



ISSN 2278 – 0211 (Online)

Prevalence of Bronchopulmonary Dysplasia among Infants Using Logistic Regression Model for UMTH and UDUTH Hospitals in Nigeria

Umar Yusuf Madaki

Lecturer, Department of Mathematics and Statistics, Yobe State University, Damaturu, Nigeria

Abdulrahman Malik

Lecturer, Department of Statistics, Mai Idris Aloomo Polytechnic Geidam, Yobe State, Nigeria

Maigana A. Bakawu

Lecturer, Department of Statistics, Mai Idris Aloomo Polytechnic Geidam, Yobe State, Nigeria

Abstract:

In this research we investigated the incidence of Bronchopulmonary Dysplasia among Infants using modeling approach. The data used were obtained from University of Maiduguri Teaching Hospital (UMTH), Maiduguri and Usmanu Danfodiyo University Teaching Hospital (UDUTH) Sokoto with sample size of Seventy (70) patients in 2014. Fifty (50) patients were from Maiduguri and Twenty (20) patients from Sokoto. Logistic regression model was employed for the analysis with the help of SPSS. The result of the analysis indicates that the logistic regression model has a perfect classification of five new cases in Sokoto while it has misclassified two Of five new cases in Maiduguri. This result shows that the prediction of Broncho-Pulmonary Dysplasia is moderately done with logistic model in Maiduguri and is more preferred for the same purpose in Sokoto. The study recommends that Doctors and Clinics should adopt the use of the model built by this research to detect the prevalence of BPD among Infants.

Keywords: Logistic regression, Bronchopulmonary Dysplasia (BPD), infants, Maiduguri, Sokoto

1. Introduction

The original definition of Bronchopulmonary dysplasia by Northway was based on radiological and pathological characteristics in prematurely born infants with respiratory distress syndrome (RDS) who were treated with mechanical ventilation and oxygen supplementation. Subsequently, the definition of BPD was changed to respiratory squeal in infants requiring oxygen supplementation more than twenty eight (28) days after birth since BPD may occur in tiny premature infants who have not previously had RDS. Bronchopulmonary dysplasia is a chronic lung disease associated with premature birth and characterized by early lung injury (Bhandari & Bhandari, 2009). The Broncho pulmonary dysplasia is a form of chronic lung disease that develops in preterm neonates treated with oxygen and positive-pressure ventilation. The pathogenesis of this condition remains complex and poorly understood; however various factors can not only injure small airways but also interfere with alveolarization (alveolar septation), leading to alveolar simplification with a reduction in the overall surface area for gas exchange. The developing pulmonary microvasculature can also be injured. Many infants born with Bronchopulmonary dysplasia exhibit signs and symptoms of respiratory distress syndrome, including the following: Tachypnea, Tachycardia, Increased respiratory effort (with retractions, nasal flaring, and grunting), frequent desaturations. These infants are often extremely immature, have a very low birth weight, and have significant weight loss during the first 10 days of life. Their requirements for oxygen and ventilatory support often increase in the first 2 weeks of life. At weeks 2-4, oxygen supplementation, ventilator support, or both are often increased to maintain adequate ventilation and oxygenation.

1.1. Objectives of the Study

The following are specific objectives:

- To build a logistic regression model that is capable of tracking Broncho-pulmonary dysplasia (BPD) infants based on their weight at birth, weight four weeks later and gender.
- To predict the Bronchopulmonary dysplasia (BPD) of some infants using the logistic regression model.

2. Literature Review

Literature review is a collection of other relevant research works that will provide bases for present research. The collection, collation and coordination of related literature are of paramount importance in every study. It is pertinent to begin every research work, particularly in this kind of statistical modelling, by outlining how other relevant literatures were consulted. A review only of those items relevant to the work has been made in this section, which has an immediate bearing to this work at hand (Vimalkumar, Moses & Padmanaban, 2011). The Bronchopulmonary Dysplasia (BPD) in the recent years on medical pediatric research is the most common cause of morbidity among extremely low birth weight (ELBW) newborns, although the incidence, risk factors and severity of the disease have changed substantially since the introduction of new treatments and scientific techniques (Press & Wilson, 1978). Moreover, (Vimalkumar, Moses & Padmanaban, 2011) conducted a study to determine the prevalence risk factors of Nephropathy in type-2 diabetic patients. Their aim was 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. The authors further discovered that as the duration of type-2 diabetes increases, the incidence of Nephropathy also increases significantly. Hence, all the type-2 diabetic patients, especially those with increased duration should be screened for Nephropathy and be made aware of the complications. (Bhering, Mochdece, Moreira, Rocco, & Sant'Anna, 2007) conducted a research while involves the building of model for the prediction of Bronchopulmonary dysplasia model for seven-day old infants. Their aim 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, the authors concluded that at the end of the first week of life, the predictive model they developed was capable of identifying newborn infants at increased risk of developing BPD with high degree of sensitivity (Oh et al., 2005). On a similar note, Proposed that adaptive control effects of exercise on glycemic control and body mass in type 2 diabetes mellitus is generally access by clinical trials. An abnormal pulmonary vasculature may be an important component of Bronchopulmonary dysplasia (BPD). The infant's lung has been examined for the endothelial cell marker and for angiogenic factors with their respective clinical receptors (Bhatt et al., 2001).

