International Journal of Statistics and Probability; Vol. 12, No. 6; November 2023
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Modelling Factors that Predict Differences in Childhood Mortality in Lagos Communities Using Prognostic Logistic and Poisson Regression Models

W. Akanji[1] R. Kareem[2], J. A. Akinyemi[2] and M. I. Ekum[2]

[1]Department of Computer Science, Lagos State University of Science and Technology, Ikorodu, Lagos, Nigeria

[2]Department of Mathematical Sciences, Lagos State University of Science and Technology, Ikorodu, Lagos, Nigeria

Correspondence: J. A. Akinyemi, Address: Department of Mathematical Science, Lagos State University of Science and Technology, Ikorodu, Lagos. Nigeria

Received: April 18, 2020 Accepted: June 1, 2020 Online Published: December 26, 2023

doi:10.5539/ijsp.v12n6p1 URL: https://doi.org/10.5539/ijsp.v12n6p1

Abstract

Lagos State is a city with one of the largest concentration of people in the world with heterogenous behaviour and cultural beliefs. There are different prognostic models in the medical sciences, yet their real life application, especially to childhood mortality is limited. There are variations in childhood mortality rate across different communities in Lagos State. Childhood mortality is a subject of interest to World Health Organization (WHO) and one of the major Millennium Development Goals. In 2014, the Special Adviser to the Lagos State Governor on Public Health, Dr. Yewande Adeshina said that under-5 and infant mortality rates in Lagos state have reduced due to various health interventions being implemented in the State. However, the truth of the matter is that childhood mortality is still high and this is an indication that we still have lots of work to do in this regard. In this paper, prognostic models were used in modelling factors that predict the differences in childhood mortality in Lagos communities. Six models, two each from logistic regression, linear regression and Poisson regression models were used. Primary data were collected from mothers that fall in the age bracket (15-49), who reside in any of the 5 divisions of Lagos State, namely Ikorodu, Badagry, Lagos Mainland/Ikeja, Lagos Island and Epe. Five variables were identified as covariates. The prognostic multi-variable models were employed. The binary logistic regression model with 5 covariates was selected as the best model for the binary response variable, while the Poisson regression model with 4 covariates was selected as the best model for the count response variable. At the end of the research, Ikorodu, Badagry and Epe communities have higher than expected childhood mortality rates. Also, we estimated childhood mortality rate in Lagos State and measured the variations in childhood mortality across Lagos communities. The factors that predict these variations were detected and control measures were recommended to reduce the difference in childhood mortality rate in Lagos State.

Keywords: Binary logistic regression, Childhood mortality, linear regression, Prognostic models, Poisson regression, Univariate data

1. Introduction

Lagos state popularly called "Centre of excellence" is the commercial city of Nigeria and of Africa. The state was created on 27th May, 1967 through decree 14 by the Federal Government (Omariba, Walter,Roderic and Fernando, 2007). Lagos state lies approximately between longitude 2042 East and 3042 East and latitude 6022 North and 6052 North. It is bounded in the South by the Guinea Coast of the 180km Atlantic Coastline, in the West by the Republic of Benin and in the North and East by Ogun State. Being the smallest state in terms of land mass in Nigeria which is 3,577 Km^2 and about 22 percent of it occupied by water. It is occupying a frontline position amongst the thirty-six states of the federation (Ojikutu, 2008). According to the World population review in 2016, Lagos state has a population estimated at 21 million makes it the largest city in Africa and one of the most populous in the world (World Population Review, 2016).

This development as observed by Oke, Adedokun, Soretire and Faweya (2001) that Lagos state occupies about 6.2 percent of the total population of Nigeria and has the most heterogeneous concentration and cultural groups such as Yoruba, Igbo, Hausa, Ibibio, Efik, Nupe,and many other Nigerians and none-Nigerians living side by side.Though, majority of the inhabitants are Yorubas. In Lagos state, child mortality which is on the decrease can be attributed to various causes: from management of childhood diseases by parents, age of mothers, preceding birth interval, poverty, to access to health care delivery and so on. In view of this, Feyisetan, Asa and Ebigbola (1997) opined that substantial blames placed on the inadequate availability of health care services, the socio-cultural determinants of people should also be considered. The knowledge of Measles and Diarrhea is pertinent in the understanding of the role of cultural beliefs in health seeking among mothers attributing the Measles attack to a variety of causes which has no link to the virus. According to the belief, Measles attack can be traditionally attributed to the break of family taboos, or an attack from witches or enemies. Most mothers attribute Diarrhea as a further growth manifestation (Ogunjuyigbe, 2004).

However, this claim was supported by Asakitipi (2007), who explained that Diarrhea alone killed about three million under 5 children annually which among the Yoruba mothers associated to attack by witches and enemies. Jegede, Ajala, Adejumo and Osunwale (2006) explained that some of these traditional practices lead to ill health in children and generally has negative effect in health of these children.

With the advent of Pentecostal Christian religion, this old belief is been replaced by those that are more refined but equally devastating to the health of children. Mazumdar (2000) explained that the relationship between culture and health has been established and that research has discovered that some of these traditional practices lead to ill health in children and generally have negative effect on health. They further highlighted the effect of forced feeding, a common practice in Yoruba land to the general well-being of children. According to them, health problems such as diarrhea, cholera, respiratory tract infection, kwashiorkor and so on have been found to be associated with forced feeding.

Many years of studies on childhood mortality have produced diverse findings on the causes and determinants of under-five deaths. While many studies on childhood mortality in Nigeria have established that individual level factors such as maternal education and other socio economic status are important predictors of childhood mortality, similar studies on the effects of neighbourhood contexts on child survival are minimal in Nigeria. Whereas, literature established that living in a deprived neighbourhood is associated with poor health outcomes of individuals (Harttgen & Misselhorn, 2006; Omariba et al., 2007; Sastry, 1997; Zanini, Moraes, Ginghani & Riboldi, 2009). For instance, Zanini et al. (2009) found that about half of the variability in infant mortality rates in Brazil was due largely to community-level characteristics. They argued that neighbourhood characteristics can aggravate or alleviate mortality risks of individuals depending on the neighbourhood contexts where individuals reside.

In 2014, the then special adviser to the Lagos state Governor on public health Dr.Yewande Adeshina said that under-5 and infant mortality rates in Lagos state have reduced due to various health interventions being implemented in the State. In the same year, 2014, she commenced the first round of Maternal, Newborn and Child Health (MNCH) week in Lagos and that various health interventions being implemented in the State have in no small measure contributed immensely to the success achieved in improving the Maternal and Child Health Indices in the State. According to her, under-5 and Infant Mortality Rates in Lagos state have reduced from 157 per 1,000 live births and 75 per 1,000 live births in 2008 NDHS to 65 per 1000 live births and 45 per 1000 live births respectively as reported in the 2011 Multiple Indicator Cluster Survey (Aigbe, 2012). It was mentioned then in 2014 that these figures are still high and this is an indication that we still have lots of work to do in this regard before the year 2015. The full implementation of the IMNCH strategy nation-wide could prevent up to 72 percent of neonatal deaths, more than 70 percent of under-5 deaths and two-thirds or 62 percent of maternal death (Adeshina, 2014).

It should be noted that infant and under-5 mortality are used interchangeably with childhood mortality. The term childhood mortality as used in this research involves all forms of mortality of a baby from age zero to age 5. Children of above 5 years were not captured in this research. The following authors have made significant contributions to childhood mortality in Africa, Nigeria and in Lagos State particularly. These authors include Schwarzer, Vach and Schumander (2000), who addressed the pattern of under-five deaths in Lagos State, Nigeria; Hand (1997) worked on combinations of host Biomarkers predict mortality among Ugandan children with severe malaria; Aigbe and Zannu (2012)examined differentials in infant and child mortality rates in Nigeria, evidence from the six geopolitical zones; and Adekanbi (2015) unravelled the effects of neighbourhood contextual influences on childhood mortality and Morbidity in Nigeria. In the heart of prediction in medical sciences is prognostic models.

Broadly speaking, prediction models are valuable for medical practice and for research purposes. In public health, prediction models may help to target preventive interventions to subjects at relatively high risk of having or developing a disease. Prognostic estimates may for example be useful for planning of remaining life-time in terminal disease; or give hope for recovery if a good prognosis is expected after an acute event such as a stroke. Classification of a patient according to his/her risk may also be useful for communication among physicians. A key condition for this type of application of a prediction model is that predictions are reliable. This means that when a 10 percent risk is predicted, on average 10 percent of patients with these characteristics should have the outcome (Abu-Hanna & Lucas, 2001). Prognostic predictions may support the weighing of harms versus individual benefits. If risks of a poor outcome are relatively low, the maximum benefit will also be relatively low. Any harm, such as a side effect of treatment, may then readily outweigh any benefits.

The claim of prediction models is that better decisions can be made with a model than without (Steyerberg & Moons, 2013). In research, prediction models may assist in the design and analysis of randomized trials. Models are also useful to control for confounding variables in observational research, either in traditional regression analysis or with modern approaches such as propensity scores (Abu-Hanna & Lucas, 2001). The following authors have applied prognostic models to various field in medicine and other areas of application. They include Vogenberg (2009), who applied predictive and prognostic models as implications for healthcare decision-making in a modern recession; Abu-Hanna and Lucas (2001) worked on prognostic models in medicine using AI and statistical approaches; Pajouheshnia, Damen, Groenwold, Moons and Peelen (2017) worked on treatment use in prognostic model research as a systematic review of cardiovascular prognostic studies; and Feng, May and Tang (2019) applied prognostic models in predicting overall survival in patients with primary gastric cancer in a systematic review.

