variable ($X_2$) was removed because it P-Value was higher than the $X_{crit}$ 0.05.

The probability of survival from cushing syndrome disease is given by the model $\hat{p} = \frac{e^y}{1+ e^y}$ , where $y = 1.120 + 0.006x_1 – 0.008x_3$

The modified Pearson Chi-square test statistic, gives

$Q^2 = 0.4265$. The fit is accepted since $Q^2 < \chi^2_{13}$ (0.05) = 22.36.

### 4.2.4 PRIMARY ALDOSTERONISM

**Table 4.13 Variables Entered for Primary Aldosteronism**

| Model | Variable Entered | Variable Removed | Method |
|---|---|---|---|
| 1 | $x_1, x_2$ | | Enter |

Log odds of the dependent Variable: Y

**Table 4.14 Model Summary for Primary Aldosteronism**

| Model | R | R Square | Adjusted R Square | Std Error of the Estimate |
|---|---|---|---|---|
| 1 | .794 | .685 | .632 | .14571 |

**Table 4.15 Analysis of variance for Primary Aldosteronism**

| Model | Sum of Squares | DF | Mean Square | F | Sig |
|---|---|---|---|---|---|
| Regression | 1.240 | 3 | .413 | 20.650 | .000 |
| Residual | .243 | 12 | .020 | | |
| Total | 1.483 | 15 | | | |

a. Predictors: (Constant) $x_1, x_2$

b. Log odds of the dependent Variable Y

**Table 4.16 Regression Coefficients for Primary Aldosteronism**

| Model | Regression coefficient | | T | Sig |
|---|---|---|---|---|
| | B | Std Error | | |
| Constant | 1.160 | .431 | 2.691 | .000 |
| $x_1$ | .001 | .027 | .037 | .039 |
| $x_2$ | .002 | .048 | .042 | .024 |

Log odds of the dependent variable: Y

We can observe that all the P- values in table 4.16 are less than the $\alpha_{crit}$ of 0.05. The variable ($X_3$) was removed because it P-Value was higher than the $X_{crit}$ 0.05.

The probability of survival from Primary Aldosteronism disease is given by the model $\hat{p} = \frac{e^y}{1+e^y}$, where $y = 1.160 + 0.001x_1 + 0.002x_2$

The modified Pearson Chi-square test statistic, gives

$Q^2 = 0.522$. The fit is accepted since $Q^2 < \chi^2_{13}(0.05) = 22.36$.

### 4.2.5 SLEEP APNEA

**Table 4.17 Variables Entered for Sleep Apnea**

| Model | Variables Entered | Variables Removed | Method |
|---|---|---|---|
| 1 | $x_2, x_3$ | | Enter |

Log odds of the dependent Variable: Y

**Table 4.18 Model Summary for Sleep Apnea**

| Model | R | R square | Adjusted R square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .858 | .788 | .701 | .17845 |

**Table 4.19 Analysis of variance for Sleep Apnea**

| Model | Sum of square | DF | Mean square | F | Sig |
|---|---|---|---|---|---|
| Regression | 1.446 | 3 | .482 | 37.656 | .000 |
| Residual | .153 | 12 | .0128 | | |
| Total | 1.599 | 15 | | | |

a. Predictors: (Constant), $x_{2,}\, x_3$

b. Log odds of the dependent Variable; Y

**Table 4.20 Regression Coefficient for Sleep Apnea**

| Model | Regression coefficient | | T | Sig |
|---|---|---|---|---|
| | B | Std Error | | |
| Constant | 3.470 | .645 | 5.382 | .000 |
| $x_2$ | -.007 | .041 | -.171 | .018 |
| $x_3$ | -.005 | .029 | -.172 | .035 |

Log odds of the dependent Variable: Y

We observe that all the P- values in table 4.20 are below the $\alpha_{crit}$ of 0.05. The variable ($X_3$) was removed because it P-Value was higher than the $X_{crit}$ 0.05.

The probability of survival from sleep apnea disease is given by the model

$\hat{p} = \frac{e^y}{1+e^y}$, where $y = 3.470 - 0.007x_2 - 0.005x_3$

The modified Pearson Chi-square test statistic, gives

$Q^2 = 0.6010$. The fit is accepted since $Q^2 < \chi^2_{13}(0.05) = 22.36$.

### 4.2.6 DISCUSSION

For patients with thyroid disease, a careful examination of table 4.2 indicates that $R^2$ = 79.3%. That is, 79.3% of the variation in Y was explained by the model. Considering table 4.4, we observed that for every unit increase in drug level ($x_1$) and cholesterol level ($x_3$), survival rate increases by 0.4% and decreases by 0.3% respectively. Furthermore, ANOVA table on table 4.3 gave a P-value of 0.000 (P<0.05). This shows that there is a linear relationship between Y and $x_1$, $x_3$.

Looking careful at renovascular disease patients, table 4.6 shows that $R^2$ = 76.4%. This means that 76.4% of the variation in the dependent variable Y was explained by the model. From table 4.12, we noticed that for every unit increase in drug level ($x_1$) and cholesterol level ($x_3$), survival rate increases by 0.4% and decrease by 0.1% respectively. The ANOVA table for renovascular disease has P-value of 0.000 (P<0.05). This is an indicator of a linear relationship between Y and $x_1$, $x_3$.

