Journal of the Nigerian Association of Mathematical Physics
Volume 60 **(April – June** ***2021 Issue*****), pp141 –148**
© J. of NAMP

PREDICTION OF BRONCHO-PNEUMONIA STATUS ON INFANTS IN NIGER STATE USING BINARY LOGISTIC REGRESSION

[1]S. S. Ahmed, [1]M. T. Gana, [2]A. S. Moh'd, [1]M. A. Mustapha, [1]A. Makada and [1]I. S. Ango

[1]Department of Mathematics and Statistics, Niger State Polytechnic, Zungeru
[2]Department of Science Lab. Tech. Niger State Polytechnic, Zungeru
School of Natural and Applied Sciences (SNAS)

Abstract

Bronchopneumonia, Lobar pneumonia, Intense respiratory tract infections are one of the leading causes of infant deaths in developing countries. This study models the effects of the significant risk factors on infants' bronchopneumonia status and also fit useful model by constructing the best model with minimum number of parameters. The data for this study consist of a random sample of 735 births with low birth weight, taken from obstetrics clinic of sampled tertiary health institutions own by Government of Niger State. These include selected General Hospital from the three geo-political zones of the state. Binary logistic regression was used to identify and model the effects of the various risk factors. The predictor variables used are; Weight at birth, Baby's gender, Mother's age and Mother's occupation. The best fitting model with minimum number of parameters was identified using likelihood ratio statistic. The odds ratio of 1.308 indicate that the likelihood effect of female infant is 31% higher than that of male infant, since male stand as reference. It was observed that baby's weight at birth and mother's occupation (housewife and civil servant categories) have significant effects on infant's bronchopneumonia status. The model built can be used to predict the new cases of Bronchopneumonia status in infants.

Keywords: Bronchopneumonia, Logistic Regression, Infant, Parameter, Weight

1.0 Introduction

Bronchopneumonia or bronchialpneumonial or bronchogenic pneumonia is a type of pneumonia characterized by multiple foci of isolated, acute consolidation, affecting one or more pulmonary lobules. It is one of two types of bacterial pneumonia as classified by gross anatomic distribution of consolidation (solidification). The other being lobar pneumonia. Bronchopneumonia is less likely than lobar pneumonia to be associated with streptococcus.

The bronchopneumonia pattern has been associated with hospital acquired pneumonia, and with specific organism's staphylococcus aurous, klebsiella coli and pseudomonas. In bacterial pneumonia, invasion of the lung parenchyma by bacteria produces an inflammatory immune response. This response leads to a filling of the alveolar sacs with exudates. The loss of air space and its replacement with fluid is called consolidation.

Pneumonia is an illness, usually caused by infection, in which the lungs become inflamed and congested, reducing oxygen exchange and leading to cough and breathlessness. It affects individuals of all ages but occurs most frequently in children and the elderly. Historically, in developed countries, deaths from pneumonia have been reduced by improvements in living conditions, air quality, and nutrition. In developing world today, many deaths from pneumonia are also preventable by immunization or access to simple, effective treatments [1].

Broncho-pulmonary dysplasia (BPD) is a chronic type of lung disease prevalent among infants. This disease if present in a pregnant mother leads to low birth weight of infants at birth. It is a serious lung condition that affects infants, it mostly

Corresponding Author**:** Ahmed S.S., Email: ahmedsule710@gmail.com, Tel: +2348036154396

