

# Modeling HIV/AIDS Prevalence in Wukari Local Government Area of Taraba State-Nigeria.

James Friday E<sup>1</sup>, Ali Michael I<sup>1</sup> and Edache Elizabeth O<sup>2</sup>

<sup>1</sup> Department of Mathematics and Statistics, Federal University, Wukari, Taraba, State, Nigeria.

<sup>2</sup> Department of Statistics, University of Ibadan, Ibadan, Oyo State, Nigeria.

## ARTICLE INFO

### Article history:

Received: 25 May 2018;

Received in revised form:

25 June 2018;

Accepted: 7 July 2018;

### Keywords

HIV/AIDS,

Logistic regression,

Mortality,

Prevalence,

Risk factors.

## ABSTRACT

The logistic regression model has had a remarkable growth in the field of statistics. It is particularly useful for predicting continuous and categorical explanatory variables using categorical dependent variable. This work studies 111 cases of HIV/AIDS infected patients and fits a binary logistic model. The results show that presence of tuberculosis, CD4 cell count and Regimens are significant risk factors of mortality of HIV/AIDS patients, obtained from the records Department of the General Hospital Wukari, Taraba State of Nigeria.

© 2018 Elixir All rights reserved.

## I. INTRODUCTION

In 1989, a new syndrome, the Acquired Immune Deficiency Syndrome (AIDS), was first recognized among homosexual men in the United States of America. [UNAIDS, 2003]. By 1983, the etiological agent, the human immunodeficiency virus (HIV) had been identified. By the mid 1980's it became clear that the virus had spread, largely unnoticed, throughout most of the world.

The HIV/AIDS epidemic has brought large increases in morbidity and mortality to many countries in sub-Saharan Africa. For some countries, the epidemic has eliminated the large gains in life expectancy that took place between 1950 and 1990 [ref no]. HIV is a virus that gradually attacks the immune system, which is our body's natural defense against illness. If a person becomes infected with the virus, they find it harder to fight off infections and diseases. The virus destroys a type of white blood cell called T-helper cell and makes copies of itself inside them [ref no]. AIDS is a syndrome caused by HIV. It is the last stage of HIV infection when the body can no longer defend itself and may develop various diseases and infections and if left untreated, leads to death [AVERT, 2017]. Laboratory measurements taken for each of the patients such as CD4 cell counts, the tuberculosis status of the patients, body mass index (obtained from their weights and heights) etc., serve as prognostic indicators of mortality among HIV/AIDS patients. Other factors may also impact the outcome.

### HIV/AIDS Prevalence and Impacts on Community of Infected Persons in Nigeria

The first case of AIDS in Nigeria was reported in 1986 thus establishing the presence of the epidemic in the country. Consequently, and in line with WHO guidelines, the government adopted ANC sentinel surveillance as the system for assessing the epidemic. Between 1991 and 2001, Nigeria witnessed an increase in the prevalence of HIV in the country.

The first HIV Sentinel Survey in 1991 showed a prevalence of 1.8%. Subsequent sentinel surveys produced prevalence of 3.8% (1993), 4.5% (1996), 5.4% (1999), 5.8% (2001), 5.0% (2003), 4.4% (2005), 4.6% (2008) and 4.1% (2010). The National HIV/AIDS and Reproductive Health Survey (NARHS) was adopted in 2003 to provide information on key HIV/AIDS and Reproductive Health knowledge and behavior related issues. In 2007, the scope was expanded to include estimation of HIV prevalence in the country. A more comprehensive survey was conducted in 2012, (NARHS plus II 2012) which showed a decline to 3.4% in HIV prevalence, indicating a reversal of the epidemic in the country, compared to the 2007 figure of 3.6%. Nigeria's epidemic is generalized, with wide variation in prevalence within the country.