Health practitioners have collected data over years on childbirth and infant mortality/under 5 child mortality in Lagos and different part of the world. These data need to be analyzed and best recommendations be given to the medical practitioners, government and international organizations in order to reduce child mortality to its minimal. This analysis can be modelled using different statistical tools by statisticians and users of statistics. More so, take for instance in Lagos State, the rate of childhood mortality in one community may be different from the rate in another community and these differences are attributed to some factors.Therefore,this study used Prognostic model to identify factors that predict differences in childhood mortality in lagos state communities. This will help to identify particular communities that need more attention as a result of high childhood mortality when compared to other communities.

This research applied prognostic model in identifying factors that predict differences in childhood mortality in Lagos communities. The multiple binary logistic regression, multiple linear regression and multivariable Poisson regression models were used as prognostic models. So, hypothesis tests such as Wald test, Likelihood Ratio tests are carried out. Also, model selection criteria such as Akaile Information Criterion (AIC) and Bayesian Information Criterion (BIC) are carried out as well as model selection criteria to evaluate the fitness of the final models.

The rest of the research is organized into three sections. Section 2 examines the materials and method applied in this research, mainly, logistic, linear and Poisson regression models. Section 3 comprises the results and discussion of findings based on the data collected and the models developed. The conclusion and recommendation are contained in Section 4.

2. Materials and Method

The prognostic models considered in this research are binary logistic regression, linear regression and Poisson regression models, where the error terms follow normal distribution and the dependent or response variables are discrete (Ogunjuyigbe, 2004; Omariba et al., 2007). The linear regression is also considered because of the behaviour of the error term, which is assumed to have the same distribution with the dependent variable. It is assumed that when the dependent variable is skewed or non-normal but has a large sample points, it has a tendency of becoming normal for some parameter values of the initial distribution of the data (Arowolo, Nurudeen, Akinyemi, Ogunsanya & Ekum, 2019). This can also be confirmed with the central limit theorem. So it is advisable to always compare performance of any model or distribution with that of normal, especially if the sample size is large. If normal is better, then there is no need of assuming the data follows the other distribution.

2.1 Prognostic Models

Prognostic models are models used in prediction of an event before its possible occurrence. Prognostic models are useful aides to clinical management (Abu-Hanna & Lucas, 2001). There are many different ways in which prognostic models can be developed. Prognostic models can be constructed by choosing from a plethora of different techniques. Commonly used techniques are simple decision rules based on the categorization of a prognostic score, Bayes rule, and logistic regression. The parameters of these models are estimated from data collected especially for the purpose of prognosis. In this research, the data of the response variable are binary and count data. For the binary data, logistic regression model would be fitted, while for the count data, the Poisson regression model would be fitted.

2.2 Binary Response Variable

2.2.1 Multiple Binary Logistic Regression Model

Let consider the linear model given by

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} + e_i \tag{1}$$

where $\beta_0, \beta_1, \beta_2, ..., \beta_k$ are the unknown parameters of the model to be estimated, $X_1, X_2, ..., X_k$ are the k covariates and e is the error term, which is distributed as Y. The expectation of Y is given in (2).

$$\pi_i = E(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} \tag{2}$$

The prevailing assumptions for the binary logistic regression are

(i) non-normal error terms, that is

$$Y_i = \begin{cases} 0; e_i = -\beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} + ... - \beta_k X_{ki} \\ 1; e_i = 1 - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} + ... - \beta_k X_{ki} \end{cases}$$

(ii) Non-constant variance
$Var(Y_i) = \pi_i(1 - \pi_i)$
$Var(Y_i) = (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki})[1 - (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki})]$

(iii) Response function constraints: $0 \leq \pi_i \leq 1$

2.2.2 Multiple Binary Logistic Response Function

The multiple binary logistic regression response function is given by

$$\pi_i = E(Y_i) = \frac{exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki})}{1 + exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki})} \tag{3}$$

Equation (3) can be linearized through log transformation to give

$$log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} \tag{4}$$

Equation (4) is known as the logit transformation. Thus, we use logit link to relate π_i with the covariates (X).

2.2.3 Parameter Estimation of the Logistic Model

Given the distribution of Y_i, the likelihood function is given by

$$L = \prod_{i=1}^{n} \pi_i^{Y_i}(1 - \pi_i)^{1-Y_i} \tag{5}$$

The log-likelihood (logL) function is derived by taking the log of equation (5) to have

$$\ell = logL = log\left[\prod_{i=1}^{n} \pi_i^{Y_i}(1 - \pi_i)^{1-Y_i}\right]$$

$$\ell = \sum_{i=1}^{n} Y_i log\pi_i + \sum_{i=1}^{n}(1 - Y_i)log(1 - \pi_i)$$

$$\ell = \sum_{i=1}^{n} Y_i log\left(\frac{\pi_i}{1-\pi_i}\right) + \sum_{i=1}^{n} log(1 - \pi_i) \tag{6}$$

Substitute equation (4) into (6) to have

$$\ell = \sum_{i=1}^{n} Y_i(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki}) - \sum_{i=1}^{n} log[1 + exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki})] \tag{7}$$

2.2.4 Maximum Likelihood Estimation of the Binary Regression Model

Equation (7) in matrix form can be written as

$$\ell = (X\beta)'y - \sum_{i=1}^{n} log[1 + exp(X\beta)] \tag{8}$$

where ℓ is the log-likelihood, X is a $n \times p$ matrix, $p = k+1, \beta$ is a $p \times 1$ matrix, that is, a column vector, y is an $n \times 1$ matrix, a column vector. In order to obtain the maximum likelihood estimation (MLE) of the multiple binary logistic regression model parameters, take the 1st derivative of the logL in equation (8) with respect to each of the parameters and equate each to zero.

$$\frac{\partial l}{\partial \beta} = \frac{\partial}{\partial \beta}\left\{(X\beta)'y - \sum_{i=1}^{n} log[1 + exp(X\beta)]\right\} = 0 \tag{9}$$

Solve equation (9) to obtain the MLE estimates, $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$.

The MLEs do not have closed forms, so numerical methods are used through the help of computer packages. In this research, the maxLik function in R was used to obtain the MLE estimates. Haven estimated the parameters of the model, we can then calculate $\hat{\pi}_i$ as

$$\hat{\pi}_i = \frac{exp(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + ... + \hat{\beta}_k X_{ki})}{1 + exp(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + ... + \hat{\beta}_k X_{ki})} \tag{10}$$

where $\hat{\pi}_i$ is the estimated probability of individual i having response $Y_i = 1$ and $\hat{\beta}_j$, $j = 1, 2, ..., k$ is no longer the slope of a linear relationship but rather the slope of the logit relationship. Thus, the odd ratio is then given as $exp(\hat{\beta}_j)$, $j = 1, 2, ..., k$.

Count Response Variable

2.3 Linear Regression Model

The k-variable linear regression model is used to study the relationship between a dependent variable and one or more covariates. The specification of such a model is given in (1) where error term, e is normally distributed, just like the dependent variable, Y (Hand, 1998). Equation (1) can be written in matrix form as

$$Y = X\beta + e \tag{11}$$

where Y is a n-dimensional vector called the response vector, X is a $n \times p$ matrix called the design matrix, where $p = k + 1$ as earlier defined. The β represents the intercept and slopes, a p dimensional vector, and e is a n-dimensional vector called the error vector (Alaba, 2012; Olubusoye, 2013).

2.3.1 Least Square Estimation of Linear Regression Model

The least square estimation (LSE) is considered the best linear unbiased estimator (BLUE) in estimating the parameters of the multiple linear regression model. To make any meaningful progress with the estimation of the vector of coefficients, β, some assumptions must be satisfied. (i) There must be a linear relationship between the response variable Y and the covariates or independent variables X. It may be linear either in the original variables or after some suitable transformation. (ii) The X columns must be linearly independent (iii) $E(Y) = X\beta$ since $E(e_i) = 0$ (iv) $Var(e_i/X) = \sigma^2$, $Cov(e_i e_j) = 0$ (v) X is unrelated to e and (vi) e $N(0, \sigma^2)$. The last assumption here shows that Y follows a normal distribution (Ekum, Akinmoladun, Aderele & Esan, 2015).

The LSE is achieved by minimizing the error term e. The error term is given by

$$e = Y - X\beta \tag{12}$$

This e is assumed to be very small, so it is minimized by first squaring (12) to have

$$e^2 = (Y - X\beta)^{'}(Y - X\beta) \tag{13}$$

Differentiating (13) with respect to β will further minimize it, Then equate it to zero and solve for β to have

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{14}$$

where $\hat{\beta}$ is a column vector that house all the parameter estimates. The response variable Y can the be estimated by

$$\hat{Y} = X\hat{\beta} \tag{15}$$

2.4 Poisson Regression Model

In statistics, Poisson regression is a generalized linear model (GLM) form of regression analysis used to model count data. Poisson regression assumes the response variable Y has a Poisson distribution, and assumes the logarithm of its expected value can be modelled by a linear combination of unknown parameters. A Poisson regression model is a log-linear model when used to model contingency tables. The Poisson regression model aims at modelling a count variable Y (rare events), counting the number of times that a certain event occurs during a given time period. The basic GLM for count data is the Poisson model with log link.

2.4.1 Poisson Distribution

Suppose Y is a discrete random variable that follows a Poission distribution, The probability mass function of Y is given as

$$P(Y = y) = \frac{\theta^y e^{-\theta}}{y!}, \theta > 0, y = 0, 1, 2... \tag{16}$$

where θ is the mean response of Y. The parameter θ is the mean of Y and it is the same as the variance of Y. That is, $E(Y) = Var(Y) = \theta$. The shape of Y with various values of θ is shown in Figure 1.

