MODELLING UNDER-FIVE MORTALITY USING THE COX PROPORTIONAL HAZARD REGRESSION MODEL: A CASE IN GHANA

Abstract

Ghana, like many other developing nations in Africa, grapples with a substantial under-five mortality rate. As a result, it falls short of fulfilling the United Nations' sustainable and millennium development goals, which are aimed at improving child health and reducing child mortality. To lay the groundwork for effective strategies and well-informed policy interventions, it is essential to utilize robust statistical models capable of pinpointing the primary factors significantly impacting under-five mortality rates. In this study, we employ the Cox proportional hazard regression model to model the under-five mortality in Ghana using data from the IPUMS-DHS from 2009 to 2014. The study's findings indicate several critical determinants that contribute to the alarming under-five mortality rate in Ghana. Specifically, marital status and the current employment status are associated with decreasing the mortality for under-five while the educational level of the mother, the year of the child's birth, the gender of the child, and the economic status of the household have been identified as substantial factors driving the high under-five mortality rate in Ghana. On the contrary, marital status and current employment status have shown associations with a reduction in under-five mortality. These findings underscore the pressing need for strategic investments in the healthcare system to enhance child health. Moreover, providing mothers with comprehensive knowledge on childcare and birth spacing emerges as a promising avenue for curtailing under-five mortality.

Keywords.

Under-five mortality; Cox proportional hazard regression model; Childcare; Ghana; Maternal factors.

I. INTRODUCTION

Children in Sub-Saharan Africa face a risk of mortality before reaching the age of 5 that is over 15 times higher than that of children in high-income nations (Apanga & Kumbeni, 2021). In Ghana and similar developing African countries, the persistent challenge of high under-five mortality rates poses a formidable barrier to achieving United Nations sustainable and millennium development goals aimed at enhancing child health and decreasing child mortality. Despite the government of Ghana's policies aimed at achieving the Millennium Development Goal 4 to reduce child mortality by 2015, the country did not meet the goal due to an upsurge in child mortality within the country. These challenges can be attributed to various factors including the maternal conditions (Azaare et al., 2022). The educational level of the mother can be a significant factor (Buor, 2003), and the income level, marital status, place of stay, being rural or urban can all influence the survivorship of under-five babies.

As a result, many researchers have tried to develop various statistical models to investigate under-five mortality including the use of the Cox proportional hazard regression model. For instance, Ayele et al. (2017) employed the Cox proportional hazard regression model for fixed and time-dependent explanatory factors to investigate under-five mortality in Ethiopia. According to their research, urban children under the age of five are at lower risk than those who live in rural areas. In a study conducted by (Nasejje & Mwambi, 2017), both the random survival forests and the Cox proportional hazards model were employed to examine the factors influencing under-five child mortality rates in Uganda, using data from the Demographic and Health Survey. The outcomes of their investigation revealed a consistent consensus between the two models. Specifically, they found that the sex of the household head, the sex of the child, and the number of births within the preceding year exhibit strong associations with under-five child mortality in Uganda, under the condition that all three covariates adhere to the proportional hazard's assumption. A study conducted by Musa et al. (2020) aimed to identify significant factors contributing to under-five child mortality. They utilized data from the Nigeria Demographic and Health Survey (NDHS) to carry out their investigation. Their study findings revealed that specific covariates hold considerable statistical significance in relation to under-five survival rates within the North Central Region of Nigeria. Notably, key factors such as the use of contraceptives by mothers, the geographical state of residence, the child's birth weight, and the type of toilet facilities used by households were found to be influential in this regard. These results emphasize the intricate connections that exist between socio-demographic variables and child survival outcomes, particularly when considered in the specific regional context specified by the study. Alotaibi et al. (2020) investigated under-five mortality determinants preceding the conclusion of the 2015 Millennium Development Goals, aligned with the target of Sustainable Development Goal 3 by 2030. Using Demographic and Health Survey (DHS) data from twelve West African countries (Ghana, Benin, Cote d'Ivoire, Guinea, Liberia, Mali, Niger, Nigeria, Sierra Leone, Burkina Faso, Gambia, and Togo), they applied Bayesian regression models including exponential, Weibull, and Gompertz methods through a gamma shared frailty model. The study's outcomes highlighted Gambia, Ghana, and Benin with the lowest under-five mortality rates per 1000 live births. Furthermore, their analysis revealed significantly higher hazard rates for male children when compared to females across the pooled data. Ogbo et al., (2019) employed the cox-regression model to analyze mortality trends in neonatal, postnatal, infant, child, and under-five categories in Tanzania spanning 2004 to 2016. Their findings revealed a stable neonatal mortality rate, whereas post neonatal and child mortality rates experienced a significant 50% reduction during this period. Asare et al. (2021) employed a prospective observational method to examine the medical records of 601 children who underwent treatment for severe acute malnutrition (SAM) at Ghanaian referral hospitals. The outcomes indicate that distinct groupings of medical conditions and clinical indicators in children diagnosed with SAM correlate with different levels of mortality risk and clinical outcomes.

While various studies have explored factors influencing under-five mortality, an essential gap remains in the literature concerning a comprehensive investigation using robust statistical methods, particularly those that account for nuanced demographic, social, and economic aspects. Although previous research has delved into some potential determinants of under-five mortality, few studies have considered the educational level of mothers, current employment status, marital status, urban-rural status, the child's sex, year of birth, and resident status of mother, which could play a significant role in the observed rising under-five mortality rates. This study aims to address these significant gaps by applying the Cox Proportional Hazard Model to scrutinize the factors contributing to under-five mortality rates in Ghana. Focusing on variables that encompass demographic, social, and economic dimensions, the research seeks to illuminate the intricate relationships between these variables and under-five mortality, employing robust statistical methodologies. Through this investigation, the study endeavors to not only contribute to the existing body of knowledge on under-five mortality but also offer a refined lens through which the determinants of child mortality in Ghana can be viewed. By revealing insights into the factors associated with both heightened and diminished under-five mortality rates, the findings aim to provide actionable insights for policymakers, healthcare professionals, and stakeholders tasked with devising targeted strategies to mitigate under-five mortality in Ghana and analogous contexts. The rest of the study's sections are structured as follows: Chapter 2 presents the materials and methods employed, chapter 3 presents the results, and a discussion and chapter 4 conclude the study with a recommendation.

II. METHODOLOGY

Data

The data used was obtained from IPUMS Global Health, whichis an online site that provides integrated international health survey data for research and educational purposes. It is made up of two data series: the Demographic Health Survey (DHS) and thePerformance Monitoring for Action. In the conventional mortality measurement, one requirement is to get information on the number of deaths and the population subject to dying. Hence the data selected satisfies the requirements for this study. The data facilitates the analysis of Demographic and Health Surveys administered in low and middle-income countries since the 1980s. We derived the 2014 demographic and health survey with Ghana as the country of interest. The sample size was 5884 with birth histories of a child at a given point. We identified the number of children born in the last five years and the age at which they die. Past information was also obtained about the number of children that died in the previous five years preceding the survey.