Of 1.8million HIV/AIDS positive persons who require treatment, Nigeria is only able to treat about 750, 000 from the support of the Global Fund and other partners. HIV/AIDS in Nigeria is funded mainly by donor money, accounting for over 90% of funds available for the intervention in the country. Common distortion to the treatment program being experienced include: drug stock-outs, introduction of various regimes of fee to access treatment by a largely poor population, several people have dropped out of the treatment program due to this reality and now, with this suspension, a climate of uncertainty hovers around the future of HIV treatment in Nigeria. Chances of getting infected with HIV/AIDS in Rivers state, Nigeria's oil-rich city is pretty high as it ranks number one in the country. There's loads of money in Rivers, so is the vulnerability of its citizens and residents to HIV/AIDS. According to the HIV/AIDS prevalence ranking by states published by National Agency for the Control of Aids (NACA) after the 2013 National Aids Reproductive Health Survey, Rivers sits pretty at the top 4.7 per cent more than number two, Taraba state.

The ranking also came as a result of test carried out on persons living with HIV/AIDS. It showed that Rivers with 15.2 per cent topped the new prevalence chart, as Taraba with 10.5 per cent is second, while Kaduna has 9.2 per cent to be on the third. Nasarawa, FCT, Akwalbom, Sokoto and Oyo are on the fourth, fifth, sixth, seventh, eighth and ninth spots respectively on the prevalence rate chart with 8.1, 7.5, 6.5, 6.4, 5.6 and 5.6 per cent respectively. On number 10 is Yobe with 5.3%. Here’s how other states rank: Cross River – 4.4%, Ondo – 4.3%, Gombe – 3.4%, Abia – 3.3%, Bayelsa – 2.7%, Osun – 2.6%, Imo – 2.5%, Borno – 2.4%, Plateau – 2.3%, Lagos – 2.2%, Jigawa – 2.1%, Adamawa – 1.9%, Kwara – 1.4%, Kogi – 1.4%, Kano – 1.3%, Enugu – 1.3%, Niger – 1.2%, Anambra – 1.2%, Ebonyi – 0.9%, Kebbi – 0.8%, Edo – 0.8%, Delta – 0.7%, Ogun – 0.6%, Bauchi – 0.6%, Zamfara – 0.4%, Ekiti – 0.2%.

**HIV/AIDS Prevalence in Taraba**

At a time HIV/AIDS prevalence rate is slowing down to a manageable level in other states of the federation, the rate is on the rise in Taraba state. With a prevalence rate of 10.5 percent since 2012, the state ranks top in the north east and second highest in the country

NACA takes a look at the factors aiding the spread in the state and government’s efforts in fighting the scourge. After the identification of the Human immune Virus, HIV by scientists in the United States of America and France in 1983, the first two cases of HIV/AIDS were diagnosed in Nigeria in 1985, few years after Taraba state was carved out of the defunct Gongola state in 1991.

The documented national prevalence rate of HIV then in the state was 1.8 percent in 1991. Perhaps with little or no attention given to the disease, the rate catapulted to 5.8 percent in 2001.

The figure was arrived at based on the National Sentinel Survey report of pregnant women attending antenatal care in the state only, excluding perhaps other women, young men and adults who were not captured in the survey.

The report as obtained by National Mirror indicates that, the rate is not only above the national average, but is also the highest in the entire north east region and second highest in the entire country since 2012. Between January and December 2015, out of 42,716 people who were tested, 5,867 people tested positive to the virus.

This research work studies the application of logistic regression analysis to determine the significant risk factor(s) of HIV/AIDS mortality in Wukari Local Government Area of Taraba State.

**II MATERIALS AND METHOD**

This study was carried out in Wukari County, the second largest local government area in Taraba state – Nigeria. The data was obtained from the records department of the central general hospital, the only medical center highly equipped with state-of-the-art medical facilities – thus a major and most probably the most reliable source of HIV/AIDS data in the local government.. The data consists of records of 111 patients infected with HIV/AIDS from 2012 to 2015.