We observed from table 4.10 for Cushing syndrome patients, $R^2$ = 85.3%. This shows that 85.3% of the variation in Y was explained by the model. An examination of table 4.12 shows that for every unit increase in drug level ($x_1$) and cholesterol level ($x_3$), survival rate increases by 0.6% and decreases by 0.8% simultaneously. Also ANOVA table for Cushing syndrome gave a P-value of 0.000 (P<0.05). This means there is a linear relationship between Y and $x_1$, $x_3$.

For patients with Primary Aldosteronism, a close observation of table 4.14 revealed that $R^2$ = 68.5%. This implies that, 68.5% of the variation in Y was explained by the model. It can be observed from table 4.16 that for a

unit increase in drug level ($x_1$) and age ($x_2$), survival rate increases by 0.1% and 0.2% respectively. The ANOVA table on table 4.15 has P-value of 0.000 ($P<0.05$). This indicates a linear relationship between Y and $x_1$, $x_2$.

We observed from table 4.18 for Sleep Apnea patients, that $R^2$ = 78.8%. This means that, 78.8% of the variation in the dependent variable was explained by the model. Looking at table 4.20, we observed that for a unit increase in age ($x_2$) and cholesterol level ($x_3$), survival rate decreases by 0.7% and 0.5% respectively. Also a P-value of 0.000 ($P<0.05$) is obtained from the analysis of variance as shown in table 4.19. This indicates a linear relationship between Y and $x_2$, $x_3$.

A test for goodness-of-fit using the modified Pearson Chi-square shows that all the fitted models are adequate.

# CHAPTER FIVE

# SUMMARY, RECOMMENDATION AND CONCLUSION

## 5.1 SUMMARY

Hypertension is a serious health problem and a major risk factor for cardiovascular disease. Cardiovascular disease is responsible for about 23,800 deaths annually (W.H.O., 2002). High blood pressure in adult has a high impact on the economy and on the quality of life of individuals with implications on resource expenditures.

The data for this study was collected from Methodist hospital Ugboju, Otukpo, Benue state. Linear relationships were developed to explain the probability of survival of patients suffering from five (5) different cardiovascular diseases.

## 5.2 FINDINGS

The study revealed that;

i. Increasing cholesterol level leads to decrease in the survival rate for renovascular, Cushing syndrome, thyroid and sleep apnea diseases.
ii. Increasing drug level will lead to increased in survival rate for thyroid disease, renovascular disease, Primary aldosteronism and Cushing syndrome.
iii. Increasing age for sleep apnea leads to decrease in survival rate.

## 5.3 RECOMMENDATIONS

In view of the findings above, the following suggestions was recommended;

i. Cholesterol level for Cushing syndrome, renovascular disease, thyroid disease and sleep apnea disease patients should be kept at the minimal level as possible in order to increase survival rate.

ii. The host community should be enlightened on the danger of increasing cholesterol level in the blood and food that build up cholesterol level.

## 5.4 CONCLUSION

This study has revealed that increasing cholesterol level has negative impact on survival. We conclude that for more survival of patients, their cholesterol level should be monitored carefully.

## REFERENCES


Akinkugbe, O .O. and Ojo, A .O. (1968). The Systemic Blood Pressure in a Rural Nigerian Population. *Tropical Geographical Medicine,* 20, 347 – 56.

Akinkugbe, O .O. (1996). The Nigerian Hypertension Programme. *Journal of Human Hypertension,* 10, 543 – 6.

Anthony Walsh (1990). Statistics for Social Sciences. *Sage Publications, Thousands Oaks.*

Adridge, S. (1984). Discrete Choice Analysis. *MIT Press, Cambridge, MA.*

*Bar*kis, G .L. (2007). Preserving Renal Function in Adults with Hypertension and Diabetes. *Journal of the American Medical Association,* 45, 15 – 22.

Balogun, M .O. and Ladipo, O.A. (1988). Circulatory Responses to Dynamic Exercise in Healthy Adult Nigerians with or without Parenteral Hypertension Tropical Cardiology, 14, 165 – 8.

Barlow, S.E. and Dietz, W.H. (1998). Obesity Evaluation and Treatment: Expert Committee Recommendations Pediatric. *Journal of Pharmacology and Experimental Therapeutics,* 262, 55 – 68.

Beaumont, J .L. and Buxtorf, J.C. (1993). Incidence of Paroxysmal Hypertension in Smokers. *British Medical Journal,* 4, 10 – 30.

Black, H .R., Elliot, W.J. and Neation, J.D. (2008). Baseline Characteristics and Elderly Blood Pressure Control in the Convince Trial. *Journal of the American Medical Association,* 37, 12 – 18.

Castelli, W. P. (1984). Epidemiology of Coronary Heart Disease. The Framingham Study. *American Journal of Medicine,* 76, 4 – 12.

