PRINT ISSN 1119-8362
Electronic ISSN 1119-8362

Full-text Available Online at
https://www.ajol.info/index.php/jasem
http://ww.bioline.org.br/ja

J. Appl. Sci. Environ. Manage.
Vol. 23 (8) 1607-1614 August 2019

Binary Logistic Regression Methods for Modeling Broncho-Pneumonia Status in Infants from Tertiary Health Institutions in North Central Nigeria

[1]YAKUBU, Y; [2]AHMED, SS; [1]AUDU, I; [1]USMAN, A

[1]*Department of Statistics, School of Physical Sciences, Federal University of Technology, Minna, Nigeria*
[2]*Department of Mathematics and Statistics, School of Natural Sciences, Niger State Polytechnic, Zungeru, Nigeria*
**Corresponding Author Email:* yisa_yakubu@yahoo.com

ABSTRACT: Acute respiratory tract infections, predominantly bronchopneumonia, are one of the leading causes of infant deaths in developing countries and around the world. This work models the effects of the significant risk factors on infants' bronchopneumonia status and also fits some reduced models and determines the best model with minimum number of parameters. The data for this study consist of a random sample of 433 births to women seen in the obstetrics clinic of two sampled tertiary health institutions in north-central Nigeria. These include University Teaching Hospital (UTH) Abuja, and Federal Medical Center (FMC) Keffi, Nasarawa State. Binary logistic regression was used to identify and model the effects of the various risk factors while stepwise regression technique was used to fit some reduced logistic regression models. Then the best fitting model with minimum number of parameters was identified using likelihood ratio statistic. It was observed that baby's weight at birth, baby's weight four weeks since birth, and mother's occupation have significant effects on infant's bronchopneumonia status. Additionally, among the four fitted reduced models, model4 is the best predictor of infants' bronchopneumonia status, followed by model3 and then model2. Therefore, community service like home visiting for health education, supplementation of vitamin A, etc., would be an advantage if provided for teenaged pregnant women as it would, in turn, reduce incidence of low birth weight and thereby reduce bronchopneumonia infection among these children.

DOI: https://dx.doi.org/10.4314/jasem.v23i8.28

Dates: Received: 12 April 2019; Revised: 13 August 2019; 28 August 2019

Keywords: Bronchopneumonia, Multiple Logistic Regression Model, Fitness, likelihood ratio test.

Acute respiratory tract infection (ARI), predominantly pneumonia, is a major cause of morbidity and mortality among young children in developing countries. ARI is an infection of any part of respiratory tract or any related structures including para nasal sinuses, middle ear and pleural cavity (Bipin *et al,* 2011). The most common form of pneumonia in infants is bronchial pneumonia, which is also known as bronchopneumonia- an infection of the bronchial tubes of the lungs, with such symptoms as high fever, productive cough, loss of appetite, weakness, wheezing and difficulty in breathing, among others (Danan, 2002). Bronchial pneumonia is one of the leading causes of infant death. This disease kills 1.8 million children under five years of age every year, more than any other illness, in every region of the world. In spite of its huge toll, relatively few global resources are dedicated to tackling this child killer (Global Action Plan for Prevention and Control of Pneumonia, 2009). Pneumonia causes 15% of all deaths in children under age 5 worldwide, 2% of which are new-born (Janelle and Rachel, 2017). Bronchial pneumonia affects infants more than adults because their respiratory immune system is still immature. The most common cause of bronchopneumonia is a bacterial lung infection, such as *Streptococcus pneumoniae* and *Haemophilus influenza* type b (Hib). Viral and fungal lung infections can also cause pneumonia (Aaron, 2018). Thus there is every need to reduce infant morbidity and mortality from pneumonia by ensuring that every child is protected through a healthy environment and access to preventive and treatment measures. This can only be achieved by studying the major causes of infant mortality and morbidity from pneumonia (risk factors) and identifying the most important factors associated with these causes and applying the findings to child health policy with the goal of reducing child morbidity and mortality. In practice, situations involving categorical outcomes are quite common and some studies have been carried out in literature on prevalence of pneumonia and other infectious diseases in children and adults. Such studies include Danbaba *et al* (2013), Beki (2012), Vitmalkumar *et al* (2011), Cornfield (2010), and Monir *et al* (2015), among others. Most of these studies only either fit multivariate logistic regression or discriminant models to the collected data so as to determine the variables with statistically