affect premature who need oxygen therapy (oxygen given through nasal prongs, a mask or a breathing tube). Most infants who develop BPD are born more than ten weeks before their due dates and weighs less than 2pounds (about 1,000grams) at birth, and have breathing problems. Infections that occur before or shortly after birth also contribute to low birth weight [2].
Low Birth Weight (LBW) is describe as a birth weight of a live born infant of less than 2,500g (5pounds 8ounces) regardless of gestational age. Subcategories include; Very Low Birth Weight (VLBW) which is less than 1500g (3pounds 5ounces), and Extremely Low Birth Weight (ELBW) which is less than 1000g (2pounds 3ounces). Normal Weight at term of delivery is 2500g - 4200g (5pounds 8ounces - 9pounds 4ounces). Most normal babies weight 5.5pounds by 37weeks of gestation. Intrauterine growth restriction refers to delayed growth within the uterus, which then leads to low birth weight. Some babies are just small and happen to weigh less than 5.5pounds at birth, just like some adults are smaller than others. Though this is considered low birth weight, in these cases, it is not abnormal or a cause for concern.
In modern days, statistics has played a significant role in Biological, Pharmaceutical and Medical Sciences [3]. The application of multivariate statistical techniques to biological and medical data has dominated the areas of evidence-based medicine. Multivariate methods are relevant in virtually every branch of applied medicine, pharmacy and public health. They come into play either when we have a medical theory to test or when we have a relationship in mind that has some importance for medical decision or policy analysis in public health. Multivariate methods are also used in other disciplines.
As researchers began to recognize that LBW and preterm birth is not synonymous, an uncomfortable new problem emerged. Term babies born at less than 2500 grams nonetheless have a high risk of mortality. What account for this risk, if not preterm delivery? The gap was bridged by the invention of a new disease: Intrauterine Growth Restriction (IUGR). The usual definition is "Small for Gestation Age" (SGA), the lightest 10% in each gestational age stratum. The creation of an entity called IUGR effectively preserved LBW as a group of babies with "preventable" ailments. Small babies who are not preterm are "growth restricted".
This convenient solution to the problem of term LBW infants led to the rapid acceptance of the concept of IUGR during the 1970s. According to Pub Med, the number of papers about IUGR increased rapidly between 1970 and 1979. In fact, it was not a new research area but a shift within LBW research from one label ("prematurity") to two ("Prematurity" and "IUGR").
The use of logistic regression dates back to 1845. It first appeared during the mathematical studies for the population growth at that time. The term logistic regression analysis comes from logit transformation, which is applied to the dependent variable. This case, at the same time, causes certain differences both in estimation and interpretation [4].
Logistic regression analysis is also called "Binary Logistic Regression Analysis", "Multinomial Logistic Regression Analysis" and Ordinal Logistic Regression Analysis", depending on the scale type where the dependent variable is measured and the number of categories of the dependent variable. Logistic regression is divided in to two; Univariate Logistic Regression and Multivariate Logistic Regression [5].
In many application areas, such as epidemiologic and biomedical studies, where outcomes may be occurrence or nonoccurrence, alive or dead, present or absent and so forth, logistic regression is the standard approach for the analysis of binary and categorical outcome data.
Logistic Regression Analysis (LRA) extends the techniques of multiple regression analysis to research situations in which the outcome variable is categorical. In practice, situation involving categorical outcomes are quite common. In a medical setting, an outcome might be presence/absence of disease. The focus of this study is on situations in which the outcome variable is dichotomous, although extension of the techniques of LRA to outcomes with three or more categories (e.g improved, same, or worse) is possible.
In some cases, especially those that involve the testing of medical theories, a formal statistical model is constructed. The model consists of statistical equations that describe various relationships. A bio-statistical analysis begins by specifying a statistical model. Once a statistical model has been specified, various hypotheses of interest can be stated and empirically tested in terms of the unknown biological or medical parameters. An empirical analysis requires data which are used to estimate model parameters and to formally test hypotheses of interest. In most cases, the model is used to make predictions in either the testing of a medical theory or the study of a policy's impact in pharmacy and public health [6].
Like any other model building technique, the goal of the logistic regression analysis is "to find the best fitting and most parsimonious, yet biologically reasonable model to describe the relationship between an outcome (dependent or response variable) and a set of independent (predictor or explanatory) variables" [7]. This statement motivates the purpose of this study: to identify risk factors for low birth weight (LBW) in newborn infants using the statistical tools of Logistic Regression Analysis.
According to [8], more than 20 million infants are born each year weighing less than 2500g (5.5 pounds), accounting for 17% of all births in the developing world, a rate more than double the level in industrialized countries of (7%).

