Science World Journal Vol. 14(No 4) 2019
www.scienceworldjournal.org
ISSN 1597-6343
Published by Faculty of Science, Kaduna State University

Full Length Research Article

A LOCALIZED LOGIT MODEL FOR PREDICTING INFANT SURVIVAL OUTCOMES

*, [1] Agada, P.O., [2] Udoumoh, E. F., [3] Odoma, P.A.

Department of Mathematics/Statistics/Computer Science, University of Agriculture Makurdi, Benue State, Nigeria

Authors Email Addresses: [1]*gadexx@yahoo.com , [2]uenobong@gmail.com , [3]odomadommzzy@gmail.com *Phone: +2348038637735

ABSTRACT

This study develops a Binary Logistic Model, local to the Maria Goretti hospital Anyigba, Kogi State, Nigeria for predicting infant birth survival outcomes. Model results shows that the odds of infant survival for normal delivery are 5 times the odds of survival by caesarean delivery, keeping the weight of the infant constant and also that the odds of infant survival increases 3 times for every unit increase in the weight of an infant when the delivery method is kept constant. It was concluded that though, mother's age which incorporate a maternal characteristic into the model was insignificant in this hospital, there is a chance that it might be significant alongside other maternal characteristics elsewhere. Hence the emphasis on developing localized models of this nature.

Keywords: Model, Logistic, Infant, Survival

INTRODUCTION

According to Fagbamigbe and Idemudia (2017), the main objective of an Antenatal Clinic (ANC) is to ensure optimal health outcomes for the mother and her baby and that appropriate care during pregnancy is vital for the health of the mother and the development of the unborn baby. The authors further established via a Binary Logistic Regression approach, the relationship between ANC utilization and the following covariates and their categories; wealth status, categorized into poorest, poorer, average, wealthier, and wealthiest; educational attainment, categorized into no formal education, Quranic, primary, secondary, and higher; marital status, categorized into never married, formerly married, and currently married. Other covariates include; location of residence, categorized into rural and urban settings, geopolitical zones, categorized into North Central, North East, North West, South East, South, and South West; age of respondent at birth, categorized into <20, 20–24, 25–34 and 35–49 years, Additional covariates include; tribe, categorized into Hausa, Yoruba, Igbo, and others; religion, categorized into Islam, Christianity, and others, birth order, categorized into 1, 2, 3, or 4 and finally; employment status, categorized into currently employed and unemployed; and some behavioural factors, categorized into need for spouse permission to visit ANC provider: yes/no; having problem with money: yes/no; and whether distance from health facility is a problem: yes/no.

According to the authors, these variables have been associated with ANC utilization in previous studies (Arthur, 2012; Dairo and Owoyokun, 2010; Fagbamigbe and Idemudia, 2015; Gage, 2007; Glei, Goldman and Rodriguez, 2003; Gymiah, Takyi, and Addai, 2006; Joshi, Torvaldsen, Hodgson, and Hayen, 2014; Kyei, Campbell and Gabrysch, 2012).

A major result of the work of Fagbamigbe and Idemudia (2017) is that, ANC use is lowest among women who need it most. These are the poor, uneducated women living in rural areas. They added that, despite the prevalent free maternal health care policy in Nigeria, poverty and literacy level influenced ANC utilization and the adequacy of number of ANC visits more than any other factor. Our take in all of these is that, since appropriate care during ANC visits is vital for the health of the mother and the development of the unborn baby and the fact that these visits are to a large extent affected by the aforementioned factors (such as the location of the mother (rural or urban) as well as hospital characteristics, it follows that spatial variability in ANC utilization should be expected. Consequently, the survival rate of infants at birth is also expected to vary spatially.

To this end, we emphasize that; global prediction model of infant survival outcomes should not be encouraged. Rather, emphasis should be on developing local or spatially dependent prediction models of infant survival rate or outcomes. The study proceeds to develop a Binary Logistic Model local to the Maria Goretti hospital Anyingba, Kogi State, Nigeria for predicting infant survival outcomes. The rest of the paper is organized as follows; Materials and Methods, Results and Discussion, Conclusion and recommendations.