**Corresponding Author Email:* yisa_yakubu@yahoo.com

significant effects on the response. However, this work goes beyond that as it also considered fitting some reduced binary logistic models. Usually in many research projects, after data are collected and a full model is fitted, some parameters appear insignificant. In such situations, a *reduced model* retaining only the significant terms is then adopted for use. Therefore in this study, effects of some risk factors on broncho-pneumonia status in infants were modeled using binary logistic regression methods. Then the fitted model was assessed for contribution of the individual factors and using stepwise regression technique, some reduced logistic regression models were fitted based on the variables with significant effects. These reduced models were then compared for their goodness of fit and the best fitting model with minimum number of parameters, that is, the one that best predicts bronchopneumonia status in infants, was identified using likelihood ratio statistic. Regression methods have become an integral component of any data analysis concerned with describing the relationship between a response variable and one or more explanatory variables. In most medical and epidemiologic studies, the outcome measure is categorical, such as occurrence or nonoccurrence of a disease, mortality (death or alive), etc., which may be coded as 1 or 0. Such studies call for evaluation of relative contribution of various factors to a single dichotomous or binary outcome variable and interest is always centered on modeling relationship between the probability of a success (which is between 0 and 1), and the explanatory variables (or risk factors). This relationship is nonlinear and modeling it by a linear function such as ordinary least squares (OLS) regression or linear discriminant function will violate the nonlinearity condition. This is due to strict statistical assumptions of these linear functions, such as linearity, normality, and continuity assumptions for OLS regression and multivariate normality with equal variances and covariances assumptions for discriminant analysis (Cabrera, 1994). Thus a nonlinear regression method, the most common of which is the Logistic Regression method, is the best approach in these kinds of studies (Anderson *et al*, 2003). This work therefore models the effects of the significant risk factors on infants' bronchopneumonia status using the logistic regression technique. The work also fits some reduced logistic regression models and determines the best model with minimum number of parameters.

MATERIALS AND METHODS

Data Collection: This research was carried out in two tertiary health institutions, which include University Teaching Hospital (UTH) Abuja, and Federal Medical Center (FMC) Keffi, Nasarawa State. The data set contains information on 433 births to women seen in the obstetrics clinic of these medical centers. All of these births were low birth weight. Of this total sample of 433 low birth weight cases, one hundred and eighty (180) cases were collected from the University of Abuja teaching hospital (UTH). Of this 180 cases, 80 (44.44%) were affected with pneumonia while the remaining 100 were not. The remaining two hundred and fifty three (253) cases were collected from the federal medical center (FMC), Keffi, in Nassarawa State, out of which 96 (37.9%) were affected with pneumonia. In all, 176 (40.6%) babies were affected with pneumonia.

The five variables identified in the code sheet in Table 1 were studied in this work. These variables have been recognized to be associated with low birth weight. Baby's weight was measured in grams, baby's sex was coded as 1 for male and 0 for female, mother's age was measured in years, and mother's occupation was coded as 0 for a housewife, 1 for a civil servant, and 2 for a business woman. Thus the baby's bronchopneumonia status was coded as 0 for a baby without the disease and 1 for a baby with the disease. Data on each of these variables were carefully and technically extracted directly from the individual client's medical folder.

The goal of this study was to determine whether these variables were risk factors in the clinic populations being served by each of the two medical centers.