Outcome variable

The primary outcome variable of interest in this study was child survival status which is the time (which is the age at death) until an event occurs, categorized as dead (as 1) and alive or censored (as 0). Observations in this study are censored, meaning that the event of interest didnot occur when the data was analyzed.

Explanatory variables

The explanatory factors, often referred to as covariates, are those variables whose impact on under-five mortality we want to examine. The maternal education level, current employment status, marital status, urban-rural status, child's sex, year of birth, and resident status all impactunder-five mortality. Categorical and continuous variables are included in the covariates for this study

The Cox Regression Model

The Cox regression model, often called the Cox proportional hazard model, is a semi-parametric survival model that measures the relationship between one or more covariates withtime (Nurmalasari et al., 2019). The Cox regression method is one method that is time- dependent and can be used to check the data between an incident and the data affected by the censor. The model is used in predicting the relationship between dependent variables and failure time with the independent variables. Based on the values of input covariates, the Cox regression model helps model the time of a specific event. The model is often used to analyzecovariate information that changes over time with the hazard proportional to the instantaneousprobability of an event at a particular time (Ngwa et al., 2016). The Cox regression model is a semi-parametric model that examines the relationship between time and an event with independent variables. It is straightforward, flexible, and can present adjusted and unadjusted hazard ratios with confidence intervals, making it suitable for analyzing daily or seasonal passage risks. The ranks of the covariate values are used to estimate hazards because the model is semi-parametric. Because of this, it is impossible to model quantitative differences between treatments using the conventional regression model.

The Cox regression model is the most used method for survival analysis because it is simple and does not requirethe assumption of the hazard function. Instead, it assumes that the hazard function is an independent variable but does not know the form of the hazard function. The formula for the Cox regression model is thus given.

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t) \cdot \exp\left(\sum_{i=1}^{p} \beta_i Z_i\right) \tag{1}$$

where, $\lambda(t|\boldsymbol{X})$ represents the hazard function at time t given the vector of covariates Z**,** the baseline hazard function $\lambda_0(t)$ also describes the hazard when all covariates are equal to zero, and β_i= (β_1, β_2, β_3, . . . , β_p are the respective coefficients.

exp ($\beta\underline{Z}$) is a relative risk function that is not time-dependent; that is, the effect of the explanatory variable either increasing or decreasing the risk remains constant and does not change given a change in time. The model is semi-parametric because it contains both parametric and non-parametric parts. The parametric part is due to the parameter in the model,which is β. Still, the failure time distribution is assumed to be known, and the non-parametricpart is the λ_0, an unspecified function in the equation above.

Assumption of the Cox Regression Model

The assumption of the Cox regression model is related to proportional hazards, and the assumption states that the hazard ratio should be constant over time. This means that the hazardratios for every individual should be proportional to each other and approximately be the sameduring follow-up. In practice, the assumption of the proportional hazard is likely to be satisfied, and a breach of the assumption may result in wrong and misinterpreted estimates (Stel et al., 2011). An example is a case where the survival curves of two groups cross; the hazard ratio for these two groups clearly will not be the same, making the use of the Cox regression model with proportional hazard inappropriate. The Cox regression model analyzes survival data whenthe effect estimate is adjusted for confounders.

The Hazard Ratio

The hazard ratio establishes the relationship between the two groups. It is defined as the ratioof the hazard rates corresponding to the conditions described by two levels of an explanatory variable. It measures the instantaneous risk associated with two groups over the time of the study. A hazard ratio of 1 means that there is no risk, a hazard ratio above 1 implies an increase in the hazard risk, meaning that the covariates are positively related to the probability of the eventoccurring, whiles a hazard ratio below one indicates a reduction in the hazard risk meaning that the covariates are negatively associated to the likelihood of the event occurring. This is summarized as,

1. HR=1: No hazard
2. HR<1: Reduction in hazard risk
3. HR>1: Increase in hazard risk.

If we take both individuals k and k+1 understudy both with different z values, the hazard ratiothat corresponds to these two individuals can be written as follows.

- The hazard function for the individual k is given as:

$$h_k(t) = h_0(t)exp\left(\sum_{i=1}^{k} \beta_i Z_i\right) \tag{2}$$

- The hazard function for the k+1 individual is given as:

$$h_{k+1}(t) = h_0(t)exp\left(\sum_{i=1}^{k+1} \beta_i Z_i\right) \tag{3}$$

- The hazard function for the two individuals

$$\frac{h_{k+1}(t)}{h_k(t)} = \frac{h_0(t)exp\left(\sum_{i=1}^{k+1} \beta_i Z_i\right)}{h_0(t)exp\left(\sum_{i=1}^{k} \beta_i Z_i\right)} = \frac{exp\left(\sum_{i=1}^{k+1} \beta_i Z_i\right)}{exp\left(\sum_{i=1}^{k} \beta_i Z_i\right)} \tag{4}$$

From equation 4 above, we can say that the hazard ratio is independent of time, and by the assumption stated earlier, we see that the hazard ratio is constant over time and proportional for two individuals or groups. Likewise, the hazard function for one individual can be written as

$$\frac{h_k(t)}{h_0(t)} = exp\ (\beta_i Z_i) \tag{5}$$

This means that the hazard of the kth individual is a fixed proportion of the baseline hazard.

We can then write the survival function for the kth individual as a constant power of the baseline survival function, that is.

$$S(t, z) = S_0\ (t)\ exp(\beta_i Z_i) \tag{6}$$

The result is interpreted as risk variables concerning baseline risk, with baseline risk defined as S_0(t) being the important life function.

Statistical tests and confidence intervals

In using the Cox regression model, we may want to test whether the null hypothesis β= β_0 is statistically significant or not; that is, we test if β_0=0 meaning that our β=0. A different test could be used. For the Cox regression model, there are three types of tests: the Wald test, the likelihood ratio test, and the score test statistic, which is explained briefly below.

The Wald Test Statistic

The Wald test is a test for estimating one model. That is, it is a test for individual hazard ratio.The Wald test is a close approximation of the likelihood ratio test. It determines how far the calculated parameters are from zero. In this case, whether the individual hazard coefficient is zero. The Wald test can test several factors simultaneously, whereas the tests reported in regression output usually only test one parameter at a time. The Wald test's null hypothesis is as follows:

H_0: $\beta_k = 0$.

The formula for the Wald test is given as

$$W=\left[\frac{\beta_k}{\text{s.e}(\beta_k)}\right] \qquad (7)$$

The Wald test is based on one degree of freedom chi-square distribution, with s.e. as the standard error.