The specific objectives of this study are; to construct a logistic regression model that will be used for tracking the Broncho-pneumonia in infants using four variables. To predict the Broncho-pneumonia status of some infants using the constructed model. To determine the prevalence of Low Birth Weight in Niger State.
The study will focuses on the major hospitals in Niger State. The sample taken from these health institutions would be used for the analysis to generate the required model on prevalence of BPn which lead to Low Birth Weight (LBW) in infants.
The assumptions are; the predictors are not highly correlated with each other. The mean and variance of a given predictors are not correlated. The binary logistic regression requires the dependent variables to be binary. This study assumed that the independent variable does not need to be multivariate normal. It is also assumed that, the information retrieved from the medical records was correct and most sources of error were put under control.

1.1 General Review

According to [9] in their research which involves the building of model for the prediction of Broncho-pulmonary dysplasia (BPD) for seven-day old infants. Their objective was to develop a predictive model capable of identifying which premature infants have the greatest probability of presenting BPD based on assessment at the end of the first week of life. They collected their data retrospectively from January 1998 to December 2001 and prospectively from August 2001 to July 2003. Their target population was infants born with less than three weeks of birth and weighing less than 1500g. They employed logistic regression for the data analysis. Four variables maintained a significant relationship with the outcome and were used to construct the formula to calculate the probability of BPD.
A study carried out by [10] on Multiple Logistic Regression analysis of risk factors in elderly pneumonia patients, QTc interval prolongation as a prognostic factor. Acute pneumonia is a serious problem in the elderly and various risk factors have already been reported, but they notice that the involvement of QTc interval prolongation remains uncertain. The aim of their study was to elucidate the prognostic factors for the development of pneumonia in elderly patients and to study the possible involvement of QTc interval prolongation. About 249 hospitalized pneumonia patients were captured at Aki-Ohta Hospital from January 2010 to December 2013, out of which the patients more than 65 year old were included, with 178 Community-Acquired Pneumonia (CAP) patients and 71 nursing care and health care-associated pneumonia (NHCAP) patients. The pneumonia severity index, vital signs, blood chemistry data and Electrocardiogram (ECG) findings were retrospectively compared using multiple logistic regression analysis.
The research conducted to determine the prevalence of Low Birth Weight (LBW) and some of its risk factors in Wushishi LGA of Niger State using Logistic Regression model. They observed that there is no significant difference in prevalence between boys and girls (14.9% versus 13.9%) i.e. $p=0.578$ [11].
A study conducted by [12] to determine the prevalence risk factors of Nephropathy in type-2 diabetic patients. A tertiary hospital was used for the study aimed to build a binary logistic model for predicting Nephropathy status among type-2 diabetic patients using age, sex, socio-economic status, and duration of Nephropathy history as covariates. They used a random sample of 200 patients suffering from type-2 diabetes where data on some risk factors like age, sex, socio-economic status were collected.
In the study on the effect of birth weight on infant mortality, it was found out in [13] that children born with low birth weight are more likely to die during the first year of life compared to children born with normal weight, independent of child's sex, birth order, (pregnancy care and delivery care), maternal education and nutritional status, household access to clean water and sanitation, as well as other factors.
The research carried out observed that the BPD is a chronic pulmonary disease which affects premature infants and contributes to their morbidity and mortality. Despite substantial changes in incidence, risk factors and severity after the introduction of new therapies and Mechanical Ventilation (MV) techniques, BPD remains common [14]. The maternal parity has a significant influence on the incidence of delivery of LBW infants in twin gestations. As in previous studies, the incidence was higher in primiparous compared with multiparous counterparts, suggesting that the uteri of multiparous women are more efficient in nurturing and promoting intrauterine growth of twins; accounting for the relatively lower incidence of LBW twin infants among them [15].
In a study of lifestyle behaviors, such as inadequate nutrition, smoking, mothers themselves who were low birth weight, low pregnancy weight gain, increasing maternal stress and/or depression, domestic violence and maternal regret and/or rejection of pregnancy to be significant factors [16]. Other socio-demographic factors according to [17] were low maternal age (under 18), high maternal age (over 35), low educational level, poverty, ethnicity and late or no antenatal care.
The reviewed of the impact of supplementation of a balanced protein/energy diet (where the protein content of diet was <25% of the total energy content) on gestational weight gain and pregnancy outcomes from 13 studies for the Cochrane Collaboration. The quality of the trials varied and often the method of randomization was not stated. It was found that there was an increase in material weight gain (17g/week) and a reduction in the risk of SGA birth (RR 0.68, 95% CI 0.57, 0.80)