MATERIALS AND METHODS

Source of Data

The data used for this research is a secondary data collected from the records department of Maria Goretti Hospital Anyigba, Kogi State Nigeria. The data was collected across the months of the year over a period of three years. A total of 713 births were documented during the period.

Model Description

Binary logistic regression is a statistical modelling procedure for building regression models used for analysing a data set in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two outcomes). The main objective of binary regression modelling is to find a model that best describe the existing relationship between the dichotomous characteristic of interest and the set of independent or predictor variables. The model achieves this by generating the coefficients of a formula to predict a logit transformation of the probability of the presence of the characteristic of interest:

$$Logit(p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K \quad (1)$$

It follows from equation (1) that the Infant Survival Rate (%) can be modelled as

$$p = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_K X_K)}} * 100\% \quad (2)$$

Where p is the probability of the presence of the characteristic of interest, the β_i's and the $X_i's$, $i = 1,2,\ldots,k$ are the regression coefficients and the independent variables respectively. In this work, p is the probability of infant survival or infant survival rate, the independent variables include, child's weight, sex (male or female), mother's age, and delivery method (normal or caesarean operation).
The logit transformation is defined as the logged odds where;

$$Odds = \frac{p}{1-p} = \frac{probability\ of\ the\ presence\ of\ the\ characteristic}{probability\ of\ the\ absence\ of\ the\ characteristic} \quad (3)$$

and

$$Logit(p) = Ln\left(\frac{P}{1-P}\right) \quad (4)$$

Unlike the ordinary regression model that chooses model parameters that minimize the sum of squares of errors, logistic models choose parameters that maximize likelihood of observing the sample values. The Forward Stepwise (Likelihood Ratio) regression method is employed in the modelling process.

Model sample size adequacy, goodness of fit and adequacy checks

Model sample size adequacy
The complexity in determining optimal sample size for logistic regression modelling made Peduzzi *et al.* (1996) suggested that , if p is the smallest of the proportions of positive or negative cases in a population and k the number of covariates, then the minimum number of cases to include is :

$$N = \frac{10*K}{P} \quad (5)$$

If the resulting $N < 100$, Long (1997) suggested that it should be rounded up to 100.

If we consider p = 0.1, with four (4) independent variables, a sample size of 400 would be required. If we substitute 713 for N and 4 for K, then the least proportion of infants who survived would be 0.056 which is approximately 0.1. In this regard, the sample size of 713 can be considered adequate.

Model goodness of fit and adequacy checks
In order to test for the goodness of fit of the model, the -2 Log likelihood Null and Full models are employed. The Null model -2 Log likelihood is given as;

$$-2 * Ln(L_0) \quad (6)$$

Where L_0 is the likelihood of obtaining the observations if the independent variables are not included in the model (i.e have no effect on the outcome) while the Full model -2 Log likelihood is given as;

$$-2 * Ln(L) \quad (7)$$

Where L is the likelihood of obtaining the observations if the independent variables are included in the model (i.e they do have effect on the outcome).

The difference in these two yields a chi-square statistic which is a measure of how well the independent variables affect the outcome or dependent variable. If the p value for the overall model fit statistic is less than 0.05, then there is evidence that at least one of the independent variables contributes to the prediction of the outcome.

The Hosmer – Lemeshow test is a test of goodness of fit employed in this work. The test divides the test data into approximately 10 groups. The chi-square statistic for this test is computed by;

$$\chi^2_{HL} = \sum_{g=1}^{G} \frac{(O_g - E_g)}{E_g(1-E_g/n_g)} \quad (8)$$

with $O_g, E_g\ and\ n_g$ defined as the observed events, expected events and number of observations for the g^th decile group and G the number of groups. The number of degree of freedom is G-2. A large value of chi-square with small p-value < 0.05 indicates poor fit while a small chi-square vale with p –value closer to 1 indicate a good logistic regression model fit.
In order to evaluate the prediction accuracy of the Binary logistic model, the classification table is employed. On this table, the observed values of the dependent variable and the predicted values at a user defined cut - off value are cross classified.
The Walds statistic tests the significance of model parameters. This helps to determine whether or not an independent variable stays in the model as it tests if the associated model parameter differs significantly from zero. The Walds statistic is computed as the regression coefficient divided by its standard error squared:

$$\left(\frac{\beta}{SE}\right)^2$$

Where
$\beta = the\ associated\ regression\ parameter$,
$SE = It'standared\ error$.
If the p-value is less than the usual $\alpha = 0.05$, then we have evidence to conclude that the independent variable differ significantly from zero. Hence it stays in the model.