The Likelihood Ratio Test Statistic

It compares the log-likelihoods of two models by looking for differences between them and seeing if they are statistically significant. This is because the difference between two log-likelihoods is equal to the log of the ratio of the two likelihoods test is called a ratio rather than a difference. As an example, the likelihood ratio test is defined as follows:

$$LR= -2\ln\left(\frac{L(m_1)}{L(m_2)}\right) = 2(LL\ (m_2) - LL\ (m_1)) \qquad (8)$$

L ($m*$): represents the likelihood of the respective model ($either\ m_1 or\ m_2$).

LL ($m*$): represents the natural log of the model's (1$either\ m_1 or\ m_2$) final likelihood.

$m_1 or\ m_2$: represent the more restrictive and the less restrictive model respectively.

The likelihood ratio has a chi-square distribution with degrees of freedom equal to the number of model parameters. It is used to obtain probabilistic values (P-values) to test for significant parameters. It is used to test the estimated model.

The Score Test Statistic

The score test also known as the Lagrange multiplier is a test that is equivalent to a logarithm rank test, except in the model estimated with it, it does not include the parameter(s) of interest. It is used to calculate how much a set of variables in the model has improved. The following is a list of the criteria for the score test:

$$SC = U'V^{-'}U \tag{9}$$

Where, $U=\frac{\partial}{\partial\beta} log\ L\ (\beta)$ (10), denotes the vector score functions, and $V= \left[\frac{\partial^2}{\partial\beta_1\beta_2}\right]$.

The score test uses a chi-square distribution with degrees of freedom equal to the estimated model's number of parameters. The confidence interval for a Cox regression model can be obtained at the relative risk as:

$$\widehat{\beta_k} \pm |Z\alpha_{/2}|\ \text{s.e}\ (\widehat{\beta_k}) \tag{12}$$

The expression above gives the upper and the lower bound which is known as the two-sided confidence interval with s.e. being the standard errors.

III. RESULTS AND DISCUSSIONS

This chapter thoroughly explains and interprets the results obtained from the analysis. The demographic features of children under five years and significant findings generated from the Cox regression model are indicated in this report. The model used for fitting the data is the Cox-regression model, and the software used was the R package. Social, economic, demographic, and geographic characteristics were included as covariates. These covariates can be summarized as follows: urban-rural status, residence status, marital status, education level, wealth/ financial status, current employment status, sex, and the child's birth year. The data were fitted with a Cox regression model to determine the factors influencing under-five child mortality in Ghana.

Descriptive Statistics

The mortality rates for both males and females, as well as the entire population, were calculated. Table 3.1 shows the mortality rates calculated for the whole population, females and males, and a brief explanation for each, with age ranges of 5 years apart. A conclusion drawn from the table below is that male under-five children had higher mortality rates than females, except for some or two cases. Taking 2010 as an example, wecan see that females have more excellent fatality rates than males. For the population, it was noticed that mortality continued to rise year after year, indicating that there are more deaths among children under the age of five in the population. This could be dueto a faulty health system, malnourished children, poverty on the part of parents, and other variables that significantly impact under-five mortality. As a result of the table, we may conclude that females under five have higher rates than males. Table 1 to 3 summarizes mortality rates for males and females, as well as the overall population of children under the age of five years for the years 2009 and 2010, 2011 and 2012, 2013 and 2014 respectively in five years intervals. From the 2009 chart in Figure 3.1, we see the mortality rates for the entire year were low. Both the males and females under five roughly have the same rate, and the mortality rate for the population in this year was likewise the same as the male and females but saw a rise for ages 31-35 months.

AGES	2009			2010		
(MONTHS)	**MALE**	**FEMALE**	**POP**	**MALE**	**FEMALE**	**POP**
15 -20	0.00	0.00	0.00	1.00	1.00	2.00
21 -25	0.00	1.00	1.00	1.00	4.00	5.00
26 -30	0.00	0.00	0.00	3.00	1.00	4.00
31 -35	1.00	1.00	2.00	1.00	4.00	3.00
36 -40	0.00	0.00	0.00	1.00	4.00	5.00
41 -45	0.00	0.00	0.00	3.00	1.00	3.00
46 -49	1.00	0.00	1.00	0.00	1.00	1.00

Table 1. mortality rates for males and females for the years 2009 and 2010.

AGES	2011			2012		
(MONTHS)	MALE	FEMALE	POP	MALE	FEMALE	POP
15 -20	1.00	1.00	2.00	0.00	0.00	0.00
21 -25	4.00	2.00	5.00	2.00	3.00	5.00
26 -30	1.00	3.00	4.00	1.00	3.00	4.00
31 -35	1.00	2.00	3.00	3.00	2.00	5.00
36 -40	1.00	4.00	5.00	3.00	2.00	5.00
41 -45	2.00	1.00	3.00	0.00	2.00	2.00
46 -49	1.00	0.00	1.00	1.00	1.00	1.00

Table 2. mortality rates for males and females for the years 2011 and 2012.

AGES	2013			2014		
(MONTHS)	MALE	FEMALE	POP	MALE	FEMALE	POP
15-20	2.00	2.00	3.00	0.00	2.00	2.00
21-25	4.00	2.00	4.00	4.00	1.00	5.00
26-30	3.00	2.00	5.00	1.00	3.00	4.00
31-35	1.00	3.00	4.00	2.00	2.00	4.00
36-40	3.00	2.00	4.00	2.00	3.00	5.00
41-45	3.00	1.00	3.00	0.00	0.00	0.00
46-49	0.00	1.00	1.00	0.00	0.00	0.00

Table 3. mortality rates for males and females for the years 2013 and 2014.

The distribution of ages by gender (males and females) from 2009 to 2014 as shown in Figure 1, reveals an interesting pattern, particularly in the age group of 31 to 35 months. In each of these years, this age group consistently exhibited the highest distribution of children. Notably, this trend was most pronounced in 2009, 2010, 2011, and 2014. In these years, the number of males in the 31 to 35-month age group was consistently higher than that of females, except for 2009 when both genders had an equal number in this age category. These findings could potentially indicate variations in child health and survival patterns among different age groups and genders during this period. Further analysis and modeling using the Cox Proportional Hazard Regression model could shed light on the factors influencing under-five mortality in Ghana during these years, potentially uncovering gender-specific disparities that warrant attention and intervention.