[18]. In the study of birth weight of babies in relation to their nutrition knowledge and place of residence found (12%) prevalence of LBW similar to national average of 12% and to 12.64%, 11.4% and 12.6% reported elsewhere. This is much higher than 8.2% reflecting the worsening economic situation in the present day Nigeria and a frightening future trend in the face of unabated current global food crises and economic meltdown [19].
The discriminant analysis and classification are multivariate techniques concern with separating district sets of objects (or observations) and with allocating new objects (or observations) to previously defined groups. Discriminant analysis is rather exploratory in nature. As a classificatory procedure, it is often employed on a one-time basis in order to investigate observed differences when casual relationships are not well understood. Classification procedures are less exploratory in the sense that they lead to well-defined rules, which can be used for assigning new objects. [20]. In their research [21] observed that breast cancer is the most common type of cancer in women, while the mortality rate of breast cancer of females over 40 years old is extremely high. If detected early, it can be treated early, and the mortality rate of breast cancer can be reduced. Therefore, the image processing technologies has been adapted to automatically breast images, select suspicious regions, and provide alerts to assist in doctors' diagnosis, reduce misdiagnosis rates due to fatigue of doctors, and improve diagnostic accuracy.
According to [22] in a similar study on proportion of low birth weight babies due to small for gestational age revealed that the incidence of LBW in south – south Nigeria was 10.1% similar to the incidence of 10.31% in Enugu, but lower than the 19.8% reported in Kano city (North West Nigeria). In South West Nigeria, the incidences ranged from 8.2% to 16.8% while in Plateau (North Central Nigeria).

2.0 Materials and Methods

This research carried out focuses on the methods and collection of data from the study area. It is necessary to critically study our methods and procedures as a precondition for achieving the desired goals. However, the research will definitely explore the prediction power of the logistic model as regards to the proper applications of biomedical modeling and to compare same for classifying the Broncho-pneumonia (BPn) status of infants. The variables considered in this research are; Baby's weight at birth, Baby's sex, Mother's age and Mother's occupation are medically adequate to elucidate the difference between a normal and Broncho-pneumonia (BPn) patient.
In this study, we shall particularly build the model. However, in classification design, the major statistical components form the basis of the research design which includes both the sampling plan and the modeling procedures. The sampling plan is the methodology used for selecting the sample from the population. The modeling procedure is the algorithms or formulae used for obtaining models of population values from the sample data and for estimating the reliability of this model.

2.1 Study Design

In order to achieve the research objectives in this study, this work is a combination, both in purpose and in design of classification analysis using baby's weight at birth, sex, mother's age and mother's occupation. In the classification design, the researcher is not interested in a mere collection of facts but models would be used to classify the BPn status of an infant whose BPn status is not known earlier.
Two variables were categorized (Sex and Mother's occupation); in sex variable, male is coded as 0 and female as 1. In Mother's occupation, House wife coded 0, Civil servant as 1 and 2 for Business mother.
The four selected predictors mentioned above are capable of characterizing Broncho-Pneumonia status in infants. These variables are believed to vary significantly between Healthy and Unheaithy (BPn) infants. They are abbreviated as follows;

$x_{bw_1} = baby's\ weight\ at\ birth\ (kg)$

$x_s = sex, where\ x_s\ is\ coded\ as\ 1\ for\ male\ and\ 2\ for\ female$

$x_{ma} = mother's\ age$

$x_{mo} = mother's\ occupation\ coded\ as\ 0\ for\ house\ wife, 1\ for\ civil\ servant$
$and\ 2\ for\ business\ mother$

This study was conducted on the available dataset from Niger State General Hospitals. The population is infinite as infants are given birth to on daily basis. Any baby with low birth weight is a potential BPn which is mainly caused through pneumonia in infant; hence, the population size cannot be specified at any point in time. A total of 730 infants' files were inspected from the General Hospitals in each geo-political zones of the state. \

2.2 Data Collection

In our research, data were carefully and technically extracted directly from the individual patient's medical folder from the randomly selected health institutions within the state, the baby's weight at birth (kg), baby's gender, Mother's Age and Mother's Occupation data were collected and tabulated as independent variables.
It is necessary to outline the framework of the major components involved in the sampling design and data collection

procedures adopted in this research. Each of these has at least two health institutions (General Hospital). The simple random sampling (SRS) scheme with size n=2 was used for data collection.