Odds ratio
Re-writing equation (1) by taking the exponential of both sides of the equation, we have;

$$Odds = \frac{p}{1-p} = e^{\beta_0} e^{\beta_1 X_1} e^{\beta_2 x_2} \ldots e^{\beta_k X_k} \quad (9)$$

It is obvious from equation 5 above that when an independent variable X_i changes by 1 unit (all other variables kept constant), the odds changes by the factor e^{β_i}. This factor is termed the odds ratio (O.R) for the independent variable X_i. It gives the relative amount by which the odds of the outcome of interest increases (O.R > 1) or decreases (O.R < 1) when the value of the independent variable is changes by 1 unit.

RESULTS AND DISCUSSION
Details of model results are presented in this section. This include model parameter values and statistics, the fitted survival rate model, the -2 Log likelihood Null and Full model statistics, contingency table for Hosmer and Lemeshow test, and the child birth state classification table.

Science World Journal Vol. 14(No 4) 2019
www.scienceworldjournal.org
ISSN 1597-6343
Published by Faculty of Science, Kaduna State University

Table 1: Model parameter values and statistics obtained after the implementation of the Forward Stepwise (Likelihood Ratio) Regression Method

		B	S.E.	Wald	df	P value	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	Age	.042	.034	1.510	1	.219	1.043	.975	1.116
	Method	1.640	.370	19.598	1	.000	5.153	2.494	10.650
	Weight	1.129	.293	14.889	1	.000	3.093	1.743	5.488
	Constant	-2.335	1.173	3.961	1	.047	.097		
Step 2[a]	Method	1.648	.370	19.881	1	.000	5.195	2.518	10.719
	Weight	1.150	.293	15.439	1	.000	3.157	1.779	5.602
	Constant	-1.259	.793	2.522	1	.112	.284		

a. Variable(s) entered on step 1: Age, Method, Weight, Sex. $Survival\ rate = \frac{1}{1+e^{-(-1.259+1.648\ Method+1.150\ Weight)}} * 100\%$,S.E = Standard error of model parameters, df = degree of freedom, Exp(β) = Odd ratio, C.I = confidence interval

The Statistical Software for Social Science (SPSS) was used to fit the Infant Survival Rate Model to the data set of maternal and infant birth characteristics captured by the dichotomous response variable; child birth state (survive, stillbirth) and the covariates; mothers age, method of delivery, weight and sex of infant. The Forward stepwise (Likelihood Ratio) method of regression was employed in fitting the model. All the covariates were entered at the start (step 0). Age and sex were found to be insignificant in the model (p value > 0.05). In step 1, sex was eliminated from the model and the result shows that the covariates mother's age is still in significant (p value > 0.05).

In step 2, mother's age was removed and method of delivery and weight of infant became the only significant covariates influencing child birth state (p < 0.05) see table 1 for details. The model regression parameters (β) their standard error (S.E), values of Wald's statistic and Exp(β) are shown on table 1. For the covariate; method of delivery, its corresponding value of Exp(β); 5.195 shows that the odds of infant survival for normal delivery are 5 times the odds of survival by caesarean delivery keeping infant weight constant. While for the covariate; weight of infant, it's corresponding value of Exp(β); 3.157 shows that for every unit increase in infant's weight, the odds in infant survival increases 3 times, when the delivery method is kept constant. Using equation (2) the Infant Survival Rate Model is obtained. This is relayed in the foot note of table 1. The model estimates in percentage, the survival rate of an infant given it's weight and delivery method

Table 2: Details of variables not in the model

			Score	df	p value
Step 1[a]	Variables	Sex	.314	1	.575
	Overall Statistics		.314	1	.575
Step 2[b]	Variables	Age	1.520	1	.218
		Sex	.130	1	.718
	Overall Statistics		1.832	2	.400

a. Variable(s) removed on step 1: Sex. b. Variable(s) removed on step 2: Age. df = degree of freedom

Table 2 provide the statistical details of variables removed from the model.