Table1: Code sheet for Bronchopneumonia Data

variables	description	codes/values	name
1	identification code	1- 433	ID
2	Recorded birth weight at birth	Grams	BW_1
3	Recorded birth weight after 4 weeks	Grams	BW_2
4	Baby's Sex	1 = male 2 = female	S
5	Age of mother	Years	MA
6	Mother's Occupation	0 = housewife 1 = civil servant 2 = business woman	MOC
7	Bronchopneumonia status	0 = absent 1 = present	Bpn

Model Fitting: If *Y* denotes an infant's bronchopneumonia status, with values "1" if the infant is infected (a *success*), and "0" otherwise (a *failure*), then, for every sampled infant, the probability that she is infected (i.e., a success) is $\pi(x) = P(Y = 1/x)$ and the corresponding probability that she is not infected (a failure) is $1 - \pi(x) = P(Y = 0/x)$.

Let the vector $\mathbf{x}' = (\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, x_p)$ denote the set of the p predictor variables in Table1, which may be categorical or continuous. The multiple logistic regression model, which relates the probability of an infant's bronchopneumonia status to the predictor variables **x** is given **by:**

$$\hat{\pi}(x_i) = \frac{e^{\hat{\beta}_0+\hat{\beta}_1 X_{BW_1}+\hat{\beta}_2 X_{BW_2}+\hat{\beta}_3 X_S+\hat{\beta}_4 X_{MA}+\hat{\beta}_5 X_{MOC_{(1)}}+\hat{\beta}_6 X_{MOC_{(2)}}}}{1+e^{\hat{\beta}_0+\hat{\beta}_1 X_{BW_1}+\hat{\beta}_2 X_{BW_2}+\hat{\beta}_3 X_S+\hat{\beta}_4 X_{MA}+\hat{\beta}_5 X_{MOC_{(1)}}+\hat{\beta}_6 X_{MOC_{(2)}}}} \quad (1)$$

Where $\hat{\pi}(x_i)$ is the predicted probability for the *i*th infant at x_i; X_{BW_1}, X_{BW_2}, X_S, X_{MA}, $X_{MOC_{(1)}}$, and $X_{MOC_{(2)}}$ denote, respectively, baby's weight at birth, baby's weight 4 weeks after, baby's sex, mother's age, mother's occupation as a civil servant, and mother's occupation as a business woman. $\hat{\beta}_0$ denotes the estimated intercept and $\hat{\beta}_h$, $h = 1,2, \dots, p$ denotes the estimated logistic regression coefficient for the *i*th predictor variable.

Since model (1) is nonlinear, the logit transformation on $\hat{\pi}(x_i)$ yields the multiple linear logistic regression model:

$$\hat{g}_{(x)} = logit\left(\hat{\pi}(x)\right) = ln\left[\frac{\hat{\pi}(x)}{1-\hat{\pi}(x)}\right]$$
$$= \hat{\beta}_0 + \hat{\beta}_1 X_{BW_1} + \hat{\beta}_2 X_{BW_2} + \hat{\beta}_3 X_S + \hat{\beta}_4 X_{MA} + \hat{\beta}_5 X_{MOC_{(1)}} + \hat{\beta}_6 X_{MOC_{(2)}} \quad (2)$$

Where all the terms are as defined above. This model was fitted to the collected data and the parameters $\beta_0, \beta_1, \dots, \beta_6$ were estimated via maximum likelihood estimation (MLE) method with the aid of the statistical package (SPSS version 22). Equation (2) is the natural log odds of an infant infected with bronchopneumonia.

Parameter estimation by MLE method is through the *likelihood function*. The likelihood for a single observation y is $\pi(x_i)$ if $y = 1$ (i.e. if the infant is infected) and $(1 - \pi(x_i))$ if $y = 0$, where the quantity $\pi(x_i)$ denotes the value of $\pi(x)$ computed at x_i, as given in equation (1). Therefore, for the pair (x_i, y_i), the contribution to the likelihood function can be expressed as (Hosmer *et al*, 2013)

$$\pi(x_i)^{y_i}[1-\pi(x_i)]^{1-y_i} \quad (3)$$

Thus for *n* independent observations, $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, the likelihood function is

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{n} \pi(x_i)^{y_i}[1-\pi(x_i)]^{1-y_i} \quad (4)$$

The log-likelihood function is:

$$\ln l(\beta) = \sum_{i=1}^{n}\{y_i ln[\pi(x_i)] + (1-y_i) ln[1-\pi(x_i)]\} \quad (5)$$

$$= \sum_{i=1}^{n}\left\{y_i ln\left[\frac{\pi(x_i)}{1-\pi(x_i)}\right] + ln[1-\pi(x_i)]\right\}$$

Estimating the value of **β,** the vector of parameters that maximizes $ln\, l(\beta)$**,** requires differentiating (5) with respect to β_0 and β_h, $h = 1,2, \dots, p$. However, the resulting expressions are nonlinear in **β** and thus require iterative methods for solution, which have been programmed in to logistic regression software.