3.0 Methodology

This study adopts logistic regression model since there are only two possible outcomes. The value 0 is used to represent a "success" or the outcome we are not interested in, and 1 represents a "failure". The mean of the dichotomous random variable Y, designated p, is the proportion of times that it takes the value 0. Equivalently;

$P_{HS} = P(HS = 0) = P(success)$, where HS stand for Health Status

The interest is to estimate the probability (p) associated with a dichotomous response (which, of course, is also its mean) for various values of an explanatory variable.

The model used is of the form;

$$Y_{HS=g(x_i)+\varepsilon_i} \qquad (1.1)$$

Where; $g -$ *is the regression coefficient for each of* x_i *predictor variable and* $\varepsilon_i -$ *is the error of the predictor.*

With $Y_{HS} \in (0,1)$

The strategy is to fit a model of the form;

$$P_{HS} = \beta_0 + \beta_i X_i + \varepsilon, \qquad (1.2)$$

$where\ X_i\ for\ i = 1, - - -, 4\ denoted\ as\ (BW,\ BS,\ MA\ and\ MO)$

This is simply the standard linear regression model in which X_i - represent the explanatory variables and Y - the outcome of a continuous normally distributed random variable, which been replaced by P_{HS}, β_0 is the intercept and β_i is its slope. However, this model is unfeasible. Since P_{HS} is a probability, it is restricted to taking values between 0 and 1. The term $\beta_0 + \beta_i x_i$, in contrast, could easily yield a value that lies outside this range. This can be resolved by fitting the model;

$$P_{HS} = e^{\beta_0 + \beta_i X_i} \qquad (1.3)$$

This equation guarantees that the estimate of P is positive. This model is also unsuitable. Although the term $e^{\alpha + \beta_i \chi_i}$ cannot produce a negative estimate of P, it can result in a value that greater than1.

To curtail this problem, we fit a model of the form;

$$\frac{P_{HS}}{1-P_{HS}} = e^{\beta_0 + \beta_i X_i} \qquad (1.4)$$

So that we have;

$$P_{HS} = \frac{e^{\beta_0 + BX_i}}{1+e^{\beta_0 + BX_i}} \qquad (1.5)$$

This expression above is called a logistic function (logit model), cannot yield a value that is either negative or greater than 1; consequently, it restricts the estimated value of P_{HS} to the required range.

The Logistic regression model for the dependence of p_i(response probability) on the values of k explanatory variables x_1, x_2, - - - x_k is given as;

$$LLogit(P_{HS}) = Log\left(\frac{P_{HS}}{1-P_{HS}}\right) = \beta_0 + \beta_1 X_{BW} + \beta_2 X_S + \beta_3 X_{MA} + \beta_4 X_{MO} \qquad (1.6)$$

Or $$P_{HS} = \frac{exp^{(\beta_0 + \beta_1 X_{BW} + \beta_2 X_S + \beta_3 X_{MA} + \beta_4 X_{MO})}}{1+exp^{(\beta_0 + \beta_1 X_{BW} + \beta_2 X_S + \beta_3 X_{MA} + \beta_4 X_{MO})}} + \varepsilon \qquad (1.7)$$

Where;

P_{HS} *denote response probability(Health Status)*
X_{BW} *denote Baby's weight at birth*
X_S *denote baby's Sex*
X_{MA} *denote Mother's Age*
X_{MO} *denote Mother's Occupation*
ε *is the error of the predictor.*

and $\left(\frac{P_{HS}}{1-P_{HS}}\right)$ is the ratio of the probability of a failure and called odds, $\beta_0, \ldots, \beta_4$ are parameters to be estimated.

In logistic model the predicted values will lie between 0 and 1 regardless of the values of the explanatory variables.

3.1 Chi-square test

In this research work, the sex and mother's occupation variables used are categorical variables. The sex has two categories, male and female, and mother's occupation has three which are house wife, civil servant and business mother. Hence, it is important to test, during the modeling process, whether BPn infection depends on sex and/or mother's occupation which is better analyzed by the chi-square statistic.