3. Materials and Methods

In this research design, having considered all the factors involved, the simple random sampling is the chosen sampling design.

Consider the three selected predictor variables which are capable of characterizing a BPD infant. From experience and records of medical practice, these variables are also believed to vary significantly between normal infants and BPD infants. The variables are;

X_1 =age

X_2 = sex, where x_2 is coded as 1 for male and 0 for female

X_3 = Birth weight (g)

Other variables are:

X_4 =socio-economic status

X_5 =duration of Nephropathy history

3.1. Logistic Regression

Logistic regression models are adequate for those situations where the dependent variable of the regression problem is binary. That is, the dependent variable has only two possible outcomes, e.g., "success/failure" or "normal/abnormal". This model may be used for classifying an object into one of two populations by letting E denote the event that the object belongs to the first population, and letting x denote a profile vector of attributes of the object to be classified (Rencher, 2002). We assume that these binary outcomes are coded as 1 and 0 in accordance with the claim that the application of linear regression models to such problems would not be satisfactory since the fitted predicted response would ignore the restriction of binary taking on values for the observed data as noted by (Hosmer Jr, Lemeshow & Sturdivant, 2013).

When studying linear regression, we attempted to estimate a population regression equation;

By fitting the model of the form;

$$Y_i = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \dots \dots \dots (1)$$

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_k X_k \dots \dots \dots (2)$$

The response Y was continuous, and was assumed to follow a normal distribution. We were concerned with predicting or estimating the mean value of the response corresponding to a given set of values for the explanatory variable. Another property of logistic regression relates to situations in which the explanatory variable X rather than the response variable Y is random (Agresti, 2018).

In this case we shall only consider a simple logistic regression that is, logistic regression models with one explanatory variable. Our first strategy might be to fit a model of the form:

$$p = \alpha + \beta x \dots \dots \dots (3)$$

This is simply the standard linear regression model in which x represents the explanatory variable and y – the outcome of a continuous, normally distributed random variable – has been replaced by (Menard, 1995). As before, α is the intercept and β is its slope. On inspection, however, this model is not feasible. Since p is a probability, it is restricted to

taking values between 0 and 1. The term $e^{\alpha + \beta x}$ in contrast, could easily yield a value that lies outside this range. Instead we might try to solve this problem by fitting the model;

This equation guarantees that the estimate of p is positive. We would soon realize, however, that this model is also unsuitable. Although the term cannot produce a negative estimate of p , it can result in a value that is greater than 1. To accommodate this final constraint, we fit a model of the form;

$$p = e^{\alpha + \beta x} \dots\dots\dots(4)$$

This expression on the right, 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 to the required range.

$$p = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}} \dots\dots\dots(5)$$

$$\frac{p}{1-p} \text{ to } 1$$

Thus, if a success occurs with probability

$$p = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}$$

the odds in favour of success are

$$\frac{p}{1-p} = \frac{e^{\alpha + \beta x} / (1 + e^{\alpha + \beta x})}{1 / (1 + e^{\alpha + \beta x})} = e^{\alpha + \beta x} \dots\dots\dots(6)$$

Taking the natural logarithms of each side of this equation

$$\ln \left(\frac{p}{1-p} \right) = \ln(e^{\alpha + \beta x}) = \alpha + \beta x \dots\dots\dots(7)$$

Where; the link function is;

$$\ln \left(\frac{p}{1-p} \right)$$

Thus, modeling the probability p with a logistic function is equivalent to fitting a linear regression model in which the Continuous response y has been replaced by the logarithms of the odds of success for a dichotomous random variable. [12]. Instead of assuming that the relationship between p and x is linear, we assume that the relationship between the link function and x is linear.

