Annals of Health Research
Volume 2, Issue No. 1
January - June 2016

ORIGINAL RESEARCH

Comparison of statistical models used for assessing factors associated with infant mortality in Nigeria

Oritogun KS*[1], Oyewole OO[2], Daniel OJ[1]

[1]Department of Community Medicine and Primary Care, Faculty of Clinical Sciences, Obafemi Awolowo College of Health Sciences, Olabisi Onabanjo University, Sagamu, Ogun State.

[2]Department of Physiotherapy, Olabisi Onabanjo University Teaching Hospital, Sagamu, Ogun State.

Abstract

Background: Infant mortality is a public health concern especially in developing countries, particularly Nigeria. Different models had been used independently to identify factors associated with infant mortality. Some of the used models sometimes violate the underlying assumption for the models. This study was designed to compare the models that have been previously used and identify the appropriate model using standard model selection criteria to analyse risk factors for infant mortality in Nigeria.

Methods: The study utilised 2008 Nigeria Demographic and Health Survey (NDHS) data with a sample size of 7107. The NDHS was a stratified two-stage cluster design where a questionnaire was used to collect data on the birth history of women aged 15-49 years. The models employed for this study were: Logit, Probit and Clog-log. The model selection criteria were Akaike Information Criterion (AIC), Residual Deviance and Vuong test. The model with the smallest criteria was considered to be the best fit.

Results: The results showed that Infant Mortality in Nigeria can be appropriately modelled by Clog-log model. The models and corresponding AIC values were: Logit (6171.1), Probit (6212.6) and Clog-log (6126.6). The residual deviance included: Logit (6135.1), Probit (6176.6) and Clog-log (6090.6). Clog-log had the smallest AIC and residual deviance values; hence, it was of the best fit. Home delivery and delivery by professionals had negative significant associations with infant mortality while women's education (primary/no education) and birth order had positive significant association, ($p < 0.05$).

Conclusion: The best model for infant mortality evaluation in Nigeria was Clog-log. Generally, improved women's education would significantly reduce Infant Mortality in Nigeria.

Keywords: *Akaike's Information Criterion, Infant mortality, Nigeria Demographic and Health Survey, Residual deviance, Vuong test, Statistical model*

Introduction

The problems of infant and childhood mortality are public health issues which cannot be over-emphasized,

especially in developing countries. Infant mortality is the probability of dying before the first birthday while under-five mortality is the probability of dying between birth and the fifth birthday. [1, 2] These mortality rates are basic indicators of quality of life and socioeconomic development in a country. [1, 3]

*
Correspondence
Department of Community Medicine & Primary Care,
Obafemi Awolowo College of Health Sciences,
Olabisi Onabanjo University,
PMB 2022, Sagamu, Nigeria.
Tel: +23480-2894-2869
E-mail: oritogunks@yahoo.co.uk

Though, some progress was made with the drive to meet the Millennium Development Goals (MDG), some six million children still die globally before their fifth birthday annually. It is hoped that, with the renewed efforts of the Sustainable Development Goals (SDG), under-5 mortality will be reduced to as low as 25 per 1000 live births by the year 2030. [4, 5]

The Nigerian Demographic and Health Survey in 2008 reported the infant and under-five mortality rates in Nigeria to be 75 per 1,000 live births and 157 per 1,000 live births respectively. [1] Researchers have worked extensively on infant and under-five mortalities with the aim of halting and reversing the increasing rates. Studies carried out on infant mortality, are rather fewer than studies of under-five mortality.

In the previous studies, different statistical methods such as univariate analysis and multiple Cox proportional hazards regression models, Chi-squared test, logistic regression models, multiple linear regression and a few others, have been employed in identifying factors associated with infant and child mortality in Nigeria. [6,9,13] However, the studies have shown that logistic regression analysis, which is Logit model and Chi-squared test are most frequently used models for the analyses of Nigeria Demographic and Health Survey 2008 data when describing infant and under-five mortalities. [6, 9, 12,13] The reliability of the findings from some of these statistical methods may be questionable since the theories/assumptions behind their uses are sometimes neglected.