Hypothesis for Chi-square Test:

H_0**:** The two variables are independent

H_1**:** The two variables are not independent

Test statistic:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-e_{ij})^2}{e_{ij}} \qquad (1.8)$$

where O_{ij} is observed value and e_{ij} is expected value.

$P = \Pr\left(\chi^2_{\upsilon} > \chi^2_0\right)$

Decision Rule:

Reject H_0 if $p<0.05$ otherwise accept H_0 at the 5% level of significance.

Reason for the Use of Chi-square Test

One of the reasons for the use of chi-square is to investigate the inter-dependency among sex and mother's occupation against BPn. The Chi-square test of independence is the appropriate statistical tool for the investigations of such inter-dependency.

3.2 Omnibus Chi-Square Test

The Omnibus Chi-square test is a log-likelihood ratio test for investigating the model coefficients in logistic regression. The test procedures are as follows:

Hypothesis for Omnibus Chi-square test:

H_0**:** The model coefficients are not statistically significant

H_1**:** The model coefficients are statistically significant

Test statistic:

$$\chi^2 = 2\left[\sum_{i=1}^{r}\sum_{j=1}^{c} O_{ij}\ln\left(\frac{O_{ij}}{e_{ij}}\right)\right] \qquad (1.9)$$

where;

O_{ij} *denote observed values and* e_{ij} *denote expected values*

Decision Rule: Reject H_0 if $p < 0.05$ otherwise accept H_1 at the 5% level of significance.

where; $P = \Pr\left(\chi^2_{\upsilon} > \chi^2_0\right)$

χ^2_0 - - - chi-square calculated and χ^2_v - - - chi-square value with vdf

Omnibus test is used to investigate the significance of the model coefficient in logistic model.

4.0 Analysis, Results and Discussions

In this chapter, the data are fitted to the logistic regression model. The results of the analyses are presented and discussed. The data were analyzed using SPSS version 21.0

Logistic regression deals with the binary cases, where the response variable consists of just two categorical values. Logistic regression model is mainly used to identify the relationship between two or more explanatory variables (X_i) and the dependent variable (Y)**.**

Table 1.1: Variables in the Equation for the Sample Data

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	95% C.I.for EXP(B) Upper
Step 1[a]	BWAB	-4.411	.336	172.144	1	.000	.012	.006	.023
	SEX(1)	.268	.187	2.063	1	.151	1.308	.907	1.885
	MA	-.018	.016	1.283	1	.257	.982	.951	1.014
	MO			6.101	2	.047			
	MO(1)	-.575	.240	5.753	1	.016	.562	.352	.900
	MO(2)	-.508	.260	3.807	1	.051	.602	.361	1.002
	Constant	9.076	.873	108.006	1	.000	8743.095		

a. Variable(s) entered on step 1: BWAB, SEX, MA, MO.

Table 1.1 shows the p-value of 0.151 > 0.05, this implies that there is no difference in the effect of male and female. The odds ratio of 1.308 indicate that the likelihood of effect of female is 31% higher than that of male, since male stand as reference. In mother's occupation, the outputs of the p-values of the first category is less than significant level (i.e. 0.016<0.05), this indicate that there is significant difference in the effect and the likelihood of effect of Civil servant is lower than the House wife effect by 44%. And the second category (Business mother) has the odds ratio of 0.602 indicate 39.8% effect lower than reference or house wife. We then conclude that there is no significant difference in the effect of the two categories (i.e. House wife, and Business mother) since 0.051>0.05.

From equation (1.7), we obtain the logistic regression model as follows:

$$P_{HS} = \frac{e^{9.076-4.411X_{BW}+0.268X_S-0.018X_{MA}-0.542X_{MO}}}{1+e^{9.076-4.411X_{BW}+0.268X_S-0.018X_{MA}-0.542X_{MO}}} \quad (1.10)$$

Alternatively, equation (1.6) becomes;

$$ln\left(\frac{\hat{p}_{HS}}{1-\hat{p}_{HS}}\right) = e^{\hat{\beta}_0+\hat{\beta}_1X_{BW}+\hat{\beta}_3X_S+\hat{\beta}_4X_{MA}+\hat{\beta}_5X_{MO}+\varepsilon_{ij}}$$

$$ln\left(\frac{\hat{p}_{HS}}{1-\hat{p}_{HS}}\right) = e^{9.076-4.411X_{BW}+0.268X_S-0.018X_{MA}-0.542X_{MO}} \quad (1.11)$$

The logistic model (1.10) will be used to predict the affected status of infants using a cut value or threshold probability of 0.5.