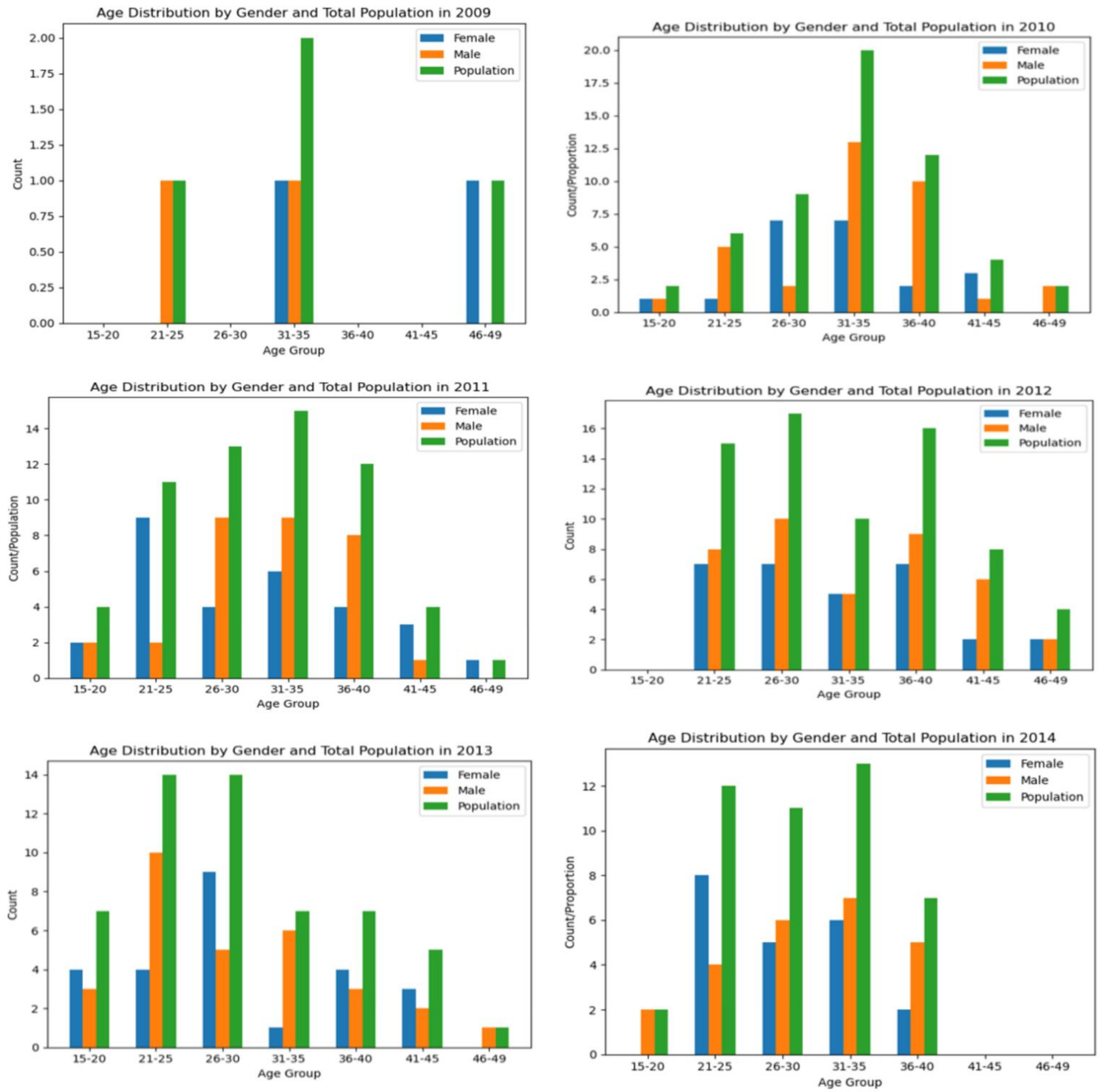

Figure 1. Age distribution by gender with respect to the population from 2009 to 2014

Survival Analysis

The study focuses on the time to event or the time from birth to death in children under five.Survival analysis was the best strategy to employ for this study. The Kaplan- Meier estimator and the Cox proportional hazard model were used for the analysis. To compare survival functions and measure their statistical significance, the likelihood ratio, the Wald test, and the score test were utilized with a p-value of 5%. The Kaplan Meier analysis was employed to investigate the survival pattern; the Kaplan Meier (KM) plot, a step function, provides insight into the under-five mortality survival distribution. The graphic indicates that if one survivor function is above another, the group described by theupper curve lived longer or had better survival outcomes than those defined by the lower curve.

Kaplan Meier

Table 4 below is a life table from the Kaplan-Meier approach, which summarizesthe number of events which is death that occurs among under-five children and the proportion of under-five survivors at a particular event time point. The time column gives the time at which death occurred, and the n.risk column also shows the number of under-fives at risk of dying before a given time, meaning that these children are either not dead or were not censored before or at a particular time. The n.event column providesan overview of children who have experienced at a time point that is a summary of under-five that are dead at a time point. Survival shows a proportion of under-five that are surviving after an event has occurred. The std. err, the lower and upper 95% CI is the standard error of the estimated survival, and the lower and upper 95% confidence bounds for the proportion of under-five surviving. The Kaplan Meier curve in Figure 1 shows a survival probability of 1 from the first year to the twentieth month, meaning that children aged between 1-20 months have a constant survival probability of 1, which neither is there a decrease in their survival nor an increase in their survival. Still, they struggle to survive afterward as their survival is not relatively stable. This is because their survival probabilitydecreases and approaches 0 when the child attains the age of 49 months. The decrease in survival may be caused by sickness due to malnutrition.

Time	n.risk	n.event	Survival	Std. err	Lower 95%CI	Upper 95% CI
16	5879	1	1.000	0.000170	0.999	1.000
18	5839	3	0.999	0.000342	0.999	1.000
19	5778	2	0.999	0.000420	0.998	1.000
20	5677	9	0.997	0.000674	0.996	0.999
21	5523	7	0.996	0.000825	0.995	0.998
22	5360	12	0.994	0.001045	0.992	0.996
23	5158	12	0.992	0.001237	0.989	0.994
24	4937	10	0.990	0.001388	0.987	0.992
25	4668	18	0.986	0.001649	0.983	0.989
26	4322	13	0.983	0.001837	0.979	0.986
27	4058	10	0.980	0.001986	0.976	0.984
28	3769	7	0.979	0.002098	0.974	0.983
29	3448	10	0.976	0.002276	0.971	0.980
30	3174	24	0.968	0.002712	0.963	0.974
31	2772	7	0.966	0.002858	0.960	0.971
32	2511	16	0.960	0.003227	0.953	0.966
33	2205	15	0.953	0.003619	0.946	0.960
34	1985	10	0.948	0.003906	0.941	0.956
35	1765	19	0.938	0.004512	0.929	0.947
36	1443	21	0.925	0.005340	0.914	0.935
37	1216	7	0.919	0.005676	0.908	0.930
38	1039	9	0.911	0.006216	0.899	0.923
39	857	6	0.905	0.006696	0.892	0.918
40	704	11	0.891	0.007832	0.875	0.906
41	497	6	0.880	0.008883	0.863	0.897
42	411	10	0.859	0.010947	0.837	0.880
44	230	1	0.855	0.011518	0.832	0.877
45	175	4	0.835	0.014830	0.806	0.864
46	105	1	0.827	0.016687	0.795	0.860
47	78	1	0.817	0.019556	0.778	0.855
48	53	2	0.786	0.028480	0.730	0.842
49	20	5	0.589	0.079037	0.435	0.744

Table 4: A life table representing the under-five mortality data in Ghana.