Some researchers have determined some of the factors associated with infant and childhood mortalities in Nigeria using the 2008 Nigeria National Demographic and Health Survey data and different statistical methods. [6 - 9] The rural areas and the northern regions in Nigeria were significantly and positively associated with neonatal deaths as more deaths occurred in these areas. Women with no formal education are also at higher risk of neonatal mortality compared to women with at least a secondary education. In addition, receiving antenatal care and receiving skilled assistance at delivery were reported to have significantly reduced neonatal deaths. [6, 9, 13] However, the conclusions from these studies may not be accurate if the method of data analysis did not observe the standard logit probability distribution assumption for the use of a logit model. This situation can be resolved by applying the criteria for model selection and goodness of fit to determine which model would be the best for the data set. The goal is to have better parameter estimates and provide evidence-based and appropriate conclusions for policy makers and other stakeholders.

Therefore, this study was aimed at comparing some Generalized Linear Models (GLM) with different link functions (Logit, Probit and Complementary log-log models) on Infant mortality and to select the best model that explains the associated factors with infant mortality.

Methods

Data and Design

The study was a retrospective analysis of the data obtained from the Nigeria Demographic and Health Survey (NDHS) for the year 2008. A total of 7107 observations were included in the analysis. The 2008 NDHS data were used to compare the output from this study with the previous studies that used the same NDHS 2008 for better comparison of the models. The subsequent studies will be compared with the results from this study to show consistency of the model in the 2013 NDHS. The 2008 NDHS used the 2006 population and Housing census of the Federal Republic of Nigeria by the National Population Commission (NPC) as the sampling frame. It was a stratified two-stage cluster design consisting 888 clusters, having 286 in the urban and 602 in the rural areas with a representative sample of 36,800 households. The questionnaires used for the 2008 NDHS were Household Questionnaire, Women's Questionnaire and Men's Questionnaire with questions pertaining to birth history, childhood mortality and other characteristics. [1] The details of the data collection methods and survey design are available in the 2008 NDHS report.

Statistical analysis

Statistical analysis was carried out with R

programming packages. [14 - 16] Microsoft Excel® was used as data spreadsheet prior to exporting to R. Recoding of variables such as breastfeeding, region, place of delivery, education and religion were done as shown in Table I.

Generalized Linear Models (GLM) such as Logit (LO), Probit (PR) and Complementary log-log (CLL) were used to analyse the data separately and the findings for each method were compared. Step AIC method in R was used to select 10 variables out of 12 variables for the initial analysis. Out of the 12 variables (Table I), assistance by doctors during delivery and availability/ non-availability of electricity were skipped. These variables were subsequently used in the final analysis. Akaike Information Criterion (AIC), Residual Deviance and Vuong test were the criteria for model selection. The model with the smallest AIC and Residual Deviance was considered to be of best fit out of the three models under consideration. P-values < 0.05 were considered to be statistically significant.

Response and Explanatory Variables

The outcome variable or dependent variable for this study was whether the child died or not between ages 0-11 months (child died = 1, child not dead=0). The explanatory variables are as shown in Table I.

Models: [17 - 19]

The Logistic model is given as:

(A) $Logit \quad \log \frac{}{1} \quad _0 \quad _1x_1 \quad _2x_2 \quad \dots \quad _jx_j \quad Z \qquad (1)$

Where: is the probability of occurrence of infant death.

$_1, \ _2, \ _3, \dots \ _j$ are the regression coefficients for covariates $x_1, x_2, x_3, \dots, x_j$.

(B) Probit model:

$(Z) \qquad (\ _0 \quad _1x_1 \quad _2x_2 \quad \dots \quad _jx_j) \qquad (2)$

$Z \qquad ^1(\ ) \quad _0 \quad _1x_1 \quad _2x_2 \quad \dots \quad _jx_j \qquad (3)$

Given that ^{1} is the inverse of the Normal Cumulative Distribution function (CDF).