**Variables in the Study**

The dependent or response variable is the patients’ status which is binary or dichotomous – Dead or Alive. The explanatory variables are simply the predictor variables used in the logistic regression model. These are the variables which influence the survival/death of patients and are given below:

**Table 1. Variables in the Study.**

Variable	Symbol	Reference Category
Age	$x_1$	
Gender	Male----- $x_{2m}$ Female---- $x_{2f}$	$x_{2f}$
Weight	$x_3$	
CD4 Count	$x_4$	
Tuberculosis status	Present----- $x_{5a}$ Absent----- $x_{5b}$	$x_{5b}$
WHO Clinical Stage	Stage 1---- $x_{6stg1}$ Stage 2---- $x_{6stg2}$ Stage 3---- $x_{6stg3}$	$x_{6stg1}$
Regimen Type	1E--- $x_{7R1E}$ 1B--- $x_{7R1B}$ 4B--- $x_{7R4B}$	$x_{7R1E}$

**The Logistic Regression Model**

Logistic regression has a wide range of application in medical and biomedical research mainly to formulate models sorting the factors that might determine whether or not an outcome happens. The distinguishing feature of logistic regression model is that the outcome variable is binary or dichotomous. Usually, patients’ data would be used to establish which attributions are influential in predicting the given outcome.

This model give estimated probabilities that lie within the range of 0 to 1. It is for this important reason that the logistic regression model is more suitable to use as a means of modeling probabilities.

Suppose that we have n binary observations  $Y_1, Y_2, Y_3 \dots, Y_n$  in which  $Y_j = 0$  or  $1, j = 1, 2, 3, \dots, n$  such that:

$$\Pr(Y_j = 1) = P_j \text{ and } \Pr(Y_j = 0) = 1 - P_j$$

$$E(Y_j) = 0. \Pr(Y_j = 0) + 1. \Pr(Y_j = 1)$$

=  $\Pr(Y_j = 1) = P_j$  which is the probability of success corresponding to the jth response or variable.

For each  $j = 1, 2, \dots, n$ , there is a row vector  $X_i = (X_1, X_2, \dots, X_k)$  of explanatory variables. The idea here is to find an equation that is related to the probability of success of the jth observation. i.e, K explanatory variables that we think may influence  $P_j$

The logistic regression model is usually formulated by relating the probability of success of the jth observation i.e  $P_j$  conditional on a vector  $X_i$  of explanatory variables through the logistic distribution functional form. Thus:

$$P_j = \frac{e^{\alpha + \sum_1^k \beta_i x_i}}{1 + e^{\alpha + \sum_1^k \beta_i x_i}}$$

where  $X_i = (X_1, X_2, \dots, X_k)$

$$\beta_i = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \beta_k \end{pmatrix}$$

$$\text{and } 1 - P_j = \Pr(Y_j = 0/X_j)$$

$$= 1 - \frac{e^{\alpha + \sum_1^k \beta_i x_i}}{1 + e^{\alpha + \sum_1^k \beta_i x_i}}$$

$$= \frac{1}{1 + e^{\alpha + \sum_1^k \beta_i x_i}}$$

The  $\beta_i$ ’s are unknown regression coefficients or parameters that are to be estimated from the data and  $X_i$  denotes the set of values of the K explanatory variables associated with the jth observation.

The linear logistic model for the dependence of  $P_j$  on the values of the K explanatory variables associated with the jth observation is :

$$\begin{aligned} \frac{P_j}{1 - P_j} &= \frac{e^{\alpha + \sum_1^k \beta_i X_i}}{1 + e^{\alpha + \sum_1^k \beta_i X_i}} \div \frac{1}{1 + e^{\alpha + \sum_1^k \beta_i X_i}} \\ &= \frac{e^{\alpha + \sum_1^k \beta_i X_i}}{1 + e^{\alpha + \sum_1^k \beta_i X_i}} \times \frac{1 + e^{\alpha + \sum_1^k \beta_i X_i}}{1} \\ &= \frac{P_j}{1 - P_j} = e^{\alpha + \sum_1^k \beta_i X_i} \end{aligned}$$

When a linear logistic model is fitted to explore the relationship between a binary response variable and one or more predictor variables as in the case of this study, the model is referred to as a logistic regression model.