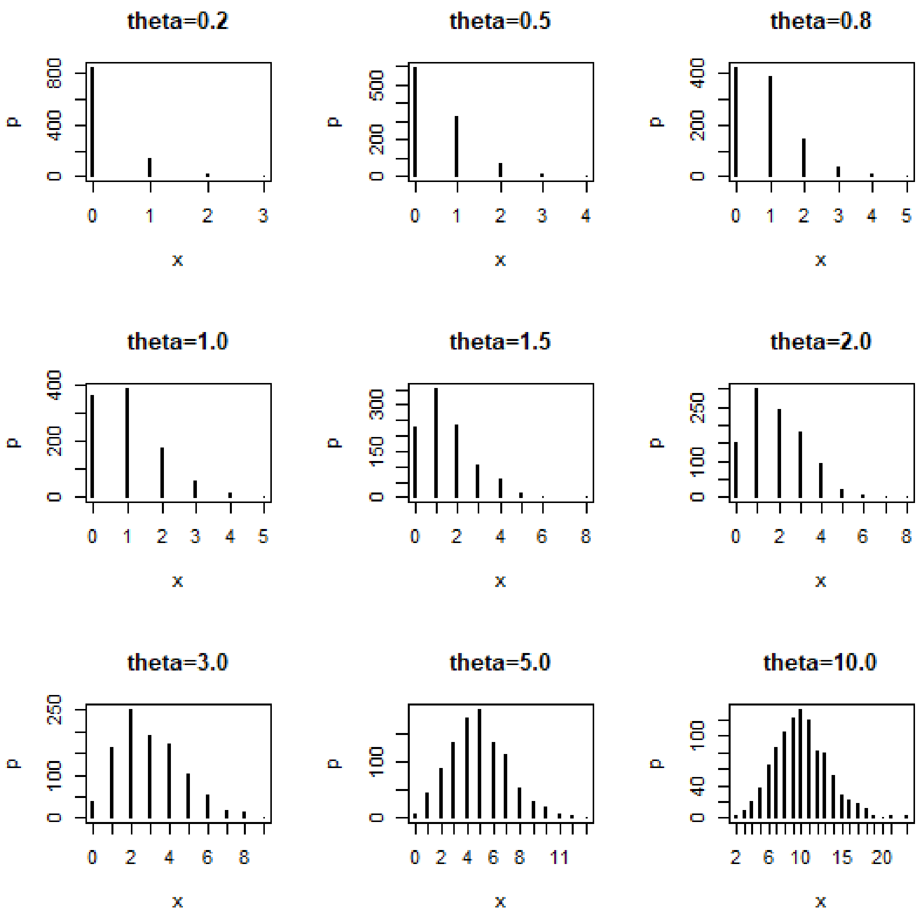


Figure 1. Poisson distribution with various values of θ

2.4.2 Multivariable Poisson Regression Model

Suppose a linear regression model with k covariates as given in (1) and can be written in matrix form as given in (11). Suppose $Y_1, Y_2, ..., Y_n$ are Poisson distributed with associated covariates $X_1, X_2, ..., X_n$. The mean response for the ith case is denoted by θ_i and is assumed to be a function of the predictor variables $X_1, X_2, ..., X_k$.

$$E(Y_i) = \theta_i = XB \tag{17}$$

The notation $\theta_i(X,\beta)$ is used to denote the function that relates the mean response θ_i to X, the values of the predictor variables for case i , and β, the values of the regression coefficients. The conditional mean function using a linear combination of the covariates is given by.

$$E(Y_i|X) = exp(XB) \tag{18}$$

2.4.3 Maximum Likelihood of Poisson Regression Parameter

It should be noted that y_i is a realization from the random variable Y_i and X is a $n \times k$ matrix of covariates.

Recall the Poisson function in equation (16), we have

$$P(Y_i = y_i|\beta, X) = \frac{\theta_i^{y_i} e^{-\theta_i}}{y_i!}, \ \theta_i > 0, y_i = 0, 1, 2...; i = 1, 2, ..., n \tag{19}$$

The likelihood function is given by

$$L(\beta) = L(y_i|\beta, X) = \prod_{i}^{n} \frac{\theta_i^{y_i} e^{-\theta_i}}{y_i!} \tag{20}$$

$$L(\beta) = L(y_i|\beta, X) = \prod_{i}^{n} \frac{[\theta_i(X'\beta)]^{y_i} e^{-[\theta_i(X'\beta)]}}{y_i!} \tag{21}$$

Take the log of (21) to have the log-likelihood function as

$$\ell = logL(\beta) = \sum_{i=1}^{n} y_i log[\theta_i(X'\beta)] - \sum_{i=1}^{n} \theta_i(X'\beta) - \sum_{i=1}^{n} log(y_i) \tag{22}$$

The log-likelihood can then be differentiated partially with respect to each of the parameters. The result of the derivatives are the be equated to zero and solve for the parameters. Note that β is a vector of parameters, to accommodate multiple covariates. Practically, iterative methods are used to estimate the parameters. In this work, we used the maxLik package in R to estimate the Poisson regression parameters.

The maximum likelihood covariance matrix of the regression coefficient is given by

$$cov(\hat{\beta}) = (X'\hat{\theta}X)^{-1} \tag{23}$$

where β is a vector of regression coefficients, $\hat{\theta}$ is an $N \times N$ diagonal matrix with ith diagonal elements $\theta_i, i = 1, 2, ..., n$, which represents the variance of Poisson distribution.

2.4.4 Deviance Residual of Poisson Regression

The deviance of a fitted model is the difference between the log-likelihood (logL) of the fitted model and a model that has a parameter (μ_i) for each observation Y_i (i.e., use all of the degrees of freedom so that the residuals will be zero). in this research $\mu_i = \theta_i$.

The Poisson regression model deviance is given by

$$Dev(x_1, x_2, ..., x_k) = -2\left[\sum_{i=1}^{n} y_i log\left(\frac{\hat{\theta}_i}{y_i}\right) + \sum_{i=1}^{n}(y_i - \hat{\theta}_i)\right] \tag{24}$$

where $\hat{\theta}_i$ is the fitted value for the ith case. The deviance residual for the ith case is given by

$$dev_i = \pm\left[-2y_i log\left(\frac{\hat{\theta}_i}{y_i}\right) - 2(y_i - \hat{\theta}_i)\right]^{-\frac{1}{2}}. \tag{25}$$

The sign of the deviance residual is selected based on whether $y_i - \hat{\theta}_i$ is positive or negative. Index plots of the deviance residuals and half-normal probability plots with simulated envelopes are useful for identifying outliers and checking the model fit. Note that if $y_i = 0$, then the term $y_i log\left(\frac{\hat{\theta}_i}{y_i}\right)$ in (24) and (25) equals zero.

2.4.5 Poisson Regression Model Specification

The Poisson regression model for this work is therefore specified as

$$\theta_i = \theta(X, \beta) = exp(X'B) = exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki}) \tag{26}$$

where β_0 is constant and β_j, are the coefficients of x_i for $j = 1, 2, ..., k$. the function that relates the mean response θ_i to X. To test whether there is a relationship between the x_j and Y, that is, to test whether $\beta_j = 0$, based on the maximum likelihood estimates of the Poisson regression model parameters given in (23), hypothesis tests such as Wald test, Lagrange Multiplier test, or Likelihood Ratio tests are carried out. Also, model selection criteria such as Akaile Information Criterion (AIC) and Bayesian Information Criterion (BIC) are carried out as well. In this research, the Wald test (Z) is used for the hypothesis test while the AIC and BIC are used as model selection criteria. Hypothesis testing and formulation of test statistics is done in similar manner to logistic regression.

2.5 Wald Test (Z) Test

A large-sample test of a regression parameter can be constructed based on the null hypothesis (H_0) that $\beta_j = 0$ against the alternative hypothesis (H_1) that $\beta_j \neq 0$. The appropriate test statistic is given by.

$$Z = \frac{\beta_j}{S(\beta_j)}, j = 1, 2, ..., k + 1 \tag{27}$$

with the decision rule that specify the rejection of H_0, if $|Z| > Z_{1-\alpha/2}$, where $Z_{1-\alpha/2}$ is a standard normal value obtained from the standard normal distribution table, α is the level of significance, $S(\beta_j)$ is the standard error of β_j and β_j are the estimates from the maximum likelihood estimator.

2.6 Model Selection Criteria

The model selection criteria considered in this work are the AIC and BIC.
The AIC is given by

$$AIC = -2logL + 2(k + 1) \tag{28}$$

where $logL$ is the log-likelihood of the model, $k + 1 = p$ is the number of parameters in the model. If the number of observations, n (sample size) is small, such that, $\frac{n}{p} < 40$, then the AIC needs to be adjusted to

$$AIC_C = nlogL + 2p + \frac{2p(p + 1)}{n - p - 1}, \tag{29}$$

where p is the number of parameters. The adjusted AIC in equation (29) shows that if $n \to \infty$, then $\frac{2p(p+1)}{n-p-1} \to 0$ in (29). The other model selection criterion used in this work is the BIC. The BIC in general is given by

$$BIC = nlogL + nlog(n) + plog(n) \tag{30}$$

where p is the number of parameters in the model and n is the sample size. The decision criterion is to select the model with the smallest values of AIC and BIC as the most suitable model (Schwarz 1978).

3. Results and Discussion

3.1 Simulation

In this section, the relation between lomax-Cauchy and lomax random v was investigated.