(C) Complementary log-log: Clog-log

$\log\{ \ \log(1 \quad)\} \quad _0 \quad _1x_1 \quad _2x_2 \quad \dots \quad _jx_j \quad Z \qquad (4)$

$\log(1 \quad) \quad e^{\ _0 \quad _1x_1 \quad _2x_2 \quad \dots \quad _jx_j} \qquad (5)$

Results

The variables used for this study were as described in Table I which shows both the outcome and explanatory variables.

Table I: Description of Variables

Variables	*Description*
Dependent/Response	Child death 0 = No; 1 = Yes
Explanatory variables	
Breast_feeding	Breastfeeding (0 = never; 1 = breastfed
Region_grp	Region (1 = north; 2 = south)
delivery_place	Place of delivery (1 = home, 2 = Govt/Public Health Centre; 3 = Private hosp/clinic; 4 = others
Education	Education (0 = primary/ no education' 1 = secondary; 3 = Higher)
hh_members	Household members
asst_nurse	Assistance by a Nurse (0 = No; 1 = Yes)
asstance_relative	Assistance by relatives (0 = No; 1 = Yes)
asst_none	Assistance by none (0 = No; 1 = Yes)
birth_order	Birth order
Religion	Religion (1= Catholic; 2 = Other Christian; 3 = Islam; 4 = Traditionalist/Others)
asst_doctor	Assistance by doctor (0 = No; 1 = Yes)
electric_present	Electricity (0 = No; 1 = Yes)

Model comparison

The Akaike Information Criterion (AIC) for Logit (6171.1), Probit (6212.6) and Complementary log-log (Clog-log (6126.6) models are shown in Table II. Clog-log model had the smallest AIC (6126.6). The residual deviance for Logit (6135.1), Probit (6176.6) and Clog-log (6090.6) are also shown in Table II, where Clog-log had the smallest value. This observation suggested Clog-log model was of best fit compared to other models which had bigger values of both AIC and Residual deviance. Vuong statistics which pairs the models to show which one of the two models paired is of good fit is shown in the same Table II. When Logit and Probit models were compared, Logit model was observed to be of good fit (Z= 7.243, P < 0.001). The comparison of Logit and Clog-log models also showed the Clog-log model to be of good fit (Z= -8.239, P < 0.001). The comparison of Clog-log model and Probit model also showed Clog-log model to be of good fit (Z= -8.243, P < 0.001).

The estimates of the three models are given in

Table III. Factors which had negative but significant association with infant mortality, as identified by Logit, Probit and Clog-log models (P < 0.0001), included breastfeeding, region (south), place of delivery (home), assistance by nurses during delivery and Islamic religion. These variables had statistically significant reduction in infant mortality. Birth order, lack of breastfeeding and assistance by relatives during delivery had statistically significant positive association with infant mortality (p < 0.0001).

However, the Clog-log model identified no assistance during delivery as an important factor that was significantly positively associated with infant mortality (P < 0.01) which was not identified by the other two models. In addition, assistance by relatives during delivery had a higher level of significance at p < 0.0001 compared with the logit and probit models at significance levels of p < 0.001 and < 0.01 respectively. The place of delivery (Government/public Health services; Private clinic), secondary education, religion (other Christians, Traditional) in Clog-log model were not significantly associated with infant mortality (p > 0.05).

