FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 4 No. 4, December, 2020, pp 401 – 408
DOI: https://doi.org/10.33003/fjs-2020-0404-496

MODELLING THE DETERMINANTS OF UNDER-FIVE CHILD MORTALITY RATES USING COX PROPORTIONAL HAZARDS REGRESSION MODEL

MUSA M. C[1]., ASIRIBO O. E[2]., DIKKO H. G[1]., USMAN M[1]. AND SANI S. S[3].

[1]Department of Statistics, Ahmadu Bello University, Zaria
[2]Department of Statistics, Federal University of Agriculture, Abeokuta
[3]Department of Agronomy, Ahmadu Bello University, Zaria
Corresponding author's email: ummohammed67@yahoo.com

ABSTRACT

An under-five childhood mortality rates in Nigeria is still high, despite efforts of government at all levels to combat the menace. This study examined some factors that significantly affect under-five child mortality. A sample of mothers with children under the age of five from Nigeria Demographic and Health Survey data (NDHS, 2013 & 2018) was used to assess the effect of some selected predictor variables (or covariates) on childhood survival. Cox proportional hazards model is essentially a regression model popularly used for investigating the association between the survival time and one or more predictor variables. The results from final fitted Cox proportional hazards regression model that the covariates, contraceptive used by the mother, state of residence, birth weight of child and type of toilet facility used by the h-ousehold were found to be significantly associated with under-five survival in the North Central Region of Nigeria. All the calculations are performed using the R software for statistical analysis.

Keywords: Child-mortality, Covariates, Cox Regression, Kaplan Meier, P-value

INTRODUCTION

Under-five childhood mortality is an important demographic health and developmental issue for a number of reasons. It is a critical element in the calculation of overall mortality, since the highest risk of death and proportion of deaths occur during childhood. It is an important indicator of countries' general medical and public health conditions and consequently its level of socio-economic development. Its increase is not only undesirable but also indicative of a decline in general living standard (Anderson *et al.*, 2002).

Many studies have investigated the determinants of childhood mortality in different socio-economic settings. Different factors may determine childhood mortality depending on the socio-economic and environmental situations. Intense research was carried out in the 1970s and 1980s which led to more focus on socio-economic and demographic characteristics, in addition to the known biomedical determinants of child mortality. Frankenberg (1995) conducted a research, and found that, building more health facilities and adding to the number of doctors in a village would significantly reduce infant and child mortality risk. But Antai *et al.*, (2009) argued that, the number of health facilities notwithstanding, the use of maternal and child health services is largely determined by mother's indigenous religious affiliation and this significantly influences the risk of infant and child mortality.

Olatunji and Adesina (2016) used Cox proportional, logistic model to investigate the determinants of infant and child mortality in Nigeria. Their findings showed that the hazard and odds ratios of infant and child mortality are significantly less frequent over specified covariates, insignificant in residence type but significant in odds ratio. Also, there is an increased risk of infant and child mortality in place of delivery. It is evident from the results obtained that social economic risk factors contribute significantly to infant and child mortality in Nigeria.

Hong *et al.*, (2009) conducted a research on infant and under-five child mortality in Rwanda using the nearly two decades of Rwandan Demographic and Health Surveys data. They found that, in the case of infant mortality, the number of children ever born, birth interval, availability of professional antenatal and delivery care, full immunization of children, mother's education, and urban-rural residence were important determinants of infant and under-five child mortality. The same study also revealed that, in the case of under-five child mortality; multiplicity of births (i.e. number of births for each pregnancy), birth intervals, antenatal care and deliveries by health professionals, full immunization of children, mother's education, and use of contraception and possession of mosquito nets were also determinants of infant and under-five mortality. Pandey *et al.,* (1998) conducted a research to examined infant and child mortality in India, they found that sex of the child, mother's residence, mother's exposure to mass media, use of clean cooking fuel, mother's literacy status, access to a toilet facility, mother's religion and ethnicity, income of the household, birth order, mother's age at birth and mother's health care were important determinants of infant and child mortality in India.

Similar findings were reported in a study by Kumar and Gemechis (2010) using data from the 2005 Ethiopian Demographic and Health Survey. The study reported that birth interval, mother's literacy, household wealth, mother's age at birth, mother's exposure to mass media, sex of the child, religion, family size, birth order and residence were important predictors of infant and child mortality.