Table 1.2: Omnibus Tests of Model Coefficients

		Chi-square	Df	Sig.
Step 1	Step	301.283	5	.000
	Block	301.283	5	.000
	Model	301.283	5	.000

The table 1.2 indicates the omnibus test for the parameter convergence at the final stage of the iteration. Here the chi-square is highly significant (chi-square=301.283 with df=5 and p<0.05)

Table 1.3: Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	10.977	8	0.203

This table suggest that, the model is a good fit to the data since p=0.203>0.05. However, the chi-square statistic shows that the BPn depend on the categorical data.

Table 1.4: Classification Results

			Predicted HS		
		Observed	healthy	unhealthy	Total
Step 1	HS	Healthy	357(86.2%)	57(13.8%)	414(100%)
		Unhealthy	111(34.6%)	210(65.4%)	321(100%)

Table 1.4: shows the practical result of using logistic regression model of all the cases used. The dataset comprises of 735 cases were used in creating the model. In the selected cases, 357 of 414 i.e. 86.2% of healthy patients were perfectly classified while, 13.8% were misclassified. 65.4% of unhealthy patients were correctly classified 34.6% misclassified.

Table 1.5: Correlation Matrix

		Constant	BWAB	SEX(1)	MA	MO(1)	MO(2)
Step 1	Constant	1.000	-.821	-.104	-.578	-.319	-.243
	BWAB	-.821	1.000	.006	.077	.151	.098
	SEX(1)	-.104	.006	1.000	-.004	-.002	.012
	MA	-.578	.077	-.004	1.000	.041	-.001
	MO(1)	-.319	.151	-.002	.041	1.000	.621
	MO(2)	-.243	.098	.012	-.001	.621	1.000

Table 1.5 show that the data set satisfy the assumptions of logistic regression which stated that, the predictors are not highly correlated with each other (that is, baby's weight at birth and baby's sex), which has the correlation as 0.006. The correlation between the self-predictor is constant across group.

5.0 Conclusion and Recommendations

From the analysis carried out it has been observed that the Logistic Model fit the data set. In comparing the outcome of the analyses, it was discovered that 77.7% of the random selected cases were correctly classified using the model built from the dataset.

In the outcome of the results, it shows that there was no perfect correlation among the independent variables. It is also discovered that the model indicate the baby's weight at birth make the highest contribution to the logistic function. While, the mother's age have low contributions.

The developed Logistic Model classified 77.7% random selected cases whose Broncho-pneumonia status is already known which prove to be very good.

One of the objectives of this research is to construct the logistic regression model that is capable of tracking BPn infants based on their variables used. The transcription and experimental method of data collection was used in this study and the research was carried out in selected General Hospitals in Niger State. The data sampled were gathered and tabulated with 735 cases of low birth weight.

The researcher recommended that the model developed in this study could assist the doctors and other health practitioners to detect and monitor the prevalence and control of BPn among infants. It is also recommended that larger sample size and health facilities be used in further study. State Government should encourage the use of the model built in this research to aid in discovering the prevalence of BPn among infants so that adequate measures for prevention and control of Broncho-Pneumonia can be taken early enough. And use of other statistical package especially those dedicated to multivariate analysis on this area in order to elucidate intensive information or results.

References

[1] Anthony, J., Brooks, W. A., Malik J. S., Douglas, H. and Kim, M. E. (2010). Pneumonia research to reduce childhood mortality in the developing world. *The Journal of Clinical Investigations*. Vol. 2, 120-128.

[2] Jobe, A. H. and Bancalari, E. (2001). Broncho-pulmonary dysplasia. *American Journal of Respiratory and Critical Care Medicine.* Vol. 163(7): 1723 - 1729.