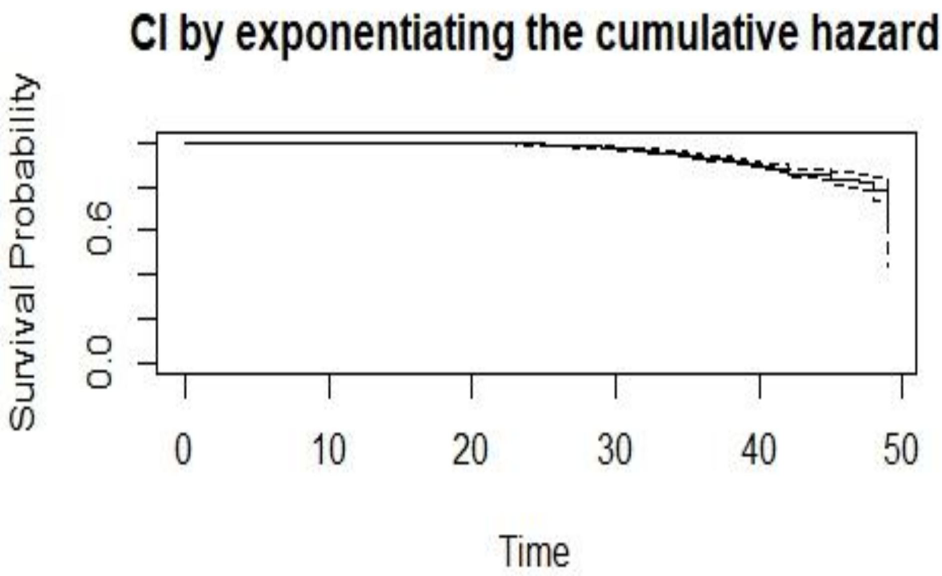

Figure 2: Kaplan Meier curve showing the survival status for children Under-five mortality.

Figure 3 through Figure 3.15 gives an overview of the survival curves for under-five mortality of children in Ghana based on the explanatory variables. The survival curve in Figure 3 shows the sex of children under-five mortality with censoring in general not based on any other factor aside from the sex of the child, where the censoring indicator is represented by the addition (plus) sign. From the curve, the survival probability for males is a little higher than that of females at the early ages, that is, from the 16th month to the 24th month, but then at later stages, females under five tend to have a higher survival probability than the males, thus from the 26th month to the 31st month. As they grow from the 33rd month onward, the female survival probabilities begin to drop while that of the males begins to rise. In some also, both genders tend to have the exact survival probabilities,for example, the 25th month and the 32nd month. This can be explained to be that those malechildren given the same age as females tend to live longer than male children, probably because male's under-five tend to have a more robust immune system than females under-five.

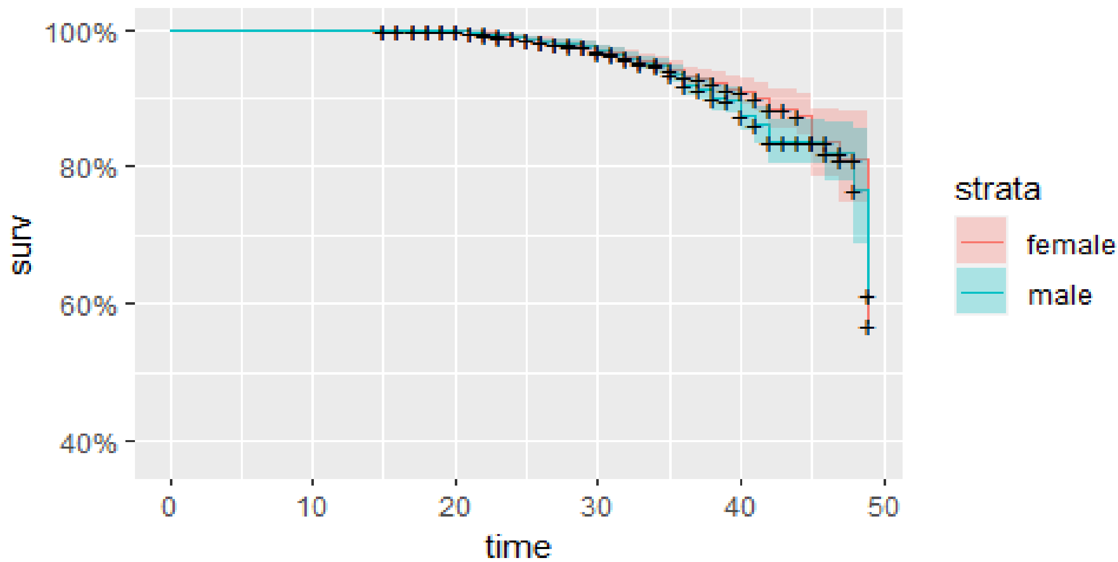

Figure 3: Survival curve for children Under-five in Ghana based on child's sex.

The survival curve (see Figure 4) shows the probabilities for those under five based on their year of death. The curve below gives us a fair idea of yearly survival for under-five children. From Figure 4, we can deduce that 2014 has a low survival probability compared to otheryears, and 2009 experienced the highest survival of children under five. As the years increase, survival probabilities also tend to decrease. Thus, the mortality of children underfive kept increasing. This tells us that Ghana has been experiencing high mortality rates forchildren under-five as the years keep growing.