Using multi-country Demographic Health Survey (DHS) data, collected between 2000 and 2005, Rutstein (2008) conducted a study and found that waiting 36 months or more to have another pregnancy substantially decreases risk to death of children and under-nutrition.

Ayele, *et al.,* (2020). This study uses Cox proportional-hazards model to identify risk factors associated with under-five mortality in Sudan. This study uses the 2014 Sudan Multiple Indicator Cluster Survey (MICS) conducted by the Central Bureau of Statistics in collaboration with several national institutions. The results show that the weight of a child at birth is positively associated with the under-five child mortality rate. Under-five children who have both small and large weights at birth are at a higher risk of dying before reaching five years.

After carefully reviewing literatures, we discovered that there has not been much studies to find determinants factors of under-five child mortality in North Central Nigeria. Therefore, there is a need to do holistic study to find out factors that affect under-five mortality rate in the North Central Zone Nigeria.

MATERIALS AND METHODS

In this study, we employed Cox proportional hazards regression model to investigate the predictor variables of under-five mortality and Kaplan Meier (K-M) method were used to estimate the survival probability of under- five mortality.

Kaplan-Meier (K-M) Estimator

The Kaplan-Meier (KM) method is a non-parametric method used to estimate the survival probability from observed survival times (Kaplan and Meier, 1958).The K-M estimator is one of the most widely used survival models in the statistical and medical research literature. It is non parametric and thus very flexible in learning survival curves, however, the estimator has the disadvantage of not incorporating patients' covariates. The model may thus be used at the population level, but not for individualised risk predictions (Rietschel, 2018).

Survival analysis techniques were used to find the effect of determinants of under-five mortality.

K-M was used to estimate and graph the survival curve. The estimator incorporates information from all of the observations, censored and uncensored by considering survival at any point in time as a series of steps defined by the observed survival and censored times (Hosmer and Lemeshow, 1999).

The K-M estimator is derived from maximum likelihood estimation of hazard function.

$$S(t) = \prod_{i:t_i \leq t}^{i} (1 - h_j) \quad (1)$$

and the likelihood function for the hazard function up to time t_i is given as

$$L\left({}^{h_j}\!/_{d_j}, n_j \right) = \prod_{j=1}^{i} h_j^{d_j} (1 - h_j)^{n_j - d_j} \quad , j \leq i \quad (2)$$

Therefore, the log likelihood will be

$$\log L = \sum_{j=1}^{i} (d_j \log h_j + (n_j - d_j) \log(1 - h_j)) \quad (3)$$

Finding the maximum of log likelihood estimation given this result, we can have

$$\frac{\partial \log L}{\partial h_i} = \frac{d_i}{\widehat{h_i}} - \frac{n_i - d_i}{1 - \widehat{h_i}} = 0 \quad (4)$$

$$\hat{h} = \frac{d_i}{n_i} \quad (5)$$

Substituting equation (5) into (1) we get

$$\hat{S}(t) = \prod_{i:t_i \le t}(1 - \hat{h_i}) \quad (6)$$

$$= \prod_{i:t_i \le t}(1 - \frac{d_i}{n_i}) \quad (7)$$

were d_i as the number of events and n_i the total individuals at risk at time t_i, discrete hazard rate h_i.

Cox Proportional Hazard Regression Model

The Cox proportional-hazards model or Cox regression (Cox, 1972) is essentially a regression model commonly used in statistics and medical research for investigating the association between the survival time of patients and one or more predictor variables.

The Cox regression model is given by

$$\mathrm{h}(t) = h_o(t)exp\left(\beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k\right) \quad (8)$$

and the Cox model can be summarised as

$$h(t) = h_o(t)\exp(\beta^T X) \quad (9)$$

where h_o(t) is the baseline hazard at time t, representing the hazard for a subject or person when all explanatory variables are 0, x_1 ... x_k are the collection of explanatory variables and β_1,…….. β_k are the collection of regression coefficients.

A positive regression coefficient (i.e. Hazard Ratio > 1) for a predictor variable, indicates that the hazard of event or death is higher, i.e., increases the risk of event or death. Conversely, a negative regression coefficient (Hazard Ratio < 1) implies a decrease risk of event or death.