The fitted model was then checked for goodness so as to know if it accurately explains the data or if it incorrectly classify cases as often as it correctly classifies them. The fitted model was assessed using a test, which is based on the *deviance* statistic (*D*),

where *D* is given as -2log Likelihood statistic, with the log-likelihood function as given in equation (5). The deviance statistic is basically a measure of how much unexplained variation there is in our fitted logistic regression model – the higher the value the less accurate the model (Hosmer *et al*, 2013). This statistic compares the difference in probability between the predicted outcome and the actual outcome for each case and sums these differences together to provide a measure of the total error in the model.

Fitting Reduced Models: Usually in many research projects, after data are collected and a full model is fitted, some parameters appear insignificant. In such situations, a *reduced model* retaining only the significant terms is then adopted for use. Part of our goal in this work is also to obtain the best fitting model with the minimum number of terms or parameters. Therefore, the contribution of each variable to the fitted full model was assessed using Wald statistic, which is the ratio of the maximum likelihood estimate of each slope parameter, $\hat{\beta}_i$, to an estimate of its standard error. The significant variables were then used to fit the reduced multiple linear logistic regression models given below to the data.

1. $\hat{g}_{(x_i)} = \hat{\beta}_0$
2. $\hat{g}_{(x_i)} = \hat{\beta}_0 + \hat{\beta}_1 X_{BW_1}$
3. $\hat{g}_{(x_i)} = \hat{\beta}_0 + \hat{\beta}_1 X_{BW_1} + \hat{\beta}_2 X_{BW_2}$
4. $\hat{g}_{(x_i)} = \hat{\beta}_0 + \hat{\beta}_1 X_{BW_1} + \hat{\beta}_2 X_{BW_2} + \hat{\beta}_5 X_{MOC_{(1)}} + \hat{\beta}_6 X_{MOC_{(2)}}$

Where X_{BW_1}, X_{BW_2}, X_S, X_{MA}, $X_{MOC_{(1)}}$, and $X_{MOC_{(2)}}$ are as defined earlier. Stepwise logistic regression technique was used to conduct the regression by including and excluding variables at various stages based on their relative importance in the fitted full model. The intercept (i.e., constant- only) model was first fitted and then the variables are added one after the other. At each step the change in the deviance (-*2LL*) statistic due to the added predictor(s) was observed, by comparing the fit of the model with and without the predictor(s) using the *likelihood ratio test* statistic (*G*), given as (Hosmer *et al*, 2013):

$$G = D(model\ without\ the\ variable) - D(model\ with\ the\ variable).$$

This can be expressed as

$$G = -2ln\left[\frac{(\text{likelihood without the variable})}{(\text{likelihood with the variable})}\right] \qquad (6)$$

Under the fitted model with *k* variables, the log-likelihood function (5) can be expressed as

$$\ln L_k(\beta) = \sum_{i=1}^{n}\{y_i ln[\tilde{\pi}(x_i)] + (1-y_i)ln[1-\tilde{\pi}(x_i)]\} \qquad (7)$$

where $\tilde{\pi}(x_i) = \tilde{y}_i/n_i$ are the fitted proportions.