From the technique of fitting a model of this form is known as logistic regression; the estimated relationship between the explanatory variable and the odd in favour of success is given below:

$$\ln \left(\frac{\hat{p}}{1-\hat{p}} \right) = \hat{\alpha} + \hat{\beta} x \dots\dots\dots(8)$$

However, we cannot apply the method of least squares, which assumes that the response is continuous and normally distributed, to fit a logistic regression model [7]; instead we use the method of maximum likelihood. The method of maximum likelihood uses the information in a sample to find the parameter estimates that are most likely to have produced the observed data [8].

3.2. The Logistic Regression Model

The logistic model is constructed for UMTH, Borno and UDUTH, Sokoto states respectively using the IBM SPSS VERSION 22 with the following output results:

		Frequency	Parameter Coding
Sex	Male	22	1.000
	Female	28	0.000

Table 1: Binary Logistic MODELUMTH, (Maiduguri) Categorical Variables Coding

The table1, above indicate that 22 of cases used in the analysis are male while the remaining 28 are female.

				Predicted		
				Status		Percentage Correct
				Healthy	MCP	
Step 0	Observed	Status	Healthy	27	0	100.0
			BPD	23	0	0.0
		Overall Percentage				54.0

Table 2: Classification Table FORA Cut Value Of 0.5

Using the estimated coefficient of the first iteration; all the 50 infants were classified healthy (code 0). Hence implying that the first iteration does not really gives the required model. Hence we continue. The table 2, above indicate that 27 of cases used in the analysis are healthy while the remaining 23 cases are BPD infants.

Iterations	-2 Log likelihood	Coefficients			
		Constant	W1	W2	Sex
1	47.868	3.974	-0.034	0.021	-0.651
2	46.319	5.405	-0.058	0.037	-0.753
3	46.243	5.808	-0.065	0.042	-0.773
4	46.243	5.835	-0.066	0.042	-0.773
5	46.243	5.835	-0.066	0.042	-0.773

Table 3: Iteration History

		-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
Step	1	46.243	0.636	0.788

Table 4: Model Summary

The table above indicates the justification of the goodness of fit test for the model

				Predicted		
				Status		Percentage Correct
				Healthy	MCP	
Step 0	Observed	Status	Healthy	21	6	77.8
			BPD	5	18	78.3
		Overall Percentage				78.0

Table 5: Classification Table for a Cut Value of 0.5

The table 5, above indicates the classification powers of the model.

		B	S.E.	Wald Sig.	df	Exp(B)
Step 1	W1	-0.066	0.057	0.000	1	0.936
	W2	0.042	0.039	0.000	1	1.043
	Sex	-0.773	0.965	0.000	1	0.461
	Constant	5.835	2.067	0.000	1	342.235

Table 6: Variables in the Equation

The table 6, above indicates parameter estimate of the model. From the table above, using the Wald's statistic, all the coefficients are statistically significant and the logistic regression is obtained as follows:

$$\hat{\alpha} = 5.835$$

$$\hat{\beta}_1 = -0.066 \quad \hat{\beta}_3 = -0.773$$

$$p = \frac{e^{\hat{\alpha} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3}}{1 + e^{\hat{\alpha} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3}} \dots \dots \dots (9)$$

Where;

$$p = \frac{e^{5.835 - 0.066x_1 + 0.042x_2 - 0.773x_3}}{1 + e^{5.835 - 0.066x_1 + 0.042x_2 - 0.773x_3}}$$

Alternatively:

$$\ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = \hat{\alpha} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$$

$$\ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = 5.835 - 0.066x_1 + 0.042x_2 - 0.773x_3$$

$$\ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = 5.835 - 0.066x_1 + 0.042x_2 - 0.773x_3$$

To compute estimates or forecasts, consider the logistic model as given below:

$$p = \frac{e^{5.835 - 0.066x_1 + 0.042x_2 - 0.773x_3}}{1 + e^{5.835 - 0.066x_1 + 0.042x_2 - 0.773x_3}}$$

That will be used to predict the BPD status of infants using a cut value or threshold probability of 0.5. Using the logistic model and the following table containing the data of additional five infants whose BPD is already known, the model is hereby tested for goodness of fit and classificatory power.