3.2 Application

This research is focused on the causes of variation in childhood mortality in Lagos State. Primary data were collected from the five divisions in Lagos State viz-a-viz Ikorodu, Badagry, Lagos Mainland/Ikeja, Lagos Island and Epe. A well structured questionnaire was used to collect the data used for the analysis. The questionnaire comprises 12 items, 7 from Section A and 5 from Section B. See the questionnaire in Appendix II. The target population are women who fall in the child bearing age and live in Lagos State. The 5 divisions in Lagos State are assumed to be equally distributed with women that falls in the age bracket (15-49), so 120 questionnaires were distributed to 120 independent individuals in each division, totalling 600 questionnaires. The sample size was based on cost and trained personnel available. Out of the 600 questionnaires distributed, 520 were retrieved and out of the 520 retrieved, after careful sorting to remove male respondents and questionnaire that were not valid, the number reduced to 476. The 476 valid questionnaires were used for the analysis. The data were entered into Microsoft Excel, saved with the extension .csv and analyzed using R 3.6.1 and SPSS 23.0 packages.

3.2.1 Exploratory Data Analysis

This subsection provides information on some hidden features that might be inherent in the data collected. Frequency distribution tables, contingency tables, bar charts and density plots are used to explore the data.

Table 1. Frequency Distribution Table

Variables	Valid Response	Frequency	Percentage (%)
Age	Less than 20	59	12.4
	20 - 30	180	37.8
	Above 30	237	49.8
Marital status	Married	315	66.2
	Divorced	100	21.0
	Single	61	12.8
Lagos divisions	Ikorodu	98	20.6
	Badagry	94	19.7
	Ikeja	97	20.4
	Lagos Island	97	20.4
	Epe	90	18.9
Religion	Christianity	268	56.3
	Islam	152	31.9
	Others	56	11.8
Education	No School	59	12.4
	Primary	144	30.3
	Secondary	150	31.5
	Tertiary	123	25.8

Table 1 shows that most of the respondents (49.8%) are women above 30 years of age and are mostly (66.2%) married women. Out of the 120 questionnaires administered to each of the 5 divisions in Lagos State, 98 were from Ikorodu division, 94 from Badagry division, 97 from Mainland/Ikeja division, 97 from Lagos Island division and 90 from Epe division with Ikorodu, Badagry, Ikeja, Island and Epe accounting for 20.6%, 19.7%, 20.4%, 20.4% and 18.9% respectively. The religion status shows that 56.3% of the respondents are Christians, 31.9% are Muslims, while the remaining 11.8% are practising other religions. Most of the respondents are secondary school holders, accounting for 31.5%, 25.8% have tertiary education certificate, 30.3% have only first school leaving certificate, while the remaining 12.4% are stacked illiterate.

The five variables in Table 1 are used as the covariate in this study. The dependent variables are (i) the number of children (under 5 years) born by a woman that are death (count variable); and the second dependent variable (ii) is binary (0 or 1), if a woman does not have any child death, 0 is assigned but if a woman has at least on child that is death, 1 is assigned. The number of children death is shown in Figure 3 and the binary variable is shown in Appendix IV. Apendix V shows the educational status of the respondents by divisions in Lagos and Appendix shows causes of infant mortality across the 5 divisions in Lagos State.

Figure 2 shows the probability of vulnerability of mothers to childhood mortality (under 5 death). The mothers (women interviewed) that are most vulnerable to childhood mortality include mothers that are above 30 years of age, mothers who live in Ikorodu community of Lagos State, mothers who are Christians and mothers who have no education (illiterate mothers). This does not mean Christians mothers are vulnerable to child mortality in the real sense of it because the instrument did not ask if they were Christians before losing their babies. Also, some women would have lost their child before relocating to Ikorodu, because the instrument did not also capture where they reside when they lost their children and how old they were when the incident happened. The only thing that is very certain is that uneducated mothers are most vulnerable to having children who may die before age 5.

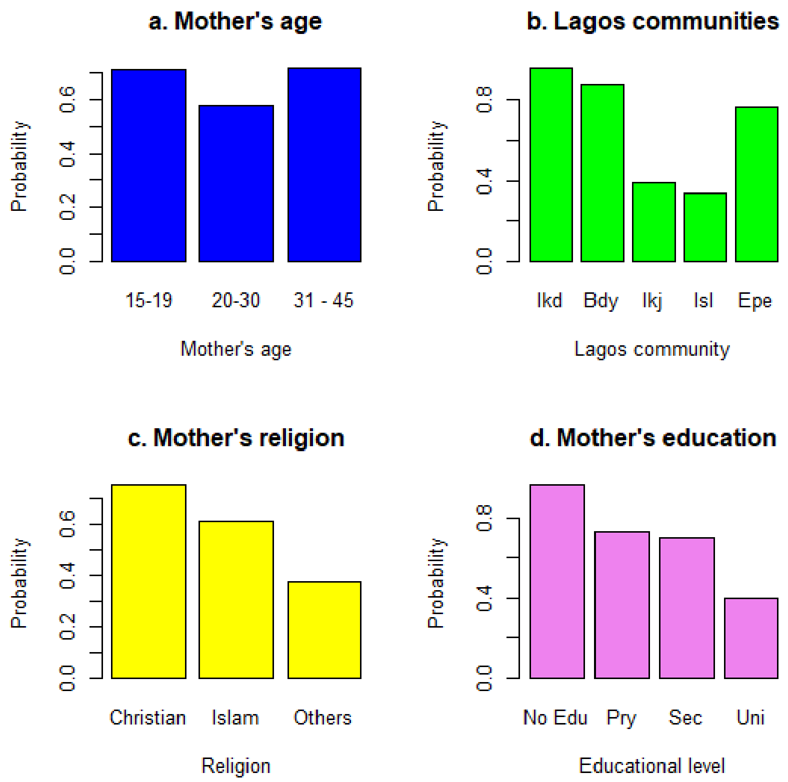


Figure 2. Probability of Under 5 Mortality

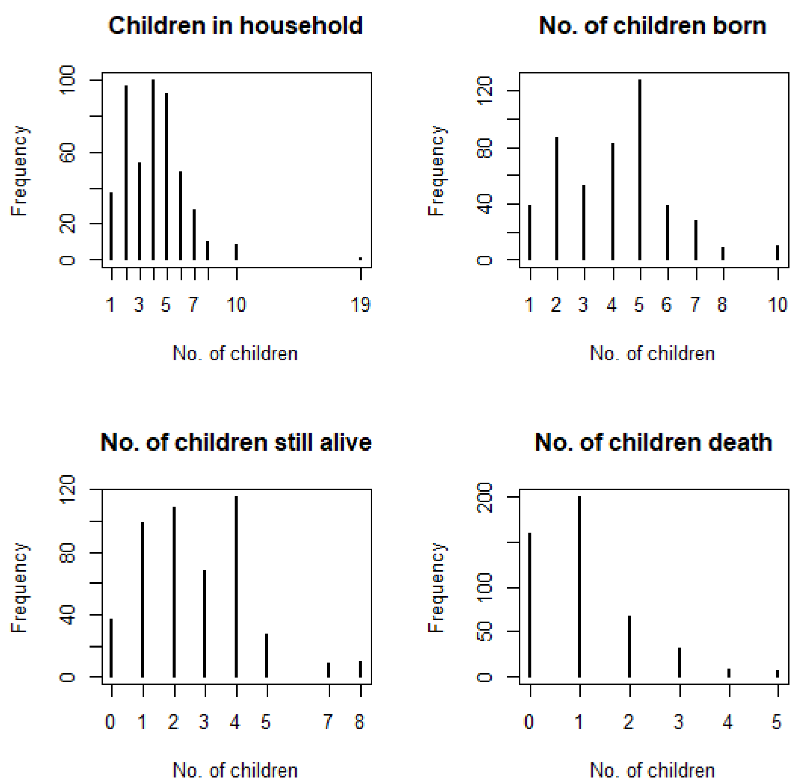


Figure 3. Count variables in study

Figure 3 depicts that most of the household where the respondents live have 4 children, most of the children born by our respondents are 5. The number of children that are still alive to women are 4, while the number of children death from a respondent is 1 (count dependent variable). The mode is used in this regard as measure of central tendency. Appendix 1 is a bar chart that depicts Table 1 clearly and also shows the count dependent variable. Appendix III shows the causes of infant mortality. The term infant mortality as used in this study comprises all children between age 0 to 5.

Table 2. Number of children death across division

		Number of children death						Total
		0	1	2	3	4	5	
Lagos divisions	Ikorodu	4	57	22	10	2	3	98
	Badagry	12	50	17	10	3	2	94
	Ikeja	59	22	15	1	0	0	97
	Lagos Island	64	28	2	3	0	0	97
	Epe	21	43	11	8	4	3	90
Total		160	200	67	32	9	8	476

Table 3. Death or Alive

		Death		Total
		Alive	Death	
Lagos divisions	Ikorodu	4	94	98
	Badagry	12	82	94
	Ikeja	59	38	97
	Lagos Island	64	33	97
	Epe	21	69	90
Total		160	316	476

Table 2 shows that the maximum number of children that are death from a mother is 5. The table shows that 200 women have lost just 1 child, 67 women have lost 2 children, 32 women have lost 3 children, 9 women have lost 4 children, and 8 women have lost 5 children. The number of children lost by a woman is spread across the five divisions in Lagos.