Cox Proportional Hazards Assumption

The proportional hazards assumption is supported by a non-significant relationship between residuals and time, and is refuted or not supported by a significant relationship between the residual and time.

Stratified Cox Proportional Hazard Regression Model

The stratified Cox regression model is a modification of the Cox proportional hazards model that allowed for control by stratification of a predictor that does not satisfy the proportional assumption. Predictors that are assumed to satisfy the assumptions are included in the model whereas the predictors being stratified is not included (Kleinbaum and Klien, 2005).

Stratification involves fitting a model that has a different baseline hazard in each stratum. This model allows for several variables being stratified through the use of a newly defined variable called Z*, whose strata (k*) consist of combinations of categories of the variables being stratified. K* is the total number of combinations formed after categorizing each of Z's

Let the variables $Z_{1}, \ldots, Z_{k}$ do not satisfy the proportional hazards assumption

Let the variables $X_{1}, \ldots, X_{k}$ satisfy the proportional hazards assumption

To create a single new variable Z^*, the following steps are adopted.

I. Categorise each Z_i

II. Form all the possible combination of the categories

III. The strata are the categories of Z^*

Therefore, the general stratified Cox regression model will be:

$$h_g(t,x) = h_{og}(t)exp\left(\beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k\right) \quad (10)$$

g = 1, 2, . . . , k*, strata defined from Z^*

The regression coefficients *(β)* are estimated by maximising the partial likelihood function obtained by multiplying likelihood functions for each strata.

RESULTS AND DISCUSSION

We used non-parametric estimator of the survival function, the Kaplan Meier (KM) method to estimate and graph survival probability as a function of time. It is also used to estimate univariate descriptive statistics for survival data, including the median survival time and compare the survival experience for two or more group of subjects. In Figure 1, the vertical axis represents estimated probability of the survival while the horizontal axis represents the time which is measured in months in this study.

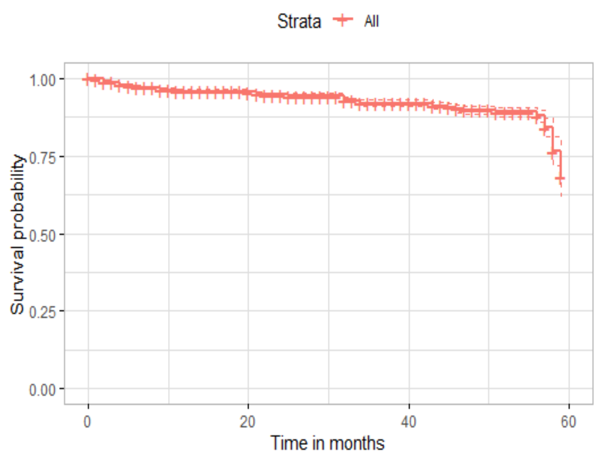

Figure 1: Kaplan Meier Curve for NDHS 2013 data

Figure 1 shows the KM survival curve of the probability of children under the age of five surviving, which was high at the first month of birth and decreases gradually as the time increases.

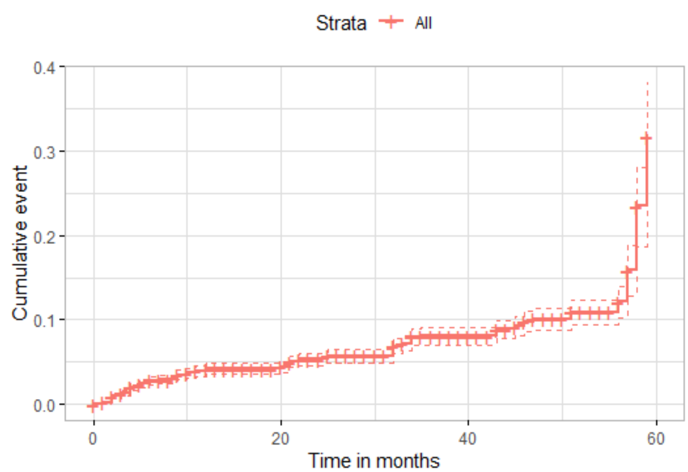

Figure 2: Cumulative Event Curve for NDHS 2013 Data

From Figure 2, the cumulative event (or death) curve rises almost immediately from 0 month until about 10 months and continue rising slowly thereafter up to around 0.12 probability on the vertical axis in 59 months on the horizontal axis.