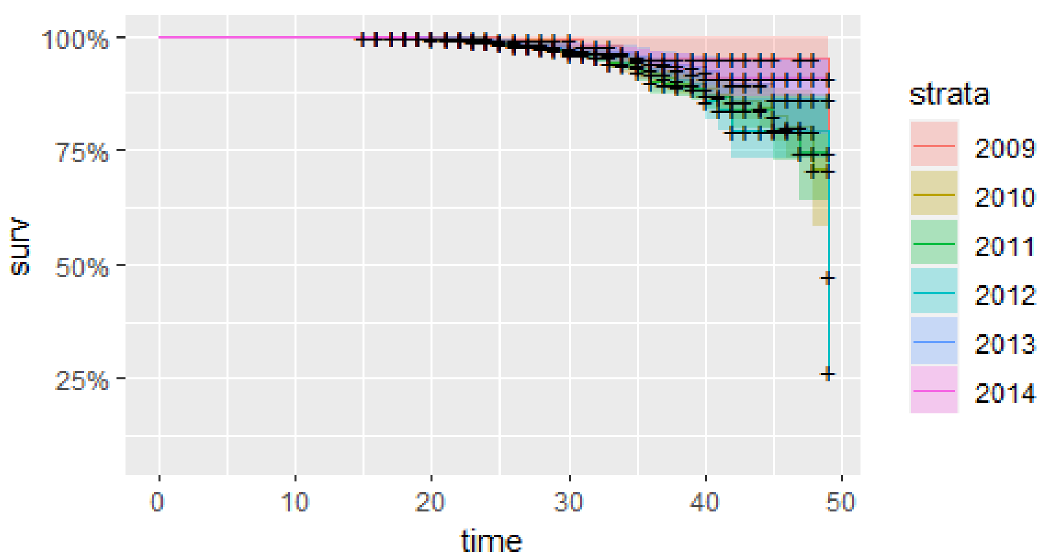

Figure 4: Survival curve for children under five in Ghana based on the year the kid was born.

Aside from the sex of the child considered in Figure 3, and the year of birth of the child inFigure 4, another factor considered to affect under-five child mortality is the current work of the mother, which is shown in Figure 5. It can be deduced from Figure 5 that mothers with current work had children who livedlonger; their child's survival probabilities were higher than those without recent work at thetime of birth. This means mothers who were not working needed more funds to look after their children properly. They could not provide the necessary medications and other things to aid the child's survival.

Moreover, the survival curve in Figure 6 also indicates that the mother's education level influences the child's survival probability. Given that the mother has no education, the child's survival probability is high, but as the child grows, the survival probability becomesunstable; that is, the survival probability starts to decrease as the child grows. This is because the mother has little knowledge of how to go about the child's welfare and the process she must go through for the child's health. Although the mother with no education has a downward sloping curve, we see that at earlier ages, the survival probability for children whose mothers have no educational background was higher than mothers with primary and secondary education but lower than mothers with higher education. As the child grows from the 31st month onwards, the survival probability for children whose mothers have no education, although low, was higher than for children whose mothers hadprimary or secondary education.

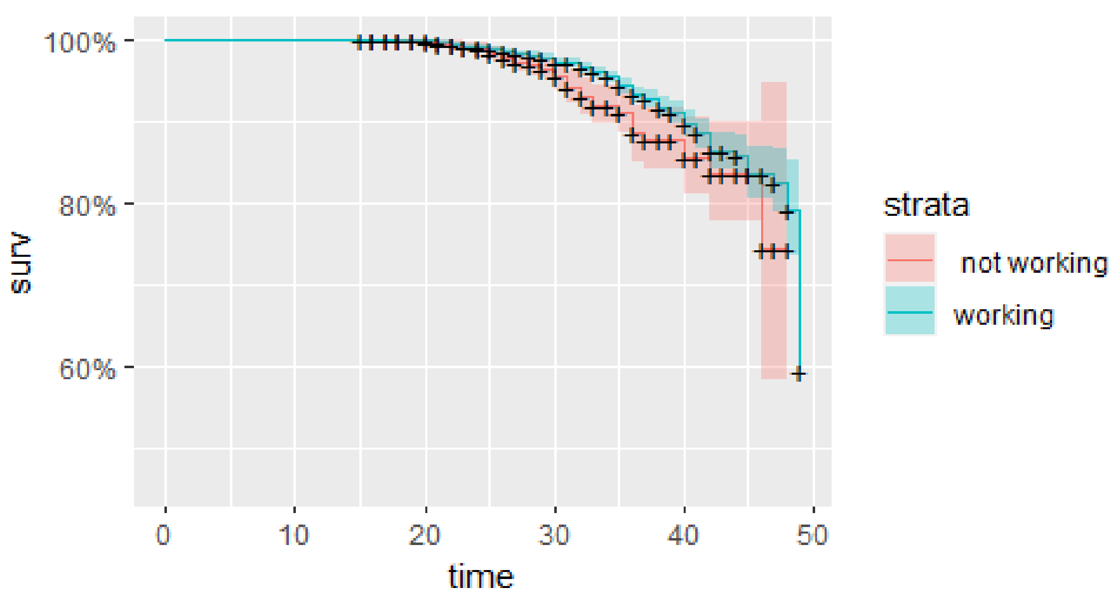

Figure 5: Survival curve for children under-five in Ghana by current work.

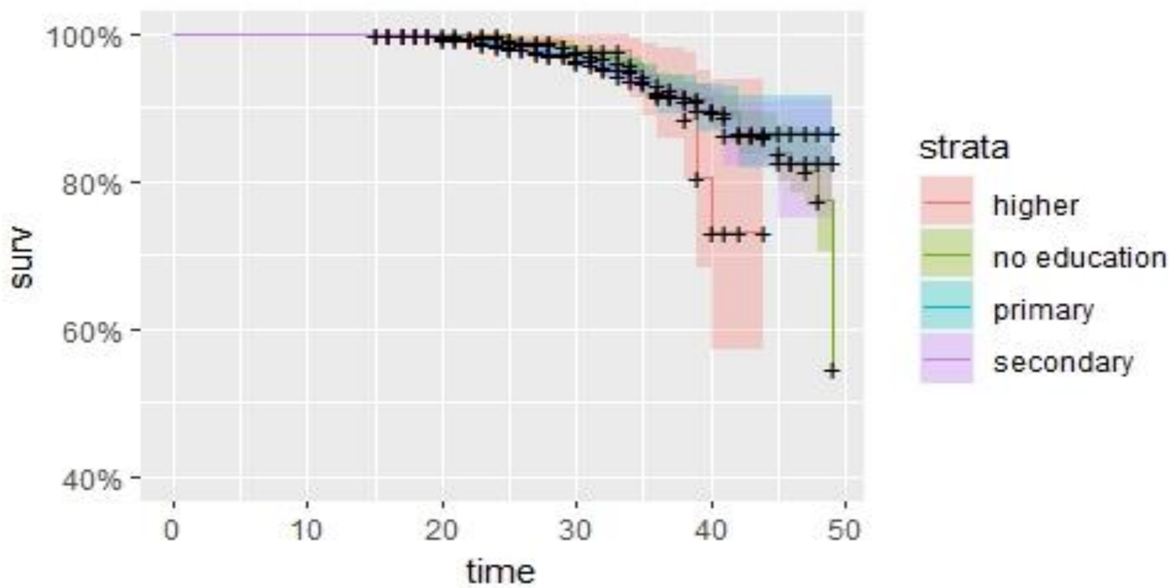

Figure 6: Survival curve for children under five in Ghana by education level.