Table 3: Change in -2 Log likelihood for each covariate at each step

Variable		Model Log Likelihood	Change in -2 Log Likelihood	df	p value. of the Change
Step 0	Age	-116.837	1.717	1	.190
	Method	-126.225	20.492	1	.000
	Weight	-122.990	14.023	1	.000
	Sex	-116.136	.315	1	.575
Step 1	Age	-116.903	1.533	1	.216
	Method	-126.250	20.227	1	.000
	Weight	-123.839	15.405	1	.000
Step 2	Method	-127.158	20.510	1	.000
	Weight	-124.983	16.161	1	.000

Step 1, -2 Log likelihood = 232.273, Step 2, -2 Log likelihood = 233.805

The -2 Log likelihood Null and Full model summary on table 3 shows the change in the -2 Log likelihood and the significance at each step of the modelling process. In step 2 (the last step), the result shows that the change is significant for both covariates; method of delivery and weight of infant (p value < 0.05). This indicates the goodness of model fit.

Table 4: Pseudo R^2 values

Step	Cox & Snell R Square	Nagelkerke R Square
1	.060	.174
2	.058	.167

What constitutes a "good" R^2 value varies between different areas of application. While these statistics can be suggestive on their own, they are most useful when comparing competing models for the same data. The model with the largest R^2 statistic is "best" according to this measure. But observe that these values are approximately equal for both models (step 1 and step 2 models). Hence, the Cox & Snell R^2 and Nagelkerke R^2 values are not considered very relevant in establishing the goodness fit of our model. Rather, the contingency table for Hosmer and Lemeshow test, and the child birth state classification table are employed. Hosmer and Lemeshow test statistic follows a chi-square distribution.

Table 5: Contingency table for Hosmer and Lemeshow test

		Child's birth state = Stillbirth		Child's birth state = Survived		
		Observed	Expected	Observed	Expected	Total
Step 1	1	15	14.449	50	50.551	65
	2	6	6.640	59	58.360	65
	3	2	3.840	64	62.160	66
	4	1	2.701	65	63.299	66
	5	1	2.009	64	62.991	65
	6	3	1.601	61	62.399	64
	7	4	1.349	63	65.651	67
	8	2	1.110	64	64.890	66
	9	1	.846	64	64.154	65
	10	0	.456	56	55.544	56
Step 2	1	14	14.425	53	52.575	67
	2	6	6.902	64	63.098	70
	3	4	4.143	71	70.857	75
	4	1	2.076	53	51.924	54
	5	1	2.037	65	63.963	66
	6	2	1.681	65	65.319	67
	7	4	1.609	76	78.391	80
	8	2	.996	60	61.004	62
	9	1	1.132	103	102.868	104

Step 1: χ^2= 10.460, df = 8, p value = 0.234, Step 2 : χ^2 = 6.010, df = 7 , p value = 0.539

The footnote of table 5 shows that the model in step 2 is a better fit of the Bayesian Logistic Regression Model with a greater p value of 0.539 > 0.05. This contingency shows close values of observed and expected frequencies for the survived and stillbirth child birth states. This further buttress the goodness of the model.

Table 6: Child's Birth State Classification Table

Observed			Predicted		
			Child's birth state		
			Stillbirth	Survived	Percentage Correct
Step 1	Child's birth state	Stillbirth	2	33	5.7
		Survived	0	610	100.0
	Overall Percentage				94.9
Step 2	Child's birth state	Stillbirth	1	34	2.9
		Survived	0	610	100.0
	Overall Percentage				94.7

a. The cut value is .500

The child birth state classification table (table 6) shows a 94.7% correct classification of the observed cases of child birth state by the model (see step 2 classification on table 6). This result is preferred to that of step 1 since the model in step 2 does not include age and sex which were earlier established as insignificant. This in addition establishes how good our model is.