Under the null (reduced) model the function can be written as

$$\ln L_{null}(\beta) = \sum_{i=1}^{n}\{y_i ln[\hat{\pi}(x_i)] + (1-y_i)ln[1-\hat{\pi}(x_i)]\} \qquad (8)$$

Using (7) and (8), equation (6) becomes

$$G = -2LL_{null} - (-2LL_k) \qquad (9)$$

That is,

$$G = -2ln\left[\frac{L_{null}}{L_k}\right] = -2\left[\frac{\sum_{i=1}^{n}\{y_i ln[\hat{\pi}(x_i)]+(1-y_i)ln[1-\hat{\pi}(x_i)]\}}{\sum_{i=1}^{n}\{y_i ln[\tilde{\pi}(x_i)]+(1-y_i)ln[1-\tilde{\pi}(x_i)]\}}\right] \qquad (10)$$

$$= -2\sum_{i=1}^{n}\left[y_i ln\left(\frac{\hat{\pi}(x_i)}{\tilde{\pi}(x_i)}\right) + (1-y_i)ln\left(\frac{1-\hat{\pi}(x_i)}{1-\tilde{\pi}(x_i)}\right)\right] \qquad (11)$$

Under the hypothesis that the coefficient(s) for the *p* excluded variable(s) are equal to zero, the statistic *G* has the chi-square distribution given by

$$G = \chi^2 = -2LL_{null} - (-2LL_k) \qquad (12)$$

with *p degrees* of freedom, where *p* equals the number of predictors added to the model (i.e., $df = k_{\text{fitted}} -$

k_{null}). We expect an improvement in fit (i.e. a significant decrease in deviance) as we add more variables to the equation depending on how significant the effect of the added variables are.

RESULTS AND DISCUSSION

The code sheet in Table1 shows two categorical predictor variables for this study, which include mother's occupation with three categories and baby's gender with two categories. Table2 reveals that the 'housewife' category of the mother's occupation and the 'male' category of the baby's sex were each used as the reference category in this work. Next the baseline (or constant-only) model was fitted to the data and the model is given in Table3. The value of the deviance statistic (-2Log Likelihood) for this model is given at the bottom of the table as $-2LL = 585.023$ indicating a high unexplained variation. The fitted baseline model is

$$logit\ (\pi(x)) = ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = -0.379 \quad (13)$$

From this table we observed that this baseline model is a significant predictor of the outcome ($p < 0.001$). We then consider the accuracy of classifying the observations of the infants' bronchopneumonia status by this model, as given in Table4. From this table we observed 100.0% correct classification of the unaffected (i.e., Bpn = 0) group and 0.0% correct classification of the affected (Bpn = 1) group with 59.4% overall percentage of correct classification. This indicates that the fitted baseline model's approach to prediction is only accurate 59.4% of the time. The multiple logistic regression model given in Equation (2) was then fitted to the data as given in Table 5.

Table 2: Categorical Variables Coding

			Parameter coding	
		Frequency	(1)	(2)
Mother's occupation	housewife	199	0.000	0.000
	civil servant	137	1.000	0.000
	business mother	97	0.000	1.000
Baby's sex	male	200	0.000	
	female	233	1.000	

Table 3: Baseline model coefficient

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-.379	.098	14.973	1	.000	.685

$-2LL = 585.023$

Table 4: *Classification Table*[a,b]

			Predicted		
			baby's health status		Percentage
Observed			unaffected	affected	Correct
Step 0	baby's health status	unaffected	257	0	100.0
		affected	176	0	0.0
	Overall Percentage				59.4

a. Constant is included in the model; b. The cut value is .500

Table 5: The model coefficients

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
BWABirth	-2.915	0.365	63.671	1	0.000	0.054	0.026	0.111
BWA4Weeks	-0.939	0.268	12.269	1	0.000	0.391	0.231	0.661
Bgender(1)	-0.209	0.237	0.778	1	0.378	0.811	0.51	1.291
Mother_age	-0.002	0.021	0.008	1	0.931	0.998	0.959	1.039
Mother_occup			7.841	2	0.020			
Mother_occup(1)	0.064	0.281	0.052	1	0.819	1.066	0.615	1.848
Mother_occup(2)	0.812	0.302	7.217	1	0.007	2.253	1.246	4.076
Constant	7.313	1.047	48.769	1	0.000	1499.784		