When the response variable has j mutually exclusive and exhaustive categories denoted by  $J = 1, 2, 3, \dots, K$  and jth category is taken as the reference category for the response variable. The choice of the reference category is arbitrary because the ordering of the categories is also arbitrary.

**The odds and the Logit of  $P_j$**

The logit of  $P_j$  is derived from the logistic function.

$$\begin{aligned} P_j &= \frac{e^{\alpha + \sum_1^k \beta_i X_i}}{1 + e^{\alpha + \sum_1^k \beta_i X_i}} \\ 1 - P_j &= \frac{1}{1 + e^{\alpha + \sum_1^k \beta_i X_i}} \\ \frac{P_j}{1 - P_j} &= e^{\alpha + \sum_1^k \beta_i X_i} \end{aligned}$$

Taking the natural logarithm (to base e) of both sides, we obtain:

$$\log_e \left( \frac{P_j}{1 - P_j} \right) = \alpha + \sum_1^k \beta_i X_i$$

The method is based on the logistic transformation or logit proportion, namely:

$$\text{logit}(P) = \frac{P_j}{1 - P_j}$$

The odds ratio is a measure of association. The odds of success are:

$$\theta = \frac{P_j}{1 - P_j}$$

The quantity  $\frac{P_j}{1 - P_j}$  is called the odds ratio denoted as  $\theta$  and the

quantity  $\log_e \left( \frac{P_j}{1 - P_j} \right)$  is called the log odds or the logit of  $P_j$

**Fitting the Linear Logistic Regression Model to Binary Data**

Let  $Y_j$  be a binary response variable (that is, patients' status—dead or alive) in which

$Y_j = 0$  (dead) and  $Y_j = 1$  (alive) for all  $j = 1, 2, 3, \dots, K$  depending on the seven explanatory variables described in section 3.2 above. If the probability of success is:

$P_j = \text{Pr}(Y = 1/j)$  means probability of  $Y = 1$  given  $j$

Hence, probability of failure = 1 – probability of success.

$$1 - P_j = q_j$$

$$q_j = \text{Pr}(Y = 0/j)$$

$$q_j = 1 - \text{Pr}(Y = 1/j)$$

From linear logistic regression model,

$$P_j = \frac{e^{\alpha + \sum_1^k \beta_i X_i}}{1 + e^{\alpha + \sum_1^k \beta_i X_i}}$$

The logistic regression function is the logit transformation of P where:

$$\begin{aligned} \text{logit}(P) &= \ln \left( \frac{P_j}{1 - P_j} \right) \\ &= \alpha + \beta_1 x_2 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k \\ &= \alpha + \sum_1^k \beta_i X_i \end{aligned}$$

where  $\alpha$  is the constant of the equation and  $\beta_i =$  the coefficient of the predictor variable i.

Using the logistic transformation in this way overcomes the problems that might arise if p was modeled directly as a linear function of the explanatory variables, in particular, it avoids fitted probabilities outside the range (0,1). The parameters in the model can be estimated by maximum likelihood estimation.

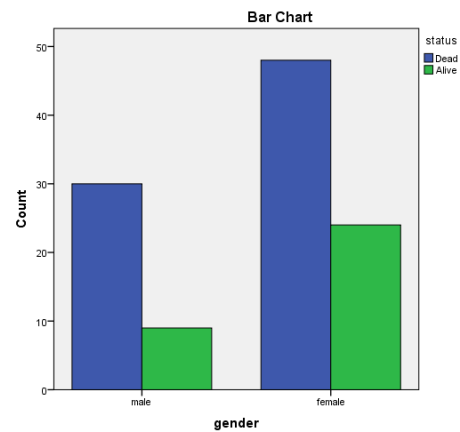
**III ANALYSIS AND RESULTS**

The data analysis was performed using SPSS 22.0 IBM version, and the following results obtained.