Chobanian, A.V. and Hill, M. (2007). National Heart, Lung and Blood Institute Workshop on Sodium and Blood Pressure; A Critical Review of Current Scientific Evidence. *Annals of International Medicine,* 71, 599 – 691.

Cooper, R. S. and Rotimi, C.N. (1997). Hypertension in Blacks. *American Journal of Hypertension,* 10, 804 – 12.

Collett, D. (1991). Modeling Binary Data. *Chapman and Hall, London.*

Freeman, V., Praserlt, K., Forrester, T. and Tony, K. (1996). A Comparative Study of Hypertension Prevalence, Awareness, Treatment and Control Rates in St. Lucia, Jamaica and Barbados. *Journal of Hypertension,* 14, 495-501.

Gil, F. and Salles, M.D. (2007). Interpretation and Evaluation of Clinical Guidelines. *The American Heart Association,* 28, 21 – 26.

Hill, M .N. and Miller, W.H. (1996). Compliance Enhancement: A Call for Multidisciplinary Team Approaches Circulation. *Journal of the American Medical Association,* 196, 66 – 106.

Hilbe J. and Joseph M. (2009). Logistic Regression Models. *Chapman and Hall/CRC Press.*

Isaac, S. and Michael, W.B. (1981). Handbook on Research and Evaluation, 2nd Edition, Edit Publishers, California, San Diego.

Izzo, J .L., Levy, D. and Black, H.R. (2006). Clinical Advisory Statement: Importance of Systolic Blood Pressure in Older Americans. *Journal of the American Medical Association,* 3, 15 – 20.

Kaufman, J. S., Owoaje, E.E., Rotimi, C.N. and Victoria, I.K. (1999). Blood Pressure Change in Africa: Case Study from Nigeria. Human Biology, 71, 641 – 5.

Kerlinger, F .N. (1973). Foundations of Behavioral Research, 2nd Edition, America, Holt, Rinehart and Winston Inc., U.S.A.

Kinner, T .C. and Taylor, J.R. (1981). Marketing Research – An Applied Approach, 3rd Edition, McGraw Hill Inc., New York.

Lawal, S. O. and Falase, A.O. (1998). The Effect of Hypertension on the Heart of Adult Nigerians. Tropical Cardiology, 14, 153 – 8.

Liao, T .F.(1994). Interpreting Probability Models: Logits, Probit and Other Generalized Linear Models. *Sage University Paper series on Quantitative Application in the Social Sciences. Sage Beverly Hills, CA*

Murray, C. J. and Lopez, A.D. (1997). Mortality by Cause for Eight Regions of the World. Global Burden of Disease. Lancet, 349, 1269 – 76.

Menard, S (1995). Applied Logistic Regression Analysis. *Sage University Paper Series on Quantitative Application in Social Sciences. Sage, Beverle Hills, CA*

Nwankwor, M.U., Adams, F., Campbell, L.L. and Ukoli, F.A. (1990). Blood Pressure in Nigeria College Males. *Journal of Human Hypertension,* 4, 72 – 3.

Nelson, F. (1984). Linear Probability, Logit and Probit Models. *Prentice Hall.*

Osuala, E. C. (1985). Introduction to Research Methodology, Lagos: Africana Publishing Limited, Onitsha.

Post, K .F. (2005). The Effect of Chronic Coffee Drinking on Blood Pressure: a Meta Analysis of Controlled Clinical Trials. *American Journal of Hypertension*, 33, 647 – 652.

Rose, B .D. (2007). Principal Result of the Controlled Onset Verapamil Investigation of Cardiovascular End Points Trial. *Journal of the American Medical Association,* 28, 207 – 222.

Sheps, S. G. and Roccella, E.J. (2004). Reflection of the Six Report of the Joint National Committee on Prevalence, Detection, Evaluation and Treatment of High Blood Pressure. *British Medical Journal,* 7, 18 - 27.

Tepper, D. (2001). Frontiers in Congestive Heart Failure: Effect of Moto Procol CR/XL in Chronic Heart Failure. *British Medical Journal,* 1, 1 – 20.

Vassan, R.S., Larson, M.G. and Leip, E.P. (2002). Assessment of Frequency of Progression to Hypertension in Non-Hypertensive Participant in the Framingham Heart Study. *American Journal of Cardiology,* 37,25-45.

Vollmen, W .M., Sacks, F.M. and Ardj, Y. (2007). The Effect of Diet and Sodium Intake on Blood Pressure. *Journal of the American Medical Association*, 10, 15 – 19.

Whelton, P. K., Hej, K. and Apple, U. (2002). Primary Premonition Hypertension; Clinical and Public Health Advisory from the National High Blood Pressure Education Progression Association, 10, 71 – 91.

Wolf, K., Maier, K., Cooper, R.S. and Banegas, J.R. (2003). Hypertension Prevalence and Blood Pressure Levels in Six European Countries,

Canada and the United States. *Journal of American Medical Association*, 289, 2363 – 9.

World Health Organization (2002). The World Health Report, Reducing Risk and Promoting Healthy Life. Geneva: World Health Organization.