$-2LL = 436.934$

From Table 5, we observed that the logistic regression coefficients for each of the variables BWABith and BWA4Weeks was significant but negative (-2.915 and -0.939, respectively) with p-values below 0.001. These coefficients indicate that for a one-unit increase in each of BWABirth and BWA4Weeks scores, we expect a decrease of 2.915 and 0.939 units, respectively, in the log odds of bronchopneumonia infection in infants. In terms of the odds ratio (Exp(B)), these coefficients (0.054 and 0.391) indicate

that, holding other variables at a fixed value, there is a 94.6% and 60.9% decrease, respectively, in the odds of getting infected with bronchopneumonia disease for a one-unit increase in the BWABirth and BWA4Weeks scores. That is, for every unit increase in the BWABirth and BWA4Weeks scores, these infants are, respectively, 94.6% and 60.9% less likely to be infected with bronchopneumonia disease. This result shows the significant impact of baby's weights both at birth and after four weeks in preventing bronchopneumonia infection in infants. The coefficient for mother's age (Mother_age) is also negative (-0.002) and not significant as the p-value is greater than 0.05. The corresponding odds ratio of 0.998 indicates that, holding other variables at a fixed value, the infants are only 0.2% less likely to be infected with bronchopneumonia disease. This shows that mother's age do not have a significant impact in preventing bronchopneumonia infection in infants. The coefficient for baby's gender (Bgender (1)) was negative (-0.209) and not significant, as the p-value is greater than an accepted alpha value of 0.05. Since the female group is our reference category, this coefficient is the log of the ratio of odds for the male group to the odds for the female group. The corresponding odds ratio (0.811) indicates that male infants are 18.9% less likely to get bronchopneumonia infection than females even after controlling for other variables. The coefficient for the mother's occupation as a civil servant (Mother_occup(1)) was positive (0.064) but not significant ($P > 0.05$) while that of the business woman (Mother_occup(2)) was positive (0.812) and significant ($P < 0.01$). Since housewife group is the reference category, this coefficient is the log of the ratio of odds of the business mother group to the odds of the housewife group. The corresponding odds ratio (2.253) indicates that, after controlling for other variables, the odds of mothers with business as their occupation are 125% higher than the odds for mothers who are housewives. This shows that infants from business mothers are 125% more likely to get bronchopneumonia infection than infants from mothers who are housewives.

The estimated logit is

$$logit\left(\pi(x)\right) = ln\left[\frac{\pi(x)}{1-\pi(x)}\right] =$$
$$7.313 - 2.915X_{BW_1} - 0.939X_{BW_2} - 0.209X_S - 0.002X_{MA} + 0.064X_{MOC_{(1)}} + 0.812X_{MOC_{(2)}} \quad (14)$$

And the associated estimated logistic probabilities can be obtained by

$$\pi(x) = \frac{e^{7.313-2.915X_{BW_1}-0.939X_{BW_2}-0.209X_S-0.002X_{MA}+0.064X_{MOC_{(1)}}+0.812X_{MOC_{(2)}}}}{1+e^{7.313-2.915X_{BW_1}-0.939X_{BW_2}-0.209X_S-0.002X_{MA}+0.064X_{MOC_{(1)}}+0.812X_{MOC_{(2)}}}} \quad (15)$$

Where X_{BW_1}, X_{BW_2}, X_S, X_{MA}, $X_{MOC_{(1)}}$, and $X_{MOC_{(2)}}$ are as defined earlier.

Assessing the Significance of the Fitted full Model: The fitted model was then assessed using the likelihood ratio test statistic *G* given in equation (11) by testing the null hypothesis that the p "slope" coefficients included are equal to zero. That is,