3.3. Classifying BPD Status Using the Logistic Model

It is pertinent to use the logistic model to classify BPD status of infants in Maiduguri using the logistic model:

SNO	Weight 1 (g) (X ₁)	Weight 2 (g) (X ₂)	Sex (X ₃)	Health Status
1	1000	1400	F	Healthy
2	1550	2440	F	Healthy
3	790	1130	F	BPD
4	1480	2140	M	Healthy
5	980	1420	F	BPD

Table 7: New Set of Observations UMTH, (Maiduguri)

The predictions are as follows;

$p_1 = \frac{e^{5.835 - 0.066x_1 + 0.042x_2 - 0.773x_3}}{1 + e^{5.835 - 0.066x_1 + 0.042x_2 - 0.773x_3}} = 1$	Healthy
$p_2 = \frac{e^{5.835 - 0.066x_1 + 0.042x_2 - 0.773x_3}}{1 + e^{5.835 - 0.066x_1 + 0.042x_2 - 0.773x_3}} = 1$	BPD
$p_3 = \frac{e^{5.835 - 0.066x_1 + 0.042x_2 - 0.773x_3}}{1 + e^{5.835 - 0.066x_1 + 0.042x_2 - 0.773x_3}} = 0$	BPD
$p_4 = \frac{e^{5.835 - 0.066x_1 + 0.042x_2 - 0.773x_3}}{1 + e^{5.835 - 0.066x_1 + 0.042x_2 - 0.773x_3}} = 0$	Healthy
$p_5 = \frac{e^{5.835 - 0.066x_1 + 0.042x_2 - 0.773x_3}}{1 + e^{5.835 - 0.066x_1 + 0.042x_2 - 0.773x_3}} = 0$	Healthy

The logistic regression has misclassified two infants in this case for Borno State.

4. Discussion of Findings

The logistic model has a perfect classification of five new cases in UDUTH, Sokoto while it has misclassified two of five new cases in UMTH, Maiduguri. Hence, from this analysis, the power of a classificatory model depends on the situation, location or even the data at hand. However, from the angle of statistical inference, if normality assumption fails, the logistic models are more preferred but if the normality assumption holds, the principle of parsimony prevails – which model fits better for a particular situation must be determined using the goodness of fit results.

5. Conclusion

In this study, Logistic model were applied to Bronchopulmonary Dysplasia (BPD) data. The result shows that the prediction of Bronchopulmonary Dysplasia BPD is better done in Logistic regression model are more preferred for the same purpose in Sokoto.

6. Recommendations

In the light of the above, it is recommended that Doctors and Clinics should adopt the use of the models built by this research to detect prevalence of BPD among infants so that adequate measures for prevention and control of BPD can be taken early enough to alert the danger of the full manifestation of the disease.

7. References

- i. Agresti, A. (2018). An Introduction to Categorical data analysis. Wiley.
- ii. Bhandari, A., & Bhandari, V. (2009). Pitfalls, problems, and progress in bronchopulmonary dysplasia. *Pediatrics*, 123(6), 1562–1573.
- iii. Bhatt AJ, Pryhuber GS, Huyck H, Watkins RH, Metlay LA, Maniscalco WM (2001). Disrupted pulmonary vasculature and decreased vascular endothelial growth factor, Flt-1, and TIE-2 in human infants dying with bronchopulmonary dysplasia. *Am J Respir Crit Care Med* 2001; 164(10 Pt 1):1971-1980.
- iv. Bhering, Carlos A., Christieny C. Mochdece, Maria EL Moreira, José R. Rocco, and Guilherme M. Sant'Anna. "Bronchopulmonary dysplasia prediction model for 7-day-old infants." *Jornal de pediatria* 83, no. 2 (2007): 163-170.
- v. Hosmer Jr, D. W., Lemeshow, S., Sturdivant R. X (2013). Applied logistic regression (Vol. 398). John Wiley & sons.
- vi. Menard, S. (2000). Coefficients of determination for multiple logistic regression analysis. *The American Statistician*, 54(1), 17–24.
- vii. Oh, W., Poindexter, B. B., Perritt, R., Lemons, J. A., Bauer, C. R., Ehrenkranz, R. A., ... Wright, L. L. (2005). Association between fluid intake and weight loss during the first ten days of life and risk of bronchopulmonary dysplasia in extremely low birth weight infants. *The Journal of Pediatrics*, 147(6), 786–790.
- viii. Press, S. J., & Wilson, S. (1978). Choosing between logistic regression and discriminant analyses. *Journal of the American Statistical Association*, 73, 699–705.
- ix. Rencher, A. C, (2002). Methods of multivariate analysis (2nded) New York: John Wiley & Sons ISBN 0-471-41889-7
- x. Vimalkumar, V. K., Moses, C. A., & Padmanaban, S. (2011). Prevalence & risk factors of nephropathy in type 2 diabetic patients. *International Journal of Collaborative Research on Internal Medicine & Public Health*, 3(8), 598