**Descriptive Statistics and Cross Tabulations of Response Variable and Categorical Variables**

**Table 2. Gender versus Status of Patients.**

		status		Total
		Dead	Alive	
gender	male	30	9	39
	female	48	24	72
Total		78	33	111

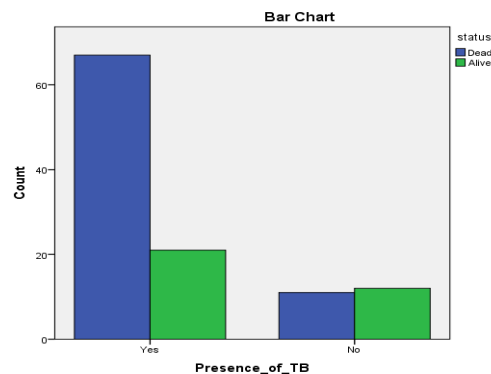


**Figure. 1**

There were 39 males and 72 females. 30 of the males and 48 of the females are dead while 9 males and 24 females remain alive as at 2015.

**Table 3: Presence of Tuberculosis versus Status of Patients.**

		status		Total
		Dead	Alive	
Presence_of_TB	Yes	67	21	88
	No	11	12	23
Total		78	33	111

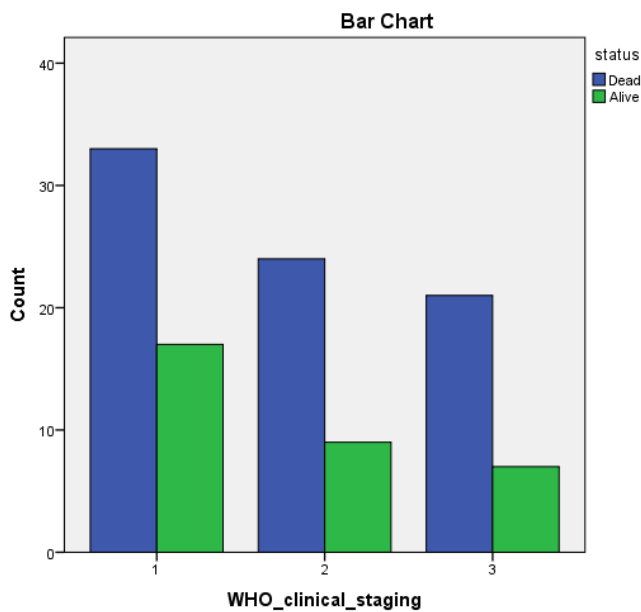


**Figure. 2**

88 of the patients are positive to tuberculosis while 23 had no tuberculosis. 67 of those with presence of tuberculosis and 11 of those with no tuberculosis died, leaving 21 of the tuberculosis and 12 of the tuberculosis-free patients alive.

**Table 4. WHO\_clinical\_staging \* status Cross tabulation.**

	status		Total
	Dead	Alive	
WHO_clinical_stage	1	33	50
	2	24	33
	3	21	28
Total	78	33	111

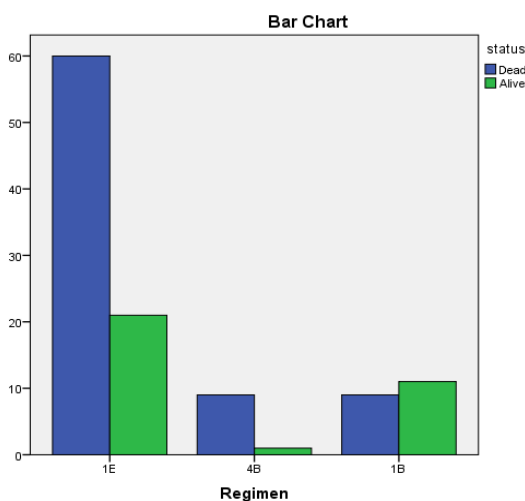


**Figure. 3**

HIV/AIDS infected patients are classified by the World Health Organization into clinical stages. From the analysis, stages 1, 2 and 3 had 50, 33, and 28 patients respectively. Of these, 33 classified into stage 1 died while 17 were alive; 24 grouped into stage 2 died leaving 9 alive and 21 belonging to stage 3 died whereas 7 patients remained alive.