Implications of Results

We earlier established that, appropriate care during ANC visits is vital for the health of the mother and the development of the unborn baby and also that these visits are to a large extent affected by factors, some among which are the location of the respondent (rural or urban). Hence spatial variability in ANC utilization should be expected and as such, the survival rate of infants at birth should also vary spatially. Consequently, emphasis should be on developing local or spatially dependent prediction models of infant survival rate while the development of its global counterparts should not be encouraged. To demonstrate the development of one of such localized models, the study developed a Binary Logistic Model, local to the Maria Goretti hospital Anyigba, Kogi State, Nigeria for predicting infant survival outcomes.

This model is hoped to compliment the efforts of the Medical Practitioners in this hospital in predicting the chance of infant survival given the weight and delivery method of infants. In fact since the odds of infant survival increases 3 times for each unit increase in the weight of the infant, then for normal delivery, doctors can determine from the model, the weights ($\geq$ the WHO approved weight of 2.5 kg) that yield high infant survival outcomes. They can therefore look for a way of regulating and monitoring the weight of the infant about a determined value to ensure that the mother gives birth normally. Though, mothers age which incorporate a maternal characteristic into the model was insignificant in this location, there is a chance that it might be significant alongside other maternal characteristics elsewhere. Hence our emphasis on developing localized models of this nature.

Conclusion

The following conclusions were drawn from the study;

(i) A Localized Binary Logistic Regression Model for predicting infant survival outcomes has been developed in this study.
(ii) The odds of infant survival for normal delivery are 5 times the odds of survival by caesarean delivery keeping the weight of the infant constant.
(iii) The odds of infant survival increases 3 times for every unit increase in the weight of an infant when the delivery method is kept constant.
(iv) Though, mothers age which incorporate a maternal characteristic into the model was insignificant in this location, there is a chance that it might be significant alongside other maternal characteristics elsewhere. Hence our emphasis on developing localized models of this nature.

Recommendations

The study recommends that;

(i) The development of local or spatially dependent prediction models of infant survival outcomes should be emphasized over their global counterparts.
(ii) The model should be used to compliment doctors' effort in determining WHO approved infant weights that will yield high survival chances of normal delivery.

REFERENCES

Arthur, E., 2012. Wealth and antenatal care use: Implications for maternal health care utilisation in Ghana. *Health Economics Review, 2*(14), 1–18.

Dairo, M., Owoyokun, K. E., 2010. Factors affecting the utilization of antenatal care services in Ibadan, Nigeria. *Benin Journal of Postgraduate Medicine, 12(1), 1–6.*

Fagbamigbe, A. F., Idemudia, E. S., 2017. Wealth and antenatal care utilization in Nigeria. Heath Care Women Int.38(1):17-37

Fagbamigbe, A. F., Idemudia, E. S., 2015. Barriers to Antenatal Care Use in Nigeria : evidences from Non – users and implications for maternal health programming. BMC pregnancy and child birth. 15(1):95-104

Gage, A.J., 2007. Barriers to the Utilization of Maternal Health Care in rural Mali. *Social Science and Medicine*, 65(8), 16661682.

Glei, D., Goldman, N., Rodriguez, G., 2003. Utilization of care during pregnancy in rural Guatemala: Does obstetrical need matter? *Social Science & Medicine, 57*, 2447– 2463.

Gymiah, S. O., Takyi, B. K., Addai, I., 2006. Challenges to the reproductive health needs of African women on religion and maternal health utilization in Ghana. *Social Science and Medicine, 62,* 2930–2944.

Joshi, C., Torvaldsen, S., Hodgson, R., Hayen, A., 2014. Factors associated with the use and quality of antenatal care in Nepal: A population-based study using the demographic and health survey data. *BMC Pregnancy and Childbirth, 14*(1), 94.

Kyei, N., Campbell, O.M.R., Gabrysch, S., 2012. The influence of distance and level of service provision on antenatal care use in rural Zambia. *PloS One, 7*(10), 1–6.

Long, J. S. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage Publications.

Peduzzi P., Concato J., Kemper E., Holford TR., Feinstein AR., 1996. A Simulation study of the number of events per variable in logistic regression analysis. Journal of Clinical Epidemology 49:1373-1379.