H_o: $\hat{\beta}_i = 0$
H_1: $\hat{\beta}_i \neq 0$

For the fitted baseline model with estimated coefficient in Table 3, the value of the deviance statistic *D* is $-2LL = 585.023$, while for the fitted full model with coefficients in Table 5, it is $-2LL = 436.934$, as shown at the bottom of the table. Thus the value of the likelihood ratio test *G* (equation 11) is

$$G = -2LL_{null} - (-2LL_k) = 585.023 - 436.934 = 148.089$$

Under the null hypothesis above, *G* has a chi-square distribution with *p* degrees of freedom, where *p* is the number of the added predictors. Thus the p-value for this test is $P[\chi^2(6) > 148.089] \leq 0.0001$. This is significant beyond the $\alpha = 0.05$ level and thus we reject the null hypothesis in this case and conclude that at least one or more of the *p* coefficients are different from zero. This indicates that the fitted full model is doing a better job at predicting bronchopneumonia status in infants than was the constant-only model. Next we assess the classification accuracy of this fitted full model, as presented in Table6. From this table we observed 84.4% correct classification of the unaffected (i.e., Bpn = 0) group and 60.8% correct

classification of the affected (Bpn = 1) group with 74.8% overall percentage of correct classification. This full model now classifies the outcome 74.8% of the cases compared to 59.4% in the baseline model. This is a marked improvement and also indicates that the fitted full model has a good fit and classifies quite well.

Fitting the Reduced Models: Table 7 presents the summary of the results of the four fitted reduced models. Each of the fitted models was assessed by testing the null hypothesis that the coefficients for each excluded variable is equal to zero using the likelihood ratio test ($G\ in\ Equation\ 2.8$), which compares each model to its parent one.

Table 6: Classification Table[a]

Observed			Predicted		
			baby's health status		Percentage Correct
			unaffected	affected	
Step 1	baby's health status	unaffected	217	40	84.4
		affected	69	107	60.8
	Overall Percentage				74.8

a. The cut value is .500

Table 7: Summary of the fitted Reduced Models

Variable	Model1		Model2		Model3		Model4	
	B	Odds Ratio	B	Odds Ratio	B	Odds Ratio	B	Odds Ratio
Constant	-0.379*	0.685	5.74*	311.149	7.135*	1255.431	7.161*	1288.833
BWABirth			-3.208*	0.04	-2.8*	0.061	-2.926*	0.054
BWA4Weeks					-0.938*	0.392	-0.934*	0.393
Mother_occp(1)							0.062	1.064
Mother_occp(2)							0.817*	2.263
$-2LL$	585.02		458.596		445.785		437.712	
Chi-square	0.000		126.428, df = 1, P<.001		12.810, df = 1, p<.001		8.073, df = 2, p<0.05	
Nagelkerke R^2	0.000		34.20%		37.10%		38.90%	
Classification accuracy	59.40%		74.60%		74.60%		73.90%	

The Table 7 reveals that intercept-only model was first fitted. This model has a deviance (-2 Log Likelihood) statistic of 585.02, and correctly classifies about 59% of the output. Baby's weight at birth (BWABirth) was then included in the model as a second variable to give model2. Effect of this variable was highly significant as it's addition caused the -2Log Likelihood statistic to drop significantly from 585.02 to 458.596 (i.e., by 126.428 as the chi-square (χ^2) value) with $p < 0.001$. This indicates that the expanded model (one with BWABirth) is doing a better job at predicting bronchopneumonia status in infants than was the constant-only model. This model accounts for about 34% of the variation in the bronchopneumonia status data and correctly classifies about 75% of the output. The baby's weight after 4 weeks (BWA4Weeks) was then included in the model as a third variable to give model3. The effect of this variable was also significant as it caused the -2Log Likelihood statistic to further drop significantly from 458.596 to 445.785 (i.e., by 12.810 as the chi-square (χ^2) value) with $p < 0.001$. This indicates that addition of BWA4Weeks variable has significantly improved the model and the new model (model3) is much better at predicting infants' bronchopneumonia status than the other two models. Also, this model has increased in explanatory power as it now accounts for about 37.1% of the variation in the data. However, its classification accuracy remains unchanged. Then mother's occupation (Mother_occp) was included as the fourth variable to give model4. The effect of this variable was also significant as it caused the -2Log Likelihood statistic to further drop significantly from 445.785 to 437.712 (i.e., by 8.073 as the chi-square (χ^2) value) with $p < 0.05$. This indicates that model4 is better at predicting infants' bronchopneumonia status than the other three models. Addition of this variable has increased the model's explanatory power to about 39% of the total variation in the bronchopneumonia status data, though there is a slight reduction in the classification accuracy of the model. Furthermore, this variable had three groups, *housewife*, *civil servant* (mother_occp(1)), and, *business woman* (mother_occp(2)). For this variable, housewife was considered as reference category and it was found that infants from mothers who are civil servants and those from mothers who are business women are, respectively, 6% and 126% more likely to be infected with bronchopneumonia than infants from mothers who are housewives. Baby's sex was entered into the model but its effect was insignificant and therefore it was excluded. These results revealed that of all the factors considered in this study, baby's weight after birth (BWABirth) is the best determinant of bronchopneumonia status in infants as it caused the