According to the survival curve in Figure 7 below, when the mother never married, thechild experienced a low survival probability compared to other cases. Children born to mothers who never married did not have a long life due to many reasons, probably the mothermight be a teenager and, as such, would not be able to fend for herself and the child as welldue to poor financial condition, low standard of living, and other psychological effects suchas poor mental health. Children born to mothers or parents who are separated or not livingtogether tend to live a little longer than children born to mothers who are not married, sameas children born to parents living together, for such children at the initial stage have an increased survival to that of children born to parents not living together, at some point bothhave the same survival probability but from age 29months to 40 months they have a low chance. At later ages, the possibilities increase. Mothers who are married, widowed, or divorced tend to have children with high survival probabilities, but those born widowed mothers tend to live longer than in all the other cases.

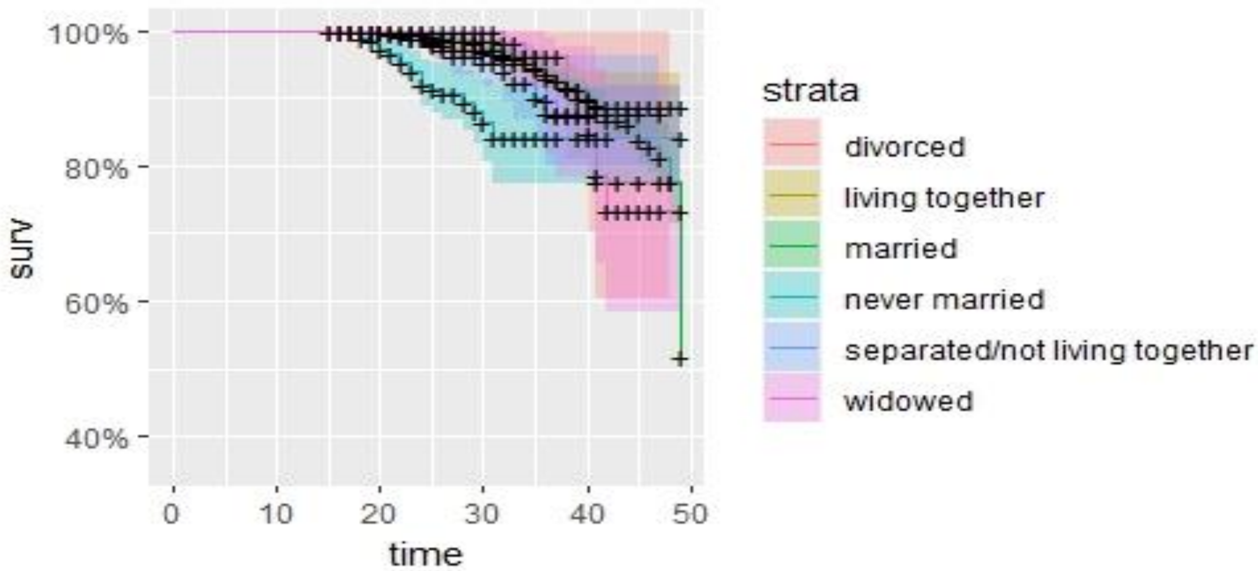

Figure 7: Survival curve for children under five in Ghana by marital status.

The survival curve in Figure 8 appears to have a low survival probability for a childborn to mothers who are visitors to the country. This may be because the mothers whoare visitors in the country may not fully be vested with the facilities in the country and, as such, may not be acquainted with how the system works and probably the weather condition will also be a factor since visitors might not be used to such weatherand climate change hence leading to their children not being able to survive longer, unlike children who are born to mothers who are usual residents of the country. Theymight have higher survival because they are used to the country's system and know how best to care for the child under conditions presented to them by how the systemworks.

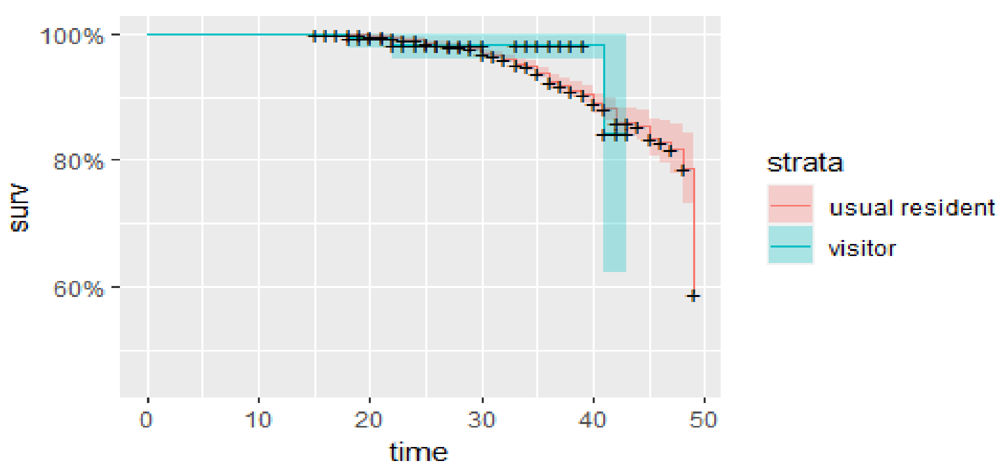

Figure 8: Survival curve for children under five in Ghana based on residence.

Children born to mothers whose place of residence is in the urban sector tend to live longerand thus have higher survival probabilities than children whose mothers live in the rural sector, as shown in Figure 9, at earlier stages but were lower as the child keeps growing.This is because, in urban places, good medical facilities exist. Adequate health facilities areavailable, improving the child's survival, which explains the high survival probability at theearly stages. However, due to the demand of the hustle and bustle in urban places, proper care and attention is not given to the children as they grow. For mothers who live in rural areas, such facilities, and amenities do not exist, leading to a decreasing and unstable survival for the child, justifying the low survival curves. Mothers who live in rural places may resort to traditional treatment and herbs, thus explaining the increase in survival probability at the early stages.