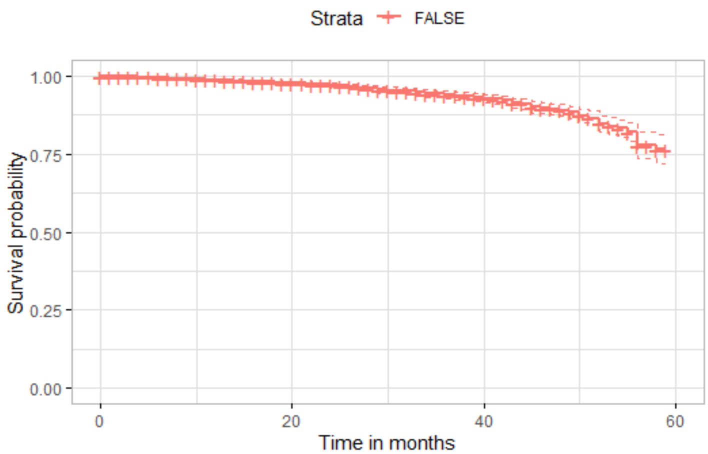

Figure 3: Kaplan Meier Curve for NDHS 2018 Data

In Figure 3 the KM survival curve shows the probability of children under the age of five surviving. The probability was high at the first month of birth and decreases slowly as the time increases.

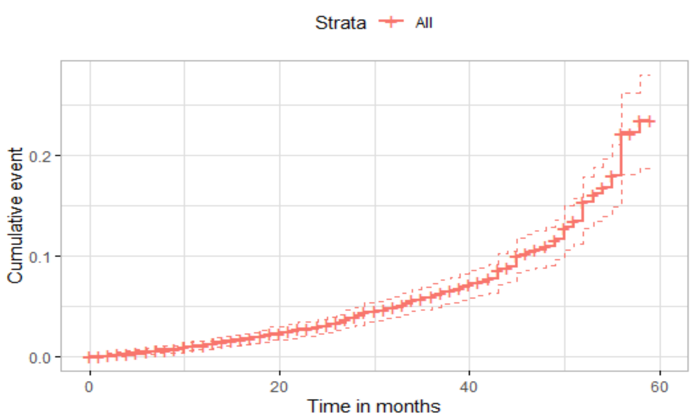

Figure 4: Cumulative Event Curve for NDHS 2018 Data

From Figure 4 the cumulative event curve rises slowly from 0 month until about 10 months and rise high thereafter up to around 0.18 probability in 59 months.

Table 1a: Results of Stratified Cox Proportional Hazard Regression Model 2013 NDHS

Factor	Coef.	ExpCoef.)	lower 0.95	Upper 0.95	Pr(|Z|)
State	0.0001580	1.0001580	0.9867	1.014	0.9818
Place_residence	0.3361615	1.3995651	0.9243	2.119	0.1123
mother_age_birth	-0.1333618	0.8751484	0.7355	1.041	0.1326
Mother_education	-0.1684762	0.8449514	0.7138	1.000	0.0502
Source_water	0.0010896	1.0010902	0.9942	1.008	0.7575
Type_toilet	0.0124590	1.0125369	1.0020	1.023	0.0192 *
household_income	0.0944950	1.0991036	0.9387	1.287	0.2404
contraceptive_use	-0.0270348	0.9733274	0.9131	1.038	0.4069
Birth_weight	-0.0008864	0.9991140	0.9862	1.012	0.8940
Child_sex	0.1184356	1.1257344	0.8541	1.484	0.4005

Table 1b: Results of Overall Test for Model Significant of Stratified Cox Proportional Hazard Regression Model 2013 NDHS

Test	Test-value	DF	P- value
Likelihood ratio test	20.57	10	0.02
Wald test	22.24	10	0.01
Score (logrank) test	23.58	10	0.009

In table 1b, the p-value for all three overall tests (likelihood, Wald, and score) are less than 0.05, indicating that the model is significant. These tests evaluate whether the coefficient of a given variable is statistically significantly different from 0. In the stratified Cox proportional hazards regression model 2013 NDHS, the test statistics are in close agreement.