Table 3 shows the number of women who have never lost a child and who have lost at least one child. In Ikorodu, out of the 98 women analyzed, 94 has recorded infant death while only 4 did not. In Badagry, out of 94 women interviewed, 82 has recorded infant death while 12 have not. In Ikeja division, out of the 97 women analyzed 38 have recorded infant death while 59 have not. In Lagos Island, out of the 97 women analyzed, 33 have recorded death, while 64 have not. In Epe division, ut of the 90 women analyzed, 69 has recorded death while 21 have not. Judging from this table, Ikorodu has the highest infant mortality rate, followed by Badagry and Epe. Lagos Island has the least infant mortality rate followed by Ikeja. So, the state government should focus on Ikorodu, Badagry and Epe in oredr to reduce under 5 mortality rate. Table 2 (count data) and Table 3 (binary data) form the basis of the models to be discussed in subsequent section.

3.3 Prognostic Regression Analysis Estimation

In this study, the response variables (Y) are (i) number of children born by a woman who died before age 5. (ii) response of a woman to whether she has lost a child (below age 5), with response 0, if she has not lost a child, and 1 if she has lost at least a child. The probability distribution $P(Y_i = 1) = \pi_i$ or $P(Y_i = 0) = 1 - \pi_i$, where response variable (Y) denotes death or alive. The independent variables or covariates are age of mother (X_1), marital status of mother (X_2), division in Lagos State where mother lives (X_3), mother's religion (X_4) and mother's education (X_5). The term mother used here refers to our respondents who have given birth to at least one child.

There are basically two different models. One is a binary response and will be modelled using binary logistic regression. The other is a count response and will be modelled by linear regression and Poisson regression. The count response is a subset of the continuous response. If the sample size is large enough, then it can be approximated by normal distribution, which is the basis for the linear regression.

? **Model A: Binary Response Variable**

1. Use all the covariates (binary logistic regression)

2. Drop a covariates that is not significant (binary logistic regression)

? **Model B: Count Response Variable**

3. Use all the covariates (linear model)
4. Drop a covariates that is not significant (linear model)
5. Use all the covariates (linear model)
6. Drop a covariates that is not significant (linear model)

Models 1 and 2 in A are competing models. Models 3, 4, 5 and 6 are competing models. Models in A are not competing with models in B. We are not only interest if a woman has lost a child or not, we are also interest in the number of children a woman lost. This will help to measure infant mortality. The formation of the models is described in the flowchart as shown in Figure 4.

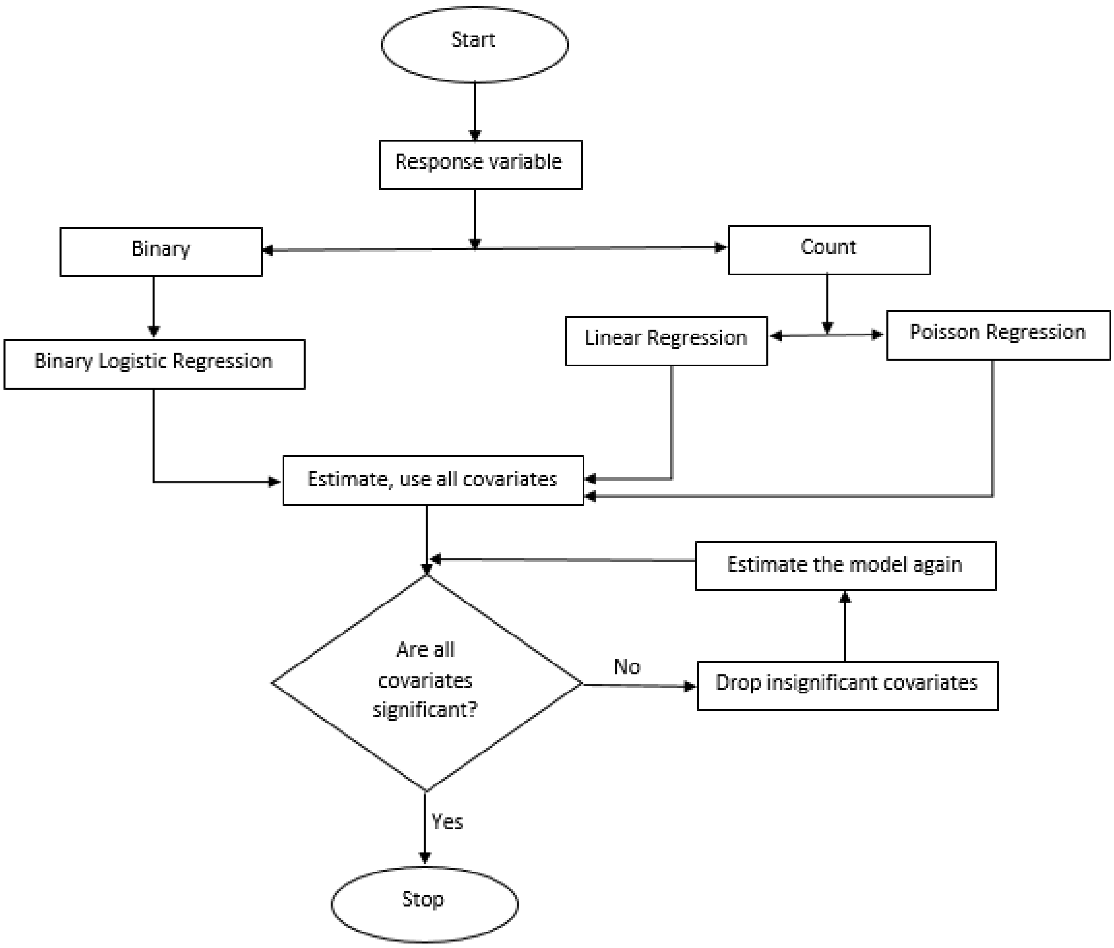


Figure 4. Flow Chart of Model Construction

3.3.1 Fitting Binary Logistic Regression Models

Table 4. Multiple Binary Logistic Regression Estimates

	Estimate	Std. Error	Z value	P-value
(Intercept)	5.43891	0.79362	6.853	0.00000
Age	0.40373	0.18765	2.152	0.03143
Marital Status	-0.31281	0.19824	-1.578	0.11458
Lagos Division	-0.52022	0.09432	-5.515	0.00000
Religion	-0.5558	0.17217	-3.228	0.00125
Education	-0.95764	0.13165	-7.274	0.00000

Model 1 regression parameters estimated in Table 4 can be expressed as

$$\hat{\pi}_i = \frac{exp(5.43891 + 0.40373x_{1i} - 0.31281x_{2i} - 0.52022x_{3i} - 0.5558x_{4i} - 0.95764x_i)}{1 + exp(5.43891 + 0.40373x_{1i} - 0.31281x_{2i} - 0.52022x_{3i} - 0.5558x_{4i} - 0.95764x_i)} \quad (31)$$

From the analysis in Table 4, it is easy to see that division of Lagos where respondent lives and education are the most significant covariates. Also, significant 5% are religion and age of the mothers. (respondents). In this model 1, the marital status of the mother is not significant at 5% level. So, there is a need to develop model 2 from binary logistic regression.

Model 2 is constructed after dropping the insignificant marital status and it is shown in Table 5.

Table 5. Binary Logistic Regression Estimates of 4 Covariates

	Estimate	Std. Error	Z value	P-value
(Intercept)	4.7152	0.6258	7.534	0.00000
Age	0.5173	0.1734	2.982	0.00286
Lagos Division	-0.4783	0.0895	-5.344	0.00000
Religion	-0.6755	0.1546	-4.368	0.00001
Education	-0.9354	0.1313	-7.125	0.00000

Model 2 regression parameters estimated in Table 5 can be expressed as

$$\hat{\pi}_i = \frac{exp(4.7152 + 0.5173x_{1i} - 0.4783x_{3i} - 0.6755x_{34i} - 0.9354x_{5i})}{1 + exp(exp(4.7152 + 0.5173x_{1i} - 0.4783x_{3i} - 0.6755x_{34i} - 0.9354x_{5i})} \quad (32)$$

3.3.2 Fitting Multiple Linear Regression Models

Table 6. Multiple Linear Regression Estimates with 5 parameters

	Estimate	Std. Error	t value	P-value
(Intercept)	1.68265	0.29317	5.740	0.000000
Age	0.30570	0.07115	4.296	0.000021
Marital Status	0.11032	0.0773	1.427	0.154225
Lagos Division	-0.11955	0.03475	-3.441	0.000632
Religion	-0.24207	0.07234	-3.346	0.000885
Education	-0.28634	0.04829	-5.930	0.000000

Model 3 regression parameters estimated in Table 6 can be expressed as

$$\hat{y}_i = 1.68265 + 0.30570x_{1i} + 0.11032x_{2i} - 0.11955x_{3i} - 0.24207x_{4i} - 0.28634x_{5i} \quad (33)$$

Table 6 shows that all the covariates are significant at 5% level, except for marital status. So, there is a need to develop model 4 by dropping marital status.