Table II: Model comparison using Akaike's Information Criterion (AIC), Residual Deviance and Vuong non-Nested hypothesis Test-statistics

Statistics	*Models*		
	Logit	**Probit**	**Clog-log**
AIC	6171.1	6212.6	6126.6
Residual Deviance	6135.1	6176.6	6090.6
Vuong non-Nested hypothesis Test - Statistic			
Model	Vuong Z- Statistics	Decision	p-values
Logit Vs Probit	7.243	Logit > Probit	P < 0.001
Logit Vs Clog-log	-8.239	Clog-log > Logit	P < 0.001
Probit Vs Clog-log	-8.243	Clog-log > Probit	P < 0.001

Table III: Estimates of Logit (LO), Probit (PR) and Clog-log (CLL) models

Variables	**Estimates (β) (Std error)**		
	LO	PR	CLL
Intercept	-0.349	-0.307**	-0.547**
Breastfeeding: breastfed	-0.406***	-0.234***	-0.338***
Breastfeeding: never breastfed	0.695***	0.422***	0.523***
Region_grp: south	-1.119***	-0.624***	-1.008***
Delivery place:			
Govt/Public health	0.040	0.030	0.014
Home (any)	-1.193***	-0.659***	-1.080***
Others	-1.217	-0.696	-1.116
Private hosp/clinic	-0.115	-0.056	-0.125
Education: primary/none	0.532**	0.316**	0.448**
Education: secondary	0.081	0.062	0.056
hh_members	-0.186***	-0.089***	-0.177***
asst_nurse: Yes	-0.557***	-0.312***	0.468***
assistance_relative: Yes	0.255**	0.124*	0.268***
ass_noneYes: no assistance	0.161	0.084	0.179*
birth_order	0.199***	0.104***	0.177***
religionislam	-0.351**	-0.213**	-0.276**
Religionotherchristian	-0.001	0.005	0.002
Religiontraditionalist/others	0.359	0.220	0.328

*Significant at 1%, **Significant at 0.1%, ***Significant at 0.001%

The odd ratios for the three models were also shown in Table IV. However, since the study suggested Clog-log model to be of best fit (AIC = 6126.6; Z = -8.239; P < 0.01), then, the study dwelt more on the interpretation of the odds ratios of the model. Table IV shows that breastfed children were less likely to record infant mortality than those who were not breastfed (OR = 0.713). The risk of infant mortality was less for children in the southern part of Nigeria compared to children in northern Nigeria (OR = 0.365). Children delivered at home (herbalist and others) were 0.34 less likely to die in infancy than children delivered in other places. Children of mothers who had no education or primary education were 1.57 more likely to die in infancy compared to those whose mothers had higher educational qualifications. An increase in one household member has 16% decreases in the odds of death in infancy. In addition, children whose births were assisted by nurses were 0.63 less likely to die in infancy compared to those whose births were not assisted by nurses. Assistance by relatives and outright lack of assistance during delivery were associated with 1.31 and 1.20 higher risks of infant mortality. The results also showed that one increase in birth order was associated with a 19.4% increase in the odds of infant mortality. Being a Muslim was associated with 0.76 less likelihood of infant mortality compared to Catholic religion.

Discussion

Generalized Linear Models (GLMs) with different link functions such as Logit, Probit and Clog-log were evaluated in this study. The smallest values

Table IV: Odds Ratio for Logit (LO), Probit (PR) and Clog-log (CLL) models

Variables	*EXP (B)*		
	LO	**PR**	**CLL**
Intercept	0.706	0.736**	0.579**
Breastfeeding: breastfed	0.666***	0.791***	0.713**
Breastfeeding: never breastfed	2.005***	1.525***	1.686***
Region_grp: south	0.327***	0.536***	0.365***
Delivery place:			
Govt/Public health	1.041	1.030	1.014
Home (any)	0.303***	0.517***	0.339
Others	0.296	0.499	0.328
Private hosp/clinic	0.891	0.945	0.883
Education: primary/none	1.703**	1.372***	1.566**
Education: secondary	1.084	1.064	1.058
hh_members	0.830***	0.915***	0.838***
asst_nurse: Yes	0.573***	0.732***	0.626***
assistance_relative: Yes	1.290**	1.132*	1.307***
ass_noneYes: no assistance	1.175	1.088	1.196*
birth_order	1.220***	1.110***	1.194***
religionislam	0.704**	0.808**	0.759**
Religionotherchristian	0.999	1.005	1.002
Religiontraditionalist/others	1.432	1.246	1.388

*Significant at 1%, **Significant at 0.1%, ***Significant at 0.001%

for AIC and Residual deviance for Clog-log model suggested it was associated with good fit. It fitted the data more appropriately than the Logit and Probit models thus, suggesting that the Clog-log model was the best of the three models for assessing the factors associated with infant mortality in Nigeria. This study also showed that Logit model was better than Probit for the data set used.