Also in table 1a, the covariate toilet facility is statistically significant (P-value < 0.05) after controlling the effects of birth interval and birth order in the model. However, the covariates (state of residence, type of residence, mother's age at birth, mother's level of education, household income, child birth weight, sex of the child, source of water) were not statistically significant (P-value > 0.05).

Table 2: Test Results for Stratified Cox Proportional Hazard Regression Model Assumption

Factor	Coef.	ExpCoef.)	lower 0.95
State	-0.0862	2.145	0.1430
Place_residence	0.0740	1.421	0.2332
mother_age_birth	-0.0608	0.896	0.3439
Mother_education	0.0292	0.208	0.6480
Source_water	-0.0343	0.254	0.6145
Type_toilet	0.0516	0.588	0.4433
household_income	-0.1188	3.115	0.0776
contraceptive_use	-0.0293	0.252	0.6155
Birth_weight	-0.0427	0.469	0.4933
Child_sex	-0.0302	0.198	0.6564
GLOBAL	NA	17.038	0.0735

From the test results of stratified Cox proportional hazard model in Table 2, the test is not statistically significant for all the covariates, and the global test is also not significant. Hence, we can assume the proportionality of the hazards model.

Table 3a: Results of Cox Proportional Hazard model for 2018 NDHS

Factor	Coef.	ExpCoef.)	Lower 0.95	Upper 0.95	Pr(>|Z|)
State _residence	0.010953	1.011014	1.0023	1.0198	0.012796 *
Place_residence	0.129034	1.137729	0.7839	1.6513	0.497243
Mother_ag_birth	-0.014840	0.985270	0.9419	1.0307	0.518628
Mother_education	0.120714	1.128302	0.9067	1.4040	0.279127
Source_water	0.003423	1.003429	0.9941	1.0129	0.473295
Type_toilet	-0.007100	0.992926	0.9725	1.0138	0.503221
Household_income	-0.039495	0.961274	0.8000	1.1551	0.673417
Contraceptive_use	-0.112246	0.893824	0.8264	0.9667	0.005020 **
Birth_interval	-0.004328	0.995681	0.9862	1.0053	0.376026
Birth_order	0.071435	1.074049	0.9877	1.1680	0.094955
Birth_weight	0.195467	1.215879	1.0849	1.3626	0.000773 ***
Child_sex	-0.109016	0.896716	0.6389	1.2586	0.528559

Table 1b: Results of Overall Test for Model Significant of Cox Proportional Hazard Regression Model 2018 NDHS

Test	Test-value	DF	P- value
Likelihood ratio test	34.03	12	0.0004
Wald test	32.32	12	0.001
Score (logrank) test	32.99	12	0.001

In table 3b, the p-value for all three overall tests (likelihood, Wald, and score) are less than 0.05, indicating that the model is significant. These tests evaluate, whether the coefficient of a given variable is statistically significantly different from 0. From the results of Cox proportional hazard model for 2018 NDHS, the test statistics are in close agreement.

From Table 3a, the covariates state of residence, contraceptive use and birth weight are statistically significant (P-value < 0.05). However, the covariates type of residence, mother's age at birth, mother's educational level, source of water, household income, sex of child, birth interval and birth order are not significant (P-value > 0.05).

The P-value of contraceptive use is 0.005020 with hazard ratio HR=0.893824 indicating a significant relationship between contraceptive use by the mother and under-five child survival. In other words, the children born from mother who use contraceptive are more likely to survive than children who are born to mother who doesn't use contraceptive.

Similarly, state of residence and birth weight are significant variables, but these variables are associated with poor survival because of their positive coefficients and thus increased the risk of under-five mortality.