Table 7. Multiple Linear Regression Estimates with 4 parameters

	Estimate	Std. Error	t value	P-value
(Intercept)	1.94707	0.22744	8.561	0.000000
Age	0.27112	0.06698	4.048	0.000060
Lagos Division	-0.13521	0.03300	-4.097	0.000049
Religion	-0.20332	0.06713	-3.029	0.002590
Education	-0.29899	0.04752	-6.292	0.000000

Model 4 regression parameters estimated in Table 7 can be expressed as

$$\hat{y}_i = 1.94707 + 0.27112x_{1i} - 0.13521x_{3i} - 0.20332x_{4i} - 0.29899x_i \quad (34)$$

3.3.3 Fitting Multivariable Poisson Regression Models

Table 8. Poisson Regression Estimates of 5 Covariates

	Estimate	Std. Error	Z value	P-value
(Intercept)	0.67483	0.29845	2.261	0.023755
Age	0.26456	0.06975	3.793	0.000149
Marital Status	0.07104	0.07478	0.950	0.342148
Lagos Division	-0.1162	0.03297	-3.525	0.000424
Religion	-0.24193	0.07541	-3.208	0.001336
Education	-0.26067	0.04666	-5.587	0.000000

- Copy.PNG

Model 5 regression parameters estimated in Table 8 can be expressed as

$$\hat{\theta}_i = exp(0.67483 + 0.26456x_{1i} + 0.07104x_{2i} - 0.1162x_{3i} - 0.24193x_{4i} - 0.26067x_{5i}) \tag{35}$$

Table 8 shows that all the covariates are significant at 5% level, except for marital status. So, there is a need to develop model 6 by dropping marital status just as the case of model 2 and model 4.

Table 9. Poisson Regression Estimates of 4 Covariates

	Estimate	Std. Error	Z value	P-value
(Intercept)	0.85774	0.22511	3.81	0.000139
Age	0.2439	0.06595	3.698	0.000217
Lagos Division	-0.12743	0.03068	-4.154	0.000033
Religion	-0.22004	0.072	-3.056	0.002241
Education	-0.27187	0.04524	-6.01	0.000000

- Copy.PNG

Model 6 regression parameters estimated in Table 9 can be expressed as

$$\hat{\theta}_i = exp(0.85774 + 0.2439x_{1i} - 0.12743x_{3i} - 0.22004x_{4i} - 0.27187x_{5i}) \tag{36}$$

Table 10. Model Selection Criteria

Models	-logL	AIC	BIC
Binary logistic regression			
Model 1 (5 covariates)	237.6006	487.2000	512.1937
Model 2 (4 covariates)	238.8635	487.7300	508.5541
Multiple linear regression			
Model 3 (5 covariates)	673.8634	1361.7270	1390.8850
Model 4 (4 covariates)	674.8924	1361.7850	1386.7770
Poisson regression			
Model 5 (5 covariates)	604.4913	1221.0000	1245.9750
Model 6 (4 covariates)	604.9396	1219.9000	1240.7060

Table 10 shows the model selection criteria. For the binary response, model 1 with 5 covariates is selected over model 2 because it has the smaller - log likelihood (-logL), AIC and BIC. The same criteria are used to select Model 6 with 4 covariates among the competing models for the count response. This selection is buttressed by the cluster bar plots in Figure 5.

- Copy.PNG

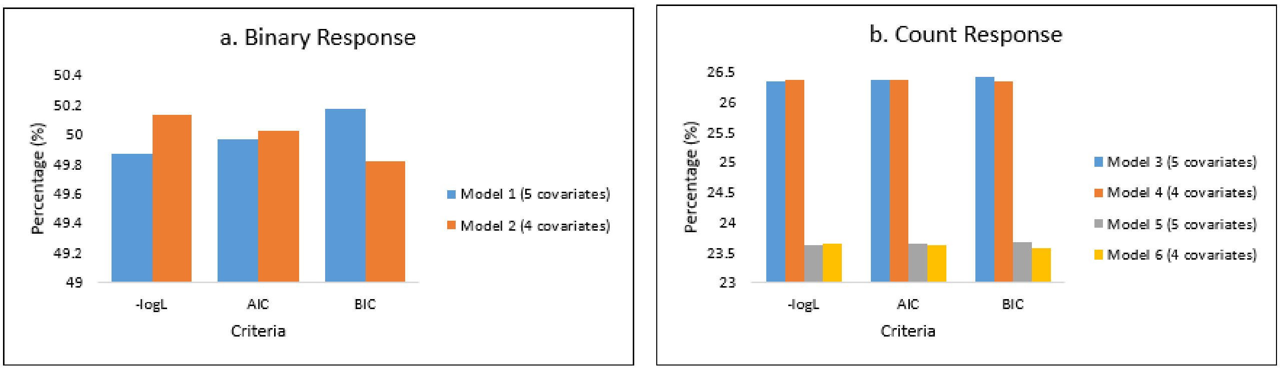


Figure 5. Model Selection Criteria

4. Conclusion and Recommendation

In this paper, we developed six competing prognostic regression models to determine factors that affect infant mortality (under 5 death) across the five divisions in Lagos State. The first two models competed for the binary response variable. Based on the -logL, AIC and BIC model selection criteria, Model 1 is selected over Model 2 because it has the smaller -logL, AIC and BIC. For the count response variable, four models competed, Model 3, Model 4, Model 5 and Model 6. Model 6 (Poisson regression with 4 parameters) is selected as the most appropriate model for fitting the count response variable because it has the smallest -logL, AIC and BIC. The estimated probability of childhood mortality in Lagos State among children of under-5 years old is 0.26. This implies that out of every 100 under 5 years children involved in this research, 26 die. Out of a total of 1,946 children born to 476 women under study 506 are death. This probability is based on this study and can be generalized if the sample is consistent with the general population of Lagos State. The study shows that out of 476 women that participated in this study 316 of them have lost at least a child under 5 years old (probability is 316/476 = 0.66). In order to identify predictors of variations of childhood mortality among Lagos communities (divisions), the multivariable poisson regression model is selected as the suitable prognostic model that identified the predictors that cause the variations of childhood mortality among Lagos State communities. These significant predictors are the age of the mother, Lagos community mother resides, mother's religion and mother's education. Women above 30 years (31-45) are most vulnerable to infant mortality, women living in Ikorodu community of Lagos State are more vulnerable to infant mortality, Christian women are more vulnerable to infant mortality and women with no education are more vulnerable to infant mortality (see Figure 2). This conclusions are based on the status of the women as at the time the interview was conducted. Some of the limitations of this research is that we did not know the women's age as at the time they lost their babies, we did not know the religion the women practiced when they lost their babies, some women would have lost their babies before becoming a Christian. We did not know the community the women resides when they lost their babies. The only factor that is very certain is the women educational level. The illiterate women are most vulnerable to infant mortality. The identified high-risk communities where childhood mortality was higher than expected in Lagos State are Ikorodu, Badagry and Epe communities. This could be traced to the fact that most of the women with no education interviewed resides in Ikorodu, Badagry and Epe in that order. Ikeja has few uneducated women while Lagos Island has none uneducated women among the women interviewed. Another factor that could be responsible is the lack of access to health care facilities. Epe, Ikorodu and Badagry in that order have the highest cause of infant mortality due to lack of access to health care facilities and the least is Lagos Island. Family size is another cause of infant mortality among Lagos communities with Epe, Badagry and Ikorodu in that order being the highest and the least is Ikeja. Appendix VI depicts causes of infant mortality in Lagos State across the five communities (divisions). Based on our findings in this research, we recommend the following (i) Female child should be given adequate education so that they can grow to become educated mothers as illiteracy among women is a major factor causing high rate of infant mortality. (ii) More health care facilities should be deployed to Ikorodu, Badagry and Epe with well trained paediatricians and gynaecologists. Mother and child units of the General Hospitals in these locations should be well equipped. The Pharmacies in these facilities should be well stocked with drugs for under 5 age. Lagos State has free health facility for the under 5 age but most women are asked to get drugs outside the General Hospitals because the drugs are not always sufficient. (iii) The life of a child should be very much important to the government so that a child taken to any of the Lagos State health facility should be treated at no cost, and all drugs administer at no cost. (iv) Poverty is another major cost of infant mortality in Lagos communities. This is as a result of large family size. A large family size would not have been a problem if the family income is large. But situations where the family size is large and the family income is low, then it can be a major cause of infant mortality. Government and all stake holders should inaugurate empowerment programmes for women, especially in Ikorodu, Badagry and Epe communities of Lagos State. If all these recommendations are honestly followed, childhood mortality will be completely eradicated in Lagos State.

Competing interest There is no competing interest to the best of authors knowledge concerning this article.

Funding

This research work was fully funded by TETfund Nigeria

Authors contribution

All authors read and approved the final manuscript before submission

Acknowledgement

The authors acknowledge with thanks the support received from Tetfund Nigeria towards the actualization of his work

References

Abu-Hanna, A., & Lucas, P. J. (2001). Prognostic models in medicine: AI and Statistical Approaches. *Methods of Information in Medicine, 40*(1), 1-5.

Adekanmbi, V. T. (2015). Unravelling the effects of neighbourhood contextual influences on childhood mortality and morbidity in Nigeria. Submitted for consideration for the degree of Doctor of Philosophy by published work C Health Sciences. Warwick Medical School University of Warwick.

Adeshina, Y. (2014). Infant mortality rates in Lagos has dropped. Press Release May 15, 2014 19:08:19pm GMT. Accessed on December 29, 2016 11:36:23am. Available at http://worldstagegroup.com/worldstagenew/index.php.

Aigbe, G. O., & Zannu, A. E. (2012). Differentials in infant and child mortality rates in Nigeria: Evidence from the Six Geopolitical Zones. *International Journal of Humanities and Social Science, 2*(16), 206-214.

Alaba, O. O. (2012). Regression and analysis of variance note. Department of Statistics, University of Ibadan, Oyo state.

Arowolo, O. T., Nurudeen, T. S., Akinyemi, J. A., Ogunsanya, A. S. & Ekum, M. I. (2019). Reduced beta skewed laplace distribution with application to failure-time of electrical Component Data. *Annals of Statistical Theory and Applications (ASTA), 1*, 31-41.

Asakitikpi, A. E. (2007). An interrogation of diarrhoea concept among Yoruba women in Ibadan Metropolis, Nigeria. *Nordic Journal of African Studies, 16*(1), 64-74.