In the literature, the commonly used models were logistic, Cox regression models and Chi-squared test but logistic regression had been most frequently featured. The present study revealed that choosing Clog-log model as the model of good fit compared to others (Logit and Probit models), breastfeeding, region, place of delivery, maternal education, household members, assistance by nurses or relatives during birth, no assistance during birth, birth order and religion were important determinants of infant mortality in Nigeria.

Apart from the Clog-log model being the model of good fit compared to others (Logit and Probit models), it also recognised an additional factor, ie no assistance during delivery, as an important factor that was significantly and positively associated with infant mortality. That means lack of assistance during delivery was one of the important factors that predicted infant mortality in Nigeria. Mothers who had no assistance during delivery had 1.2 times more likelihood of infant mortality and this buttressed the importance of assistance during delivery, especially when conducted by a health care worker (OR=0.6), in the reduction of infant mortality in Nigeria. Further, this study had buttressed the role of supervision of births by trained health personnel in the reduction of infant mortality. [6] The difference in the roles of the factors associated with infant mortality in the present study compared with the previous studies could be explained in terms of the model of analysis adopted; previous studies used logit models while the present study used the Clog-log model. Clog-log model also identified "delivery assisted by relative" to have a slightly higher risk (OR=1.3) and at higher p-value (0.01%) compared to other models.

Most of the factors associated with infant mortality recognised by Clog-log model in the present study had earlier been reported in previous studies in Nigeria. [7 -9, 11 - 13] In addition, Clog-log model recognised such factors at higher levels of significance. The implication of this is that NDHS 2008 data were better described as having asymmetric distribution which is one of the properties of Clog-log model rather than the claimed standard logit probability distribution from previous studies. More so, place of delivery was distinctly recognised as a factor associated with infant mortality by Clog-log model unlike other studies from Nigeria using 2008 NDHS data.

The limitations to the study included factors such as memory recall bias by mothers or caregivers, given the retrospective nature of the data. The other limitation is that the respondents might not want to recall the events leading to the death of their child or children and this may lead to underestimation of infant mortality.

Conclusion

The study showed that Clog-log model was of best fit for modelling 2008 Nigeria Demographic and Health Survey data. Infant mortality in Nigeria from NDHS 2008 data followed Clog-log asymmetric distribution. Therefore, Clog-log model is recommended for researchers studying Infant mortality in Nigeria. The following factors: place of delivery (home), adequate breastfeeding ,assistance by professionals during delivery and household members significantly reduced infant mortality while women with primary or no education and birth order had statistically significantly increased effect on infant mortality in Nigeria as revealed by the model (Clog-log) used. In addition, only the Clog-log model identified assistance during birth as an important factor associated with infant mortality.

Authors' contributions: OKS conceived the research, extracted the data, did the analysis and interpretation, and wrote the first draft of the manuscript with part of the discussion. OOO participated in drafting this manuscript. DOJ made substantial input into the intellectual content of the manuscript.

Conflicts of Interests: None declared

Funding: None

Publication History: Submitted 14-December 2015; Revised 20-February 2016; Accepted 05-April 2016

Acknowledgement

The authors appreciate the National Population Commission and ICF Macro Calverton, Maryland, USA, for making the NDHS data available for use.