**Table 5. Regimen \* status Crosstabulation.**

	Regimen	status		Total
		Dead	Alive	
Regimen	1E	60	21	81
	4B	9	1	10
	1B	9	11	20
Total		78	33	111



**Figure. 4**

There are three classes of regimens – 1E, 4B and 1B. There were 81 patients on regimen 1E, 10 on regimen 4B and 20 on 1B. However, there were 60, 9 and another 9 dead patients on these regimens while 21, 1 and 11 patients on these respective regimens are alive.

**Table 6. Descriptive Statistics of the Continuous Explanatory Variables.**

	N	Range	Minimum	Maximum	Mean	Std. Deviation
CD4_count	111	1006	19	1025	255.21	200.577
Weight	111	70.80	8.00	78.80	48.7288	14.69463
Age	111	63	2	64	29.45	11.506
Valid N	111					

There was an average CD4 Cell counts of 255.21 cells, 48.7288 kilograms of the patients' weight and a mean age of 29.45 years of the patients studied.

**LOGISTIC REGRESSION ANALYSIS AND RESULTS**

**Case Processing Summary.**

Unweighted Cases <sup>a</sup>		N	Percent
Selected Cases	Included in Analysis	111	100.0
	Missing Cases	0	.0
	Total	111	100.0
Unselected Cases		0	.0
Total		111	100.0

**Dependent Variable Encoding.**

Original Value	Internal Value
Dead	0
Alive	1

**Omnibus Tests of Model Coefficients.**

	Chi-square	df	Sig.
Step 1 Step	22.801	9	.007
Block	22.801	9	.007
Model	22.801	9	.007

**Model Summary**

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	112.299 <sup>a</sup>	.186	.264

**Classification Table<sup>a</sup>**

	Observed status	Predicted status		Percentage Correct
		Dead	Alive	
Step 1 status	Dead	70	8	89.7
	Alive	19	14	42.4
Overall Percentage				75.7

	B	S.E.	Wald	df	Sig.	Exp (B)
gender()	-.242	.555	.191	1	.662	.785
CD4_cell_count	-.003	.002	4.041	1	.044	.997
Weight	.009	.020	.181	1	.671	1.009
Presence_of_TB()	-1.483	.538	7.585	1	.006	.227
WHO_clinical_stage			1.634	2	.442	
WHO_clinical_stage(1)	.787	.628	1.573	1	.210	2.197
WHO_clinical_stage(2)	.688	.701	.964	1	.326	1.991
Age	-.018	.028	.411	1	.521	.982
Regimen			9.379	2	.009	
Regimen(1)	-1.720	.614	7.832	1	.005	.179
Regimen(2)	-2.900	1.381	4.409	1	.036	.055
Constant	2.024	1.222	2.741	1	.098	7.565

- The classification table indicates the number of dead patients and patients that are alive which are observed in the dependent variable
- The null model has predicted that all the patients are dead, and the overall percentage of cases for which the dependent variable was correctly predicted given the model is 70.3%.
- The Wald Chi-Square test that the null hypothesis that the constant equals zero.

Under the full model (Block 1)

- The Omnibus Test is a Chi-Square goodness of fit test. It has the null hypothesis that intercept and all coefficients are zero. Based on our result, our intercept and coefficients are not zeros, thus we reject the null hypothesis. The test shows that the model parameters are significantly different from zero.
- The Cox and Snell R square and Nagelkerke R square model shows that the model is good, though not great.
- The classification table shows the predicted values of the dependent variable based on the full logistic regression model. This table shows how many cases are correctly predicted (70 patients were observed to be dead and are correctly predicted to be dead, 14 patients were observed to be alive and are correctly predicted to be alive, 8 patients were observed to be dead but are predicted to be alive and 19 patients were observed to be alive but are predicted to be dead. The overall percent of patients correctly predicted by the model is 75.7%.
- The Column B shows the values for the logistic regression equation for predicting the dependent variable from the independent variable. They are in log odds units. The prediction equation is:

$$\begin{aligned} \text{logit}(P) &= \ln\left(\frac{P_j}{1 - P_j}\right) \\ &= \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k \\ \text{logit}(P) &= 2.024 + (-0.018)x_1 + (-0.242)x_{2m} \\ &\quad + 0.009x_3 + (-0.003)x_4 \\ &\quad + (-1.483)x_{5a} + 0.787x_{6stg2} \\ &\quad + 0.688x_{6stg3} + (-1.720)x_{7R1B} \\ &\quad + (-2.9)x_{7R4B} \end{aligned}$$