[3] Cornfield, A. (2010). Binary logistic model for predicting Myocardia infection. *American Journal of Public Health*, Vol. 81: 15-21.

[4] Anderson, A.A. (2008). Tracking Broncho-pulmonary dysplasia (BPD) using the logistic regression model. *International Journal of Statistics* Vol. 65: 59-66.

[5] Stephenson, B. (2008), Binary Response and Logistic Regression analysis. http://www.public.iastate.edu/~stat415/Stephenson/stat415 chapter3.pdf

[6] Rencher, A. C. (2002). Method of Multivariate Analysis (2nd ed.) John Wiley & Sons. New York.

[7] Hosmer, D. W. and Lemeshow, S. (1989). Applied logistic regression. New York: Wiley.

[8] UNICEF, (2008) Global Database on Low Birth Weight. Low Birth Weight Incidence by Country (1996-2006).

[9] Carlos, A.B, Christieny, C.M, Moreira, E.L.M, Jose, R.R and Guilherme, M. S. (2007). Binary logistic model for predicting Broncho- Pulmonary Dysplasia (BPD). *J. Pediar (Rio) 200,* Vol. 83(2):163-170.

[10] Yasuyuki T, Gen T, Miki O, Akihisa S. and Takeshi Isobe (2014). Multiple Logistic Regression Analysis of risk factors in elderly pneumonia patients: Multidisciplinary Respiratory Medicine; Vol. 9(59).

[11] Danbaba A, Audu M, Mahmud A. M (2013). Investigation of risk factors of low birth weight using multivariate logistic regression analysis: *Journal of humanities, science and technology;* Vol. 3(2): 69-77.

[12] Vitmalkumar, V.K, R.C Moses, Padmanaban S. (2011). Binary logistic model for detecting Prevalence & Risk Factor of Nephropathy. In Type 2 Diabetic patients. *International Journal of collaborative Research on internal medicine and public Health* Vol. 3(8): 598-615.

[13] Uthman, O. A. (2008). Effect of Low Birth Weight on Infant Mortality; *the Internet Journal of Epidemiology.* Vol. 6(1).

[14] Tapia, J. I., Agost, D., Alegria, A. and Standen, J. (2006). Broncho-pulmonary Dysplasin, Incidence, Risk Factors and Resource Utilization in a Population of South America Very Low Birth Weight Infants. *Journal de pediatric*. Vol. 82(1): 15-20.

[15] Onyiriuka A. N. (2010). Incidence of Low Birth Weight Infants in Twin Gestations. *Niger J Clin Pract;* Vol. 13: 365 – 70.

[16] Rees, G., Brook, Z., Doyle, W ., and Costeloe, K.(2005). The Nutritional Status of Women in the First Trimester of Pregnancy attending an inner-city Antenatal Department in the UK. *Journal of Royal Society for the Promotion of Health;* Vol. 125(5): 232 – 238.

[17] Lasker, J., Coyle, B., Kuang, L, and Ortynsky, M. (2005). Assessment of risk factors for Low Birth Weight deliveries. *Health Care of Women International;* Vol. 26: 262 – 280.

[18] Kramer, M.S. (2002). Balanced protein/energy supplementation in pregnancy. *Cochrane Database Sys Rev. Issue 1.*

[19] Ngwu, E. K. (2009). Birth weight of babies in relation to their mothers' nutritional knowledge and place of residence. *Nigerian Journal of Nutritional Science.* Vol. 30(2): 36 – 42.

[20] Vishwa, N. M., Madaki, U. Y., Vijay, V. S. and Babagana, M. (2015). Application of Discriminant Analysis on Bronch-Plumonary Dysplasia among Infants: *American Journal of Theoretical and Applied Statistics* Vol. 4(2): 44-51.

[21] Yu-shun, C., Chiun, L. and Wang, K. (2011). Based on Fuzzy Linear Distribution Analysis for Breast Cancer Mammography Analysis. *doi.ieeecomputersociety.org*

[22] Eneh, A.U. and Ugwu, R.O. (2011). The Proportion of Low Birth Weight Due to Small for Gestational age (SGA) and Prematurity. *The Internet Journal of Paediatrics and Neanatology,* Vol. 13(1).