Ekum, M. I., Akinmoladun, O. M., Aderele, O. R.,& Esan, O. A. (2015). Application of multivariate analysis on the effects of World Development indicators on GDP Per Capita of Nigeria (1981-2013). *International Journal of Science and Technology (IJST), 4*(12), 524-534.

Erdman, L. K., Dhabangi, A., Musoke, C., Conroy, A.L., Hawkes, M., Higgins, S., & Kain, K. C. (2011). Combinations of host biomarkers predict mortality among Ugandan children with severe malaria: A Retrospective Case-Control Study. *PLoS ONE, 6*(2), e17440.

Feng, Q., May, T. M., & Tang, J. (2019). Prognostic models for predicting overall survival in patients with primary gastric cancer: A Systematic Review. *Biomed Res Int.*, 5634598. https://doi.org/10.1155/2019/5634598

Feyisetan, J. B., Asa, S. & Ebigbola, A. J. (1997). Mothers management of childhood diseases in Yorubaland: the influence of cultural beliefs. *Health Transition Review, 7*, 221-34.

Hand, D. J. (1998). Data mining: Statistics and more? *Am Stat, 52*(2), 112-128. https://doi.org/10.2307/2685468

Hand, D. J. (1997). *Construction and assessment of classification rules*. Chichester: John Wiley and Sons 1997.

Harttgen, K., & Misselhorn, M. (2006). A multilevel approach to explain child mortality and undernutrition in South Asia and Sub-Saharan. *Ibero-America Institute for Economic Research, Nr, 156*, 1-38.

Ijomah, M. A., Biu, E. O., & Mgbeahurike, C. (2018). Assessing logistic and poisson regression model in analyzing count data. *International Journal of Applied Science and Mathematical Theory*, 4(1), 42C68.

Jegede, S. A., Ajala, A. S., Adejumo, O. P., & Osunwale, S. O. (2006). Forced feeding practice in Yoruba community of southwestern Nigeria: Evidence from Ethnographic Research. *Anthropologist, 8*(3), 171-179. https://doi.org/10.1080/09720073.2006.11890955

Mazumdar, M. J. (2000). Glassman, Tutorial in biostatistics, categorizing a prognostic variable: review of methods, code for easy implementation and applications to decision-making about cancer treatments. *Stat Med, 19*, 113-32. https://doi.org/10.1002/(SICI)1097-0258(20000115)19:1¡113::AID-SIM245¿3.0.CO;2-O

Michael, H. K., Christopher, J. N., John, N., & William, L. (2005). *Applied linear statistical model, fifth Ed.*. McGraw Hill International, New York. 555-623.

Neuberger, J. (1997). Prognostic models - what is their future? *Eur J Gastroenterol Hepatol, 9*, 1145-1152.

Ogunjuyigbe, P. O. (2004). Under-five mortality in Nigeria: Perception and Attitude of the Yoruba towards the Existence of Abiku. *Demographic Reserarch, 11*(2), 43-56. https://doi.org/10.4054/DemRes.2004.11.2

Ojikutu, R. K. (2008). Pattern of under-five Deaths in Lagos State, Nigeria. *Sudanese Journal of Public Health, 3*(4), 1-15.

Oke, A., Adedokun, L., Soretire, F., & Faweya, O. (2001). Lagos State, Nigeria: Report of the In-Depth assessment of the HIV/AIDs situation-Implementing Aids prevention and Care project (IMPACT), USAID, family Health International.

Olubusoye, E. O. (2013). Econometric model Centre for Econometric and Allied Research, Ibadan. Oyo state.

Omariba, D., Walter, R., Roderic, B.,& Fernando, R. (2007). Determinants of infant and child mortality in Kenya: an analysis controlling for frailty effects. *Popul Res Policy Rev, 26*, 299C321.

Pajouheshnia, R., Damen, J. A. G., Groenwold, R. H. H., Moons, K. G. M., & Peelen, L. M. (2017). Treatment use in prognostic model research: a systematic review of cardiovascular prognostic studies. *AM Health Drug Benefits, 15*. $https://link.springer.com/article/10.1186/s41512-017-0015-0$.

Salzberg, S. L. (1999). On comparing classifiers: a critique of current research and methods. *Data Mining Knowl Discov, 1*, 1-12.

Sastry, N. (1997). Family-level clustering of childhood mortality risk in Northeast Brazil. *Population Studies, 51*, 245-261. https://doi.org/10.1080/0032472031000150036

Schwarzer, G., Vach, W., & Schumacher, M. (2000). On misuses of artificial neural networks for prognostic and diagnostic classification in oncology. *Stat Med, 19*, 541-61. https://doi.org/10.1002/(SICI)1097-0258(20000229)19:4¡541::AID-SIM355¿3.0.CO;2-V

Steyerberg, E. W., Moons, K. G. M., & for the PROGRESS Group (2013). Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLoS Med., 10*(2), e1001381.

Vogenberg, F. R. (2009). Predictive and Prognostic Models: Implications for Healthcare decision-making in a modern recession. *AM Health Drug Benefits, 2*(6), 218-222.

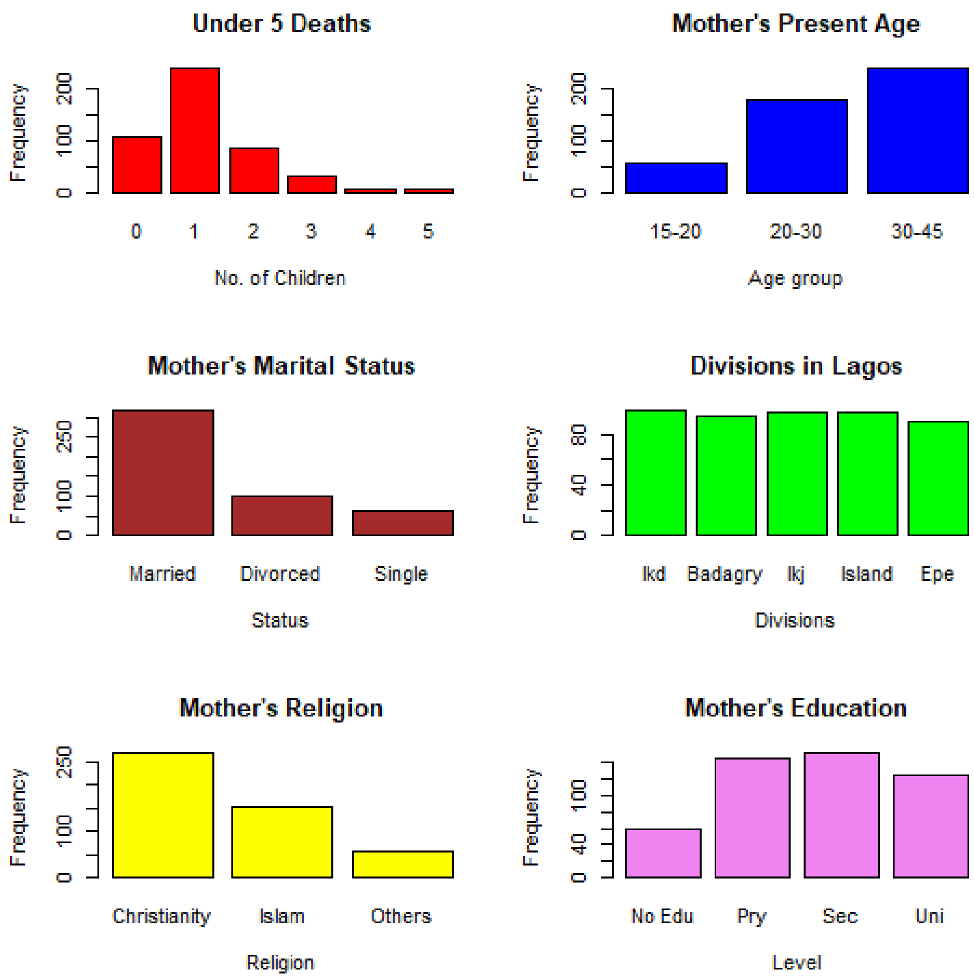
World Population Review (2016). Accessed on 22nd December 2016. Avaialble at worldpopulationreview.com/world-cities/lagos-population.

Zanini, R. B., Moraes, A. J., Giugliani, E. R., & Riboldi, J. (2009). Infant mortality trends in the state of Rio Grande do Sul, Brazil, 1994-2004: a multilevel analysis and community risk factors. *Cad. Saude Publica, Rio de Janerio, 25*(5), 1035-1045.

Zhu, B. P., Lemeshow, S., Hosmer, D. H., Klaar, J., Avrunin, J., & Teres, D. (1996). Factors affecting the performance of the models in the Mortality Probability Model II system and strategies of customization: a simulation study. *Crit Care Med, 24*(1), 57-63.

Appendix

Appendix I. Under 5 Deaths and Mothers Characteristics



Appendix II. Questionnaire

QUESTIONNAIRE ON INFANT MORTALITY IN THE DIVISIONS OF LAGOS STATE (QOIM)

The aim of this questionnaire is to elicit information on the causes of infant mortality in Lagos state.

Please complete the questionnaire below as accurate as possible. All information given will be treated as confidential and used only for research purposes. The success of this study depends on your cooperation and so you are implored to respond to the items appropriately.