Table 4: Test Results for Cox Proportional Hazard Regression Model Assumption

Factor	Rho	Chisq	P- value
State _residence	0.0886	1.2619	0.2613
Place_residence	-0.0205	0.0591	0.8079
Mother_ag_birth	-0.0261	0.1129	0.7369
Mother_education	-0.2189	6.4076	0.0114
Source_water	0.0377	0.2307	0.6310
Type_toilet	-0.0303	0.0715	0.7892
Household_income	0.2075	4.4914	0.0341
Contraceptive_use	0.1309	2.8609	0.0908
Birth_interval	-0.0925	1.6455	0.1996
Birth_order	-0.0711	0.5681	0.4510
Birth_weight	-0.0841	1.0803	0.2986
Child_sex	-0.0293	0.1191	0.7301
GLOBAL	NA	16.7274	0.1601

From the test results of Cox proportional hazard model in Table.4, the test is not statistically significant for the global test (P value > 0.05). Hence, we can assume the proportionality of the hazards model.

CONCLUSION

In conclusion, this study was able to identify the factors which affect under-five mortality in North Central Geo-Political Zone of Nigeria, using Nigeria Demographic and Health Survey Data of 2013 and 2018 respectively. The results of both model show that contraceptive used of mother, birth weight and type of toilet facility being used by the family and place of residence were significantly associated with under-five child mortality. Based on the research findings, there should be concerted efforts to address these factors in order to reduce the under-five mortality. There should also be new policies and interventions to focus together with existing policies and intervention strategies that need to be strengthened in order to reduce under-five child mortality to the lowest level.

REFERENCES

Anderson, B. A., Romani, J. H., Phillips, H. E. and Van Zyl, J. A. (2002). Environment, Access to Health Care, and Other Factors Affecting Infant and Child Survival in South Africa, 1989–1994. *Population and Environment*, 23(4), 349-364.

Antai., D., Ghilagaber, G, Wedrén, S, Macassa, G, and Moradi, T. (2009). Inequities in under-five mortality in Nigeria: differentials by religious affiliation of the mother. *Journal of Relig Health.* , 48(3), 290-304

Ayele, D. G., Satty, A., and Zewotir, T. (2020). Understanding the Correlates of Under-five Mortality in Sudan Using Survey Survival Models. *Asian Journal of Research in Infectious Diseases*, *5*(1), 55-68. https://doi.org/10.9734/ajrid/2020/v5i130160

Cox D. R. (1972). Regression Model and Life Tables. *Journal of Royal Statistical Society* (B), 34: 187-220

Frankenberg E. (1995).The effects of access to health care on infant mortality in Indonesia. *Health Transition Review*, 5, 143-163.

Hong, R., Ayad, M., Rutstein, S and Ren, R. (2009). Childhood Mortality in Rwanda Levels, Trends, and Differentials. *Further Analysis of the Rwanda Demographic and Health Surveys*, ICF Macro Calverton, Maryland, USA.

Hosmer D. W. and Lemeshow S. (1999). Applied Survival Analysis: New York USA.

Justice M. K. A. (2019) Predictive model and determinants of under-five child mortality: evidence from the 2014 Ghana demographic and health survey. *BMC public health* 19:16

Kaplan E. L. and Meier P. L (1958). Non Parametric Estimation from Incomplete Observation. *Journal of American Statistical Association* **53: 4**57-481

Kleinbaum D. G. and Klien M. (2005). Statistics for Biology and Health: Survival Analysis *A Self-Learning Text* Second Edition.

Kumar, P. P. and Gremechis F. (2010). Infant and Child Mortality in Ethiopia: A Statistical Analysis Approach, *Ethiopian Journal of Education and Science*, 5(2):51-7

National Population Commission. (2013) Nigerian Demographic and Health Survey NDHS

National Population Commission. (2018) Nigerian Demographic and Health Survey NDHS

Olatunji A. P and Adesina O. A. (2016). Effect of Determinants of Infant and Child Mortality in Nigeria: Hazard and Odds Ratio Models. *West African Journal of Industrial & Academic Research Vol.16 No.1 December 2016*

Pandey, A, Choe, M.K., Luther, N.Y., Sahu, D. and Chand, J. (1998). Infant and Child Mortality in India. Mumbai, India and Honolulu. USA:NFHS Subject Report No. 11. IIPS and East-West Center, *Population and Health Studies*.

Rietschel, C. (2018). Automated feature selection for survival analysis with deep learning [Master's thesis].

Rutstein, S. O. (2008).Further Evidence of the Effects of Preceding Birth Intervals on Neonatal, Infant, and Under-Five-Years Mortality and Nutritional Status in Developing Countries: Evidence from the Demographic and Health Surveys.