References

1. National Population Commission (NPC) [Nigeria] and ICF Macro. Nigeria Demographic and Health Survey 2008; Abuja, Nigeria: National Population Commission and ICF Macro, 2009.
2. National Population Commission (NPC) [Nigeria] and ORC Macro. Nigeria Demographic and Health Survey 2003; Calverton, Maryland: National Population Commission and ORC Macro. 2004: 107-108.
3. United Nations Development Programme (UNDP). Measuring human development: A primer. New York: United Nations Development Programme. 2007.
4. Countdown to 2015: Maternal, Newborn and Child Survival .Available at www.countdown2015mnch.org/documents/2012report/Nigeria. Accessed September 2015.
5. United Nations (UN). Sustainable Development Goals: 17 goals to transform our world. 2015. Available at www.un.org/sustainabledevelopment/health/ Accessed February 2015.
6. Akinyemi JO, Bamgboye EA, Ayeni O. Trends in neonatal mortality in Nigeria and effects of bio-demographic and maternal characteristics. BMC Pediatr 2015: 15 (36)
7. Kayode GA, Adekanmbi VT, Uthman OA. Risk factors and a predictive model for under-five mortality in Nigeria: evidence from Nigeria demographic and health survey. BMC Preg Childbirth 2012: 12(10). http://www.biomedcentral.com/1471-2393/12/10.
8. Izugbara C. Whose Child is dying? Household characteristics and U5M in Nigeria. South Afr J Child Health 2014: 8(1): 16-22.
9. Fagbamigbe AF, Alabi O. Differentials and correlates of Infant Mortality in Nigeria: A comparative survival analysis between North East and South West Nigeria. Int J Trop Dis Health 2014: 4(8): 869-886.
10. Adedini SA, Odimegwe C, Bamiwuye O, Fadeyibi O, De Wet N. Barriers to accessing health care in Nigeria: Implications for child survival. Global Health Action 2014; 7: 23499. Available at www.globalhealthaction.net/index.phplgha/article/view/23499 Accessed 2014.
11. Akinyemi JO, Bamgboye EA, Ayeni O. New trends in Under-five mortality determinants and their effects on child survival in Nigeria: A review of childhood mortality data from 1990-2008. Afr Pop Stud 2013: 27 (1): 25-42.
12. Adetoro GW, Amoo EO. A statistical Analysis of child mortality: Evidence from Nigeria. Journal of Demographic and Social Statistics 2014:1: 110-120.
13. Adebowale AS, Yusuf BO, Fagbamigbe AF. Survival probability and predictors for woman experience childhood death in Nigeria: "analysis of north-south differentials". BMC Public Health 2012:12:430.

14. Jackman S. Vuong's non-nested hypothesis test. R documentation. www.rdocumentation.org/packages/pscl/functions/vuong Accessed May 2016.
15. Robert I, Kabacoff. R- Generalized Linear Models. 2012:1-4. Available at www.statmethods.net/advstats/glm.html Accessed May 2016.
16. Brian S. Everitt, Torsten Hothorn. A handbook of Statistical Analyses using R 2015: 1-15. Available at https://cran.r-project.org/web/packages/HSAVR/index.html Accessed 2015.
17. Gujarati DN, Porter DC. Basic Economics. 5th Ed. Mc Graw-Hill International Edition; 2009: 553-573.
18. Christopher Manning. Logistic regression (with R) 2007:1-8. nlp.stanford.edu/~Manning/courses/ling289/logistic.pdf.
19. McCullagh P, Nelder JA. Generalized Linear Models. 2nd Ed. FRS London: Chapman and Hall.1989: 21-31.
20. Park, Hun Myoung. Regression Models for Binary Dependent Variables Using Stata, SAS, R,LIMDEP, and SPSS. Working Paper. The University Information Technology Services (UITS) Center for Statistical and Mathematical Computing, Indiana *University.* 2009. *Available at http://www.indiana.edu/~statmath/stat/all/cdvm/index.html.*
21. Murray L, Munz P. Binomial Link functions. http://www.stats.uwo.ca/faculty/bellhouse/links.pdf. Accessed December 2015.
22. Applied Categorical and Non-normal Data Analysis. Ed 231c: Complementary Log-Log models. www.Philender.com/courses/categorical/notes2/clog.html. Accessed December 2015.