SECTION A: PERSONAL INFORMATION

1. SEX: MALE ( ) FEMALE ( )

2. PRESENT AGE: BELOW 20 YEARS ( ) 20-30 ( ) ABOVE 30 YEARS ( )

3. STATUS: MARRIED ( ) DIVORCED ( ) SINGLE ( )

4. DIVISION OF LAGOS OF RESPONDENT: IKORODU ( ) BADAGRY ( ) IKEJA ( ) LAGOS ISLAND ( ) EPE ( )

5. RELIGION: CHRISTIANITY ( ) ISLAM ( ) OTHERS ( )

6. EDUCATION STATUS: NO SCHOOL ( ) PRIMARY ( ) SECONDARY ( ) TERTIARY ( )

7. NUMBER OF CHILDREN IN HOUSEHOLD: 1 ( ) 2 ( ) 3 ( ) 4 ( ) 5 ( ) 6 ( ) More than 6 ( )

SECTION B: CAUSES OF INFANT MORTALITY

1. How many children have you given birth to? ____________
2. How many are still living? ___________________
3. If there is a difference between 1 and 2 above, how many died before age 5? _______
4. Please pick out of the following the cause of the death

   i. Education status of mother ( ) ii. Family size ( ) iii. Birth type ( )

   iv. Environment of mother/child ( ) v. Abortion ( ) vi. Poverty ( )

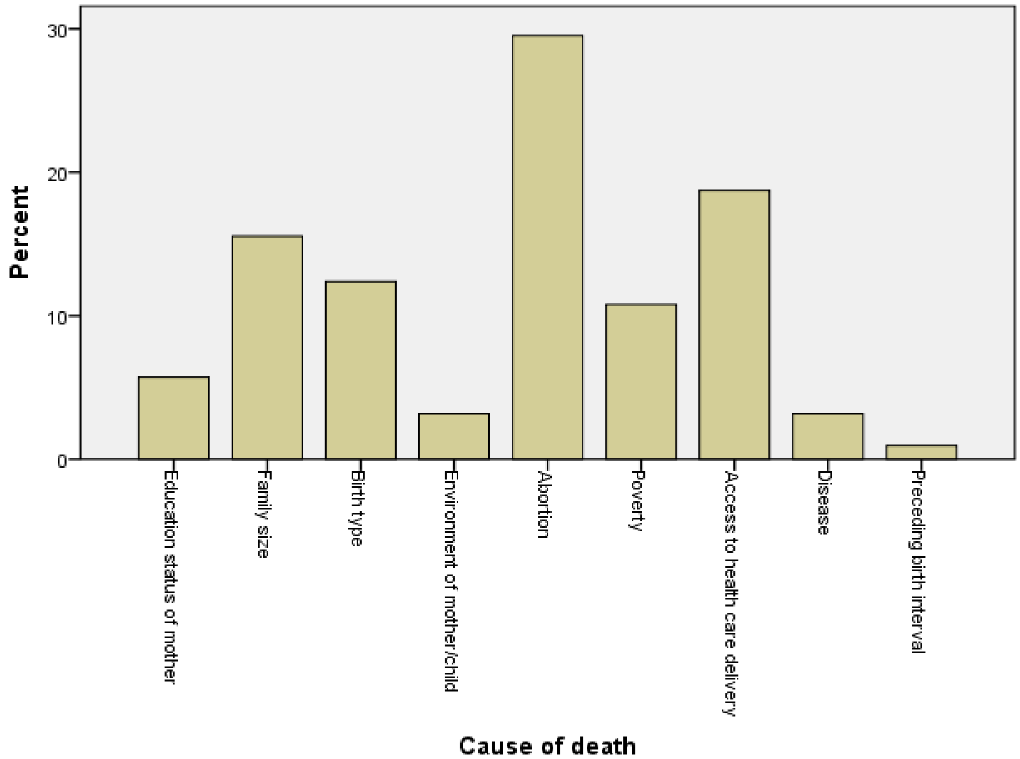
   vii. Access to health care delivery ( ) viii. Disease ( ) ix. Family history ( )

   x. Preceding birth interval ( ), xi. Breastfeeding status ( ),

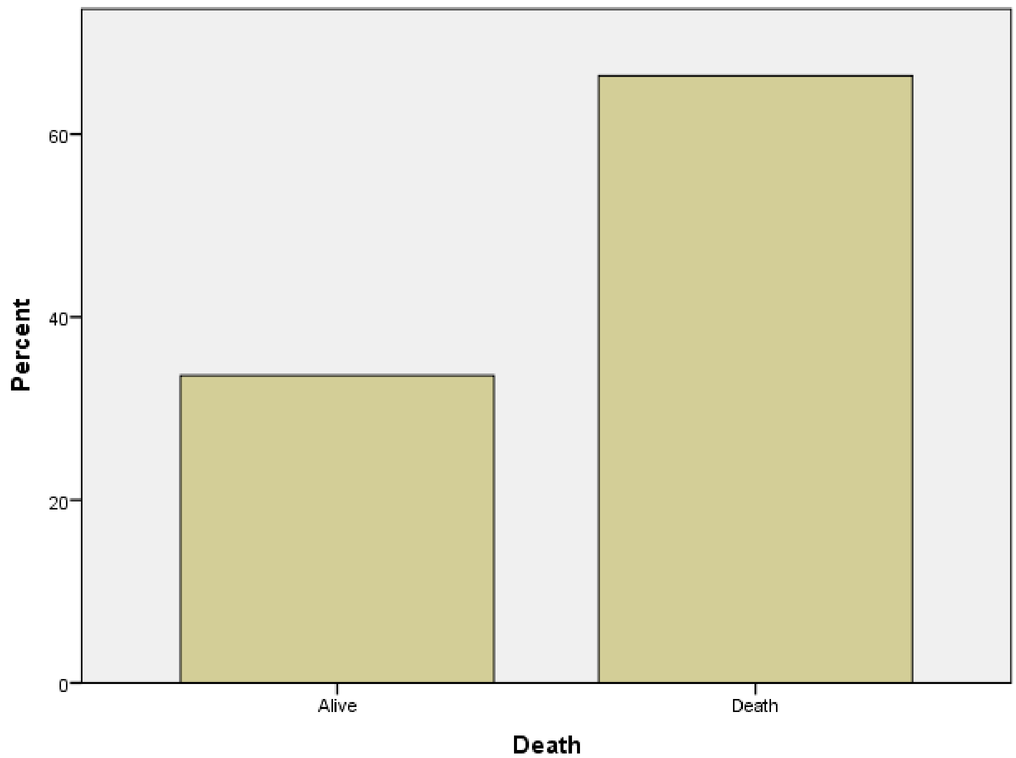
   xii. Source of drinking water ( ) xiii. Income status of household ( )

   xix. Prolong labour of mother ( ) xv. Others ( )
5. If others, please specify ____

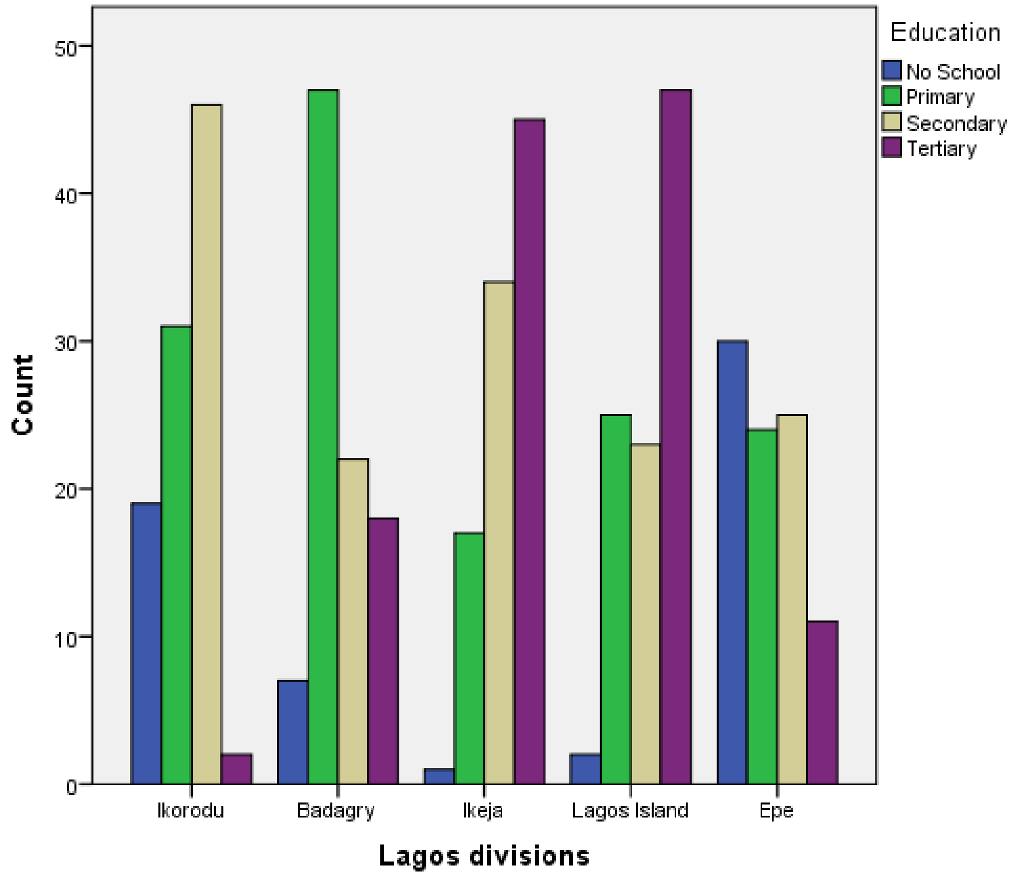
Appendix III. Causes of death



Appendix IV. Children death or alive



Appendix V. Education of Mother by Division



Appendix VI. Cause of Infant Mortality by Division