- The output indicates that CD4 Cell count, the presence of Tuberculosis and Regimen of patients are significant predictors of mortality.
- The Exp(B) column shows the odd ratios. The model depicts that the male patients are 0.785 times more likely to die than the female patients.
- Patients with Tuberculosis are 1.009 times more likely to die than those without tuberculosis. Thus, they have an increased probability of 0.009 chance of dying than those without tuberculosis.
- The patients on clinical stages 2 and 3 are 2.197 and 1.991 times more likely to die than patients on stage 1. Although, result shows that it is not a significant predictor of mortality.
- The HIV/AIDS patients on Regimen 1B and 4B are 0.179 and 0.055 times more likely to die than those on Regimen 1E

#### IV CONCLUSION AND RECOMMENDATION

The research leaves us with the following conclusions:

1. The logistic regression model provides an adequate means of identifying risk factors of mortality from diseases and

could be applied to other areas such as social sciences and engineering.

2. The major prognostic factors of mortality of HIV/AIDS patients in Wukari are identified as Tuberculosis presence, Regimen of patients and the CD4 Cell counts of the patients

From the work done, the logistic model is strongly recommended for use in cases where the response variable is categorical.

#### REFERENCES

1. "HIV and Its Transmission". Centers for Disease Control and Prevention. 2003. Archived from the original on February 4, 2005. Retrieved May 23, 2006.
2. "HIV/AIDS Fact sheet N°360". WHO. November 2015. Retrieved 11 February 2016.
3. Fidelis M (2007). 'Why Taraba Leads North-East in HIV/Aids Prevalence'. Daily Trust
4. Madhu B, Ashok N.C and Balasubramaniam (2014): A Multinomial Logistic Regression Analysis to study the influence of residence and socio-economic status on Breast Cancer incidences in Southern Karnataka, *International Journal of Mathematics and Statistics Invention*, Vol 2(5), pp 01 – 08
5. Ogunsaki R.E and Adebayo A.B (2014): Performance of Logistic Regression in Tuberculosis Data, *International Journal of Scientific and Research Publications*, Vol 4(1), ISSN 2250-3153
6. Shi-Qiang Jiang and Qing Liu (2009): Application of Logistic Regression in Combination with multiple diagnostic tests for auxiliary diagnosis of nasopharyngeal Carcinoma, *Chinese Journal of Cancer*, Vol 28(2), Sun Yat-Sen University Cancer Centre, pp 177 – 180.
7. Trauma Score and the Injury Severity Score". *The Journal of trauma* 27 (4): 370–378.
8. UNAIDS (2011): *Global HIV/AIDS response - Epidemic update and health sector progress toward universal access - Progress report 2011*. UNAIDS, Geneva; 2011
9. UNAIDS (May 18, 2012). "The quest for an HIV vaccine".
10. UNAIDS(2012): UNAIDS Global Report on the Global AIDS epidemic 2012,
11. UNAIDS, WHO (December 2007). "2007 AIDS epidemic update" (PDF). Retrieved March 12, 2008.url: <http://www.avart.org> (accessed on 9<sup>th</sup> July, 2016)
12. Weiss, R.A (1993): "How does HIV cause AIDS?" *Science*, 260(5112), 12739, WHO. 2015. p. 13. ISBN 9789241509565.
13. Yusuff H, Mohammad N, Ngah U. K and Yahaya A.S (2012), Breast Cancer Analysis using Logistic Regression, *International Journal of Research and Reviews*, Vol 10(1), pp 14 – 22