highest reduction of 126.428 in the -2Log Likelihood statistic. That is, bronchopneumonia infection in a newly-born baby significantly depends on the weight at birth and a child with normal birth weight is less likely to be infected with bronchopneumonia. This was followed by the weight of the baby at 4 weeks since birth (BWA4Weeks) while occupation of the mother (mother_occp) has the least effect. Comparing the four fitted reduced models from this Table in terms of the -2Log Likelihood statistic, model4 turns out to be the best predictor of infants' bronchopneumonia status, followed by model3 and then model2. This is also true when we look at each of the models' explanatory power.

Conclusion: This study has established the potentials of logistic regression technique in modeling bronchopneumonia status in infants. It has identified baby's weight at birth and his/her weight four weeks after birth as the best determinants of his/her bronchopneumonia status. Each of these was observed to have a highly-significant impact in preventing bronchopneumonia infection in infants. The study has also demonstrated the power of likelihood ratio statistic in the fitting and identification of the reduced logistic regression models that best predicts bronchopneumonia status in infants.

REFERENCES

AAP. Undated. Pneumonia. American Academy of Pediatrics. Accessed on 27th September, 2018

Aaron K. (2018) via https://www.medicalnewstoday.com/articles/323167.php Accessed on 25th January, 2019

Anderson, AA (2008). Tracking Broncho-pulmonary dysplasia (BPD) using the logistic regression model. *International Journal of Statistics* 65, 59-66.

Beki, OD (2012).The Use of Discriminant Model and Logistic Regression for Tracking the Incidence of Broncho-Pulmonary Dysplasia among Infants. An unpublished Dissertation Submitted to Usman Danfodio University, Sokoto Nigeria.

Bipin, P; Nitiben, T; Sonaliya, KN (2011). A study on prevalence of acute respiratory tract infections (ari) in under five children in urban and rural communities of ahmedabad district, Gujarat. *Nation. J. of Comm. Med.* Vol 2 Issue 2.

Cabrera, AF (1994). Logistic regression analysis in higher education: An applied perspective. Higher Education: Handbook of Theory and Research. 10: 225–56.Cornfield, A (2010). Binary logistic model for predicting Myocardia infarction. *Amer. J. Pub. Health* 81,15-21.

Danan, C (2002). Gelatinase activities in the airways of premature infants and development in bronchopulmonary dysplasia. *Amer. J. Physiol., Lung cell MOI Physiol.* 283: L1086-L1093.

Danbaba, A; Audu, M; Mahmud, AM (2013). Investigation of risk factors of low birth weight using multivariate logistic regression analysis. *J. of Hum, Sci. Technol.* A publication of Niger State Polytechnic. 3, 59-77

Global Action Plan for Prevention and Control of Pneumonia (2009). World Health Organization/The United Nations Children's Fund (UNICEF).

Hosmer, DW; Lemeshow, S (2000). Applied Logistic Regression. (2nd ed.) Wiley and Sons, inc. New York.

Janelle, M; Rachel, N (2017). Bronchopneumonia: Symptoms, Risk Factors, and Treatment.

Vimalkumar, VK; Moses, RC; Padmanaban, S (2011). Binary logistic model for detecting prevalence and risk factor of Nephropathy in type 2 diabetic patients. *Int. j. Collabor. Res. Internal Med. Pub. Health.* 3 (8): 598-615.