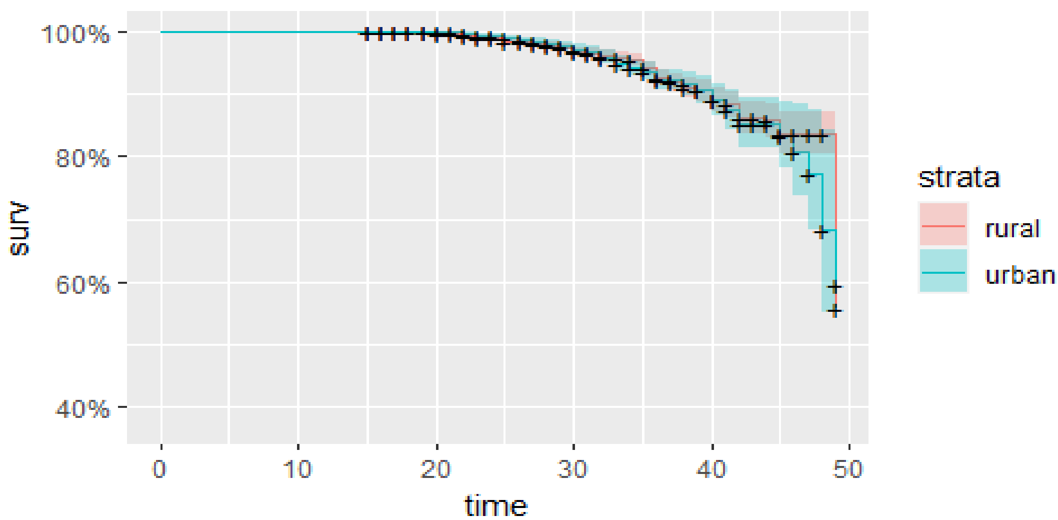

Figure 9: Survival curve for children under five in Ghana based on mother's urban-rural status.

The wealth status of the mother is another factor that determines the survival and longevityof their child. As shown in Figure 10, mothers considered highly rich tend to have children with higher survival probabilities in their early stages since, during these periods, the mothers have time to fully take good care of and look after the children themselves. Due to their status of being highly rich, they leave the children to caretakers and nannies to look after them during their 35th month onwards to go to their various jobs. In such instances, the adequate proportion of food given to the child might not be enough, thus making them malnourished and, in the long run, leading to low survival probabilities,as indicated by the curve. From the curve, averagely rich mothers (i.e., they're in the middleclass), those that are poorer, and those considered very poor (i.e., poorest) roughly have thesame survival probabilities for their children.

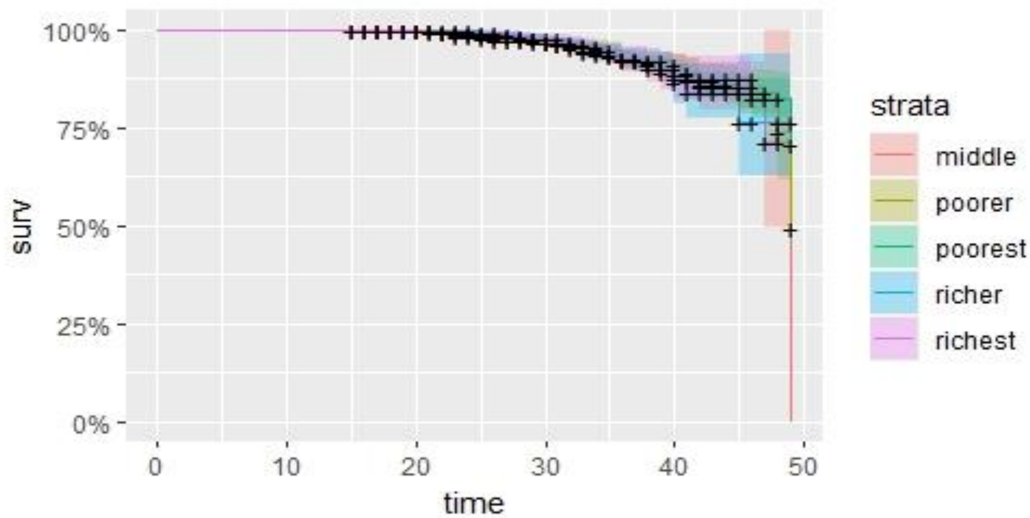

Figure 10: Survival curve for children under-five in Ghana by wealth status.

The Cox-regression Model

The results from the cox-regression model show the covariates in the model that were statistically significant with time. It also shows the direction of the significance by the coefficient values given; see Table 5 and Table 6 for a display of results. Considering all the covariates and running the Cox regression analysis on them, we see that some of the covariates were significant while others were not. After running the model, the results showed that sex was not statistically significant since it had a p-value greater than 5% and a positive direction, indicating that survival based on sex is low. This implies that the hazard risk for children under the age of five is high when sex is considered. Although resident-based status has a negative direction, indicating a high and good survival chance for children under-five, it is statistically insignificant because the p-value is greater than 0.05. The wealth status had a negative coefficient value for all its categories (i.e., poorest, poorer, middle, richer, and richest), with their p-values all more significant than the one employed in the study, indicating that the covariate wealth status is not statistically significant with age, even though it showed a negative direction indicating a high survival rate for under-five children. It was excluded from the analysis because its p-value was greater than 0.05. The results show that urban-rural status suggests a high survival rate for children under-five, as shown by the negative coefficient given in Table 5. However, the p-value indicates that it is not statistically significant.

The education level had positive coefficient values for all its categories, indicating that education level when measured against age, predicts a high survival rate for children under-five. However, its p-value was greater than 5%, suggesting that it is not statistically significant. The child's birth year was also not statistically significant, indicating that the child's survival rate decreased as the years passed. As shown in Table 5 by the various p-values (Pr(>|z|), marital status and currentemployment status covariates were statistically significant for age. The marital status was significant, with a negative direction indicating a high survival for under-five children bornto parents who are living together and those who are married, and a positive direction indicating high mortality and thus low survival for those born to parents who were never married, separated/ not living together and widowed. However, the order in the table meansthat under-five children born to such parents have a low survival rate that increases over time, i.e., it is very low at first. However, despite being low, the coefficient value for widowed gives a smaller positive value, indicating that it is better than the value before it but still shows a low survival rate. On the other hand, the current employment status exhibited statistical significance in a negative direction, indicating it correlates to high survival for children under the age of five. The hypothesis that the hazard in the covariates associated with age over time becomes proportional to each other and, thus, the hazard becomes the same over the follow-up time can then be said to satisfy the Cox regression assumption. The covariates' marital status and current employment status satisfy this hypothesis. The various p-values obtained from the likelihood ratio, Wald, and score (log-rank) tests in Table 7 provide the impression that all the covariates satisfy the assumption and are statistically significant with time.

This is a strong indicator that adding non-statistically significant covariates to the Cox regression analysis will result in misrepresentation and, as a result, incorrect estimates. To avoid the work being congested, the Cox regression was conducted for individual covariatesto evaluate which covariates are significant with time and which ones are not. Results for the respective analysis may be found in the appendix section. The analysis and results derived from running the Cox regression model on the individual covariates and by basing the interpretation on the various tests (i.e., the Wald test, the likelihood ratio test, and the score (log-rank) tests, the study employs the Wald test as the primary statistical test used for the analysis), the various covariates with the p-value of thetests mentioned earlier showed that the wealth status, the education level, the resident status,the rural-urban status, the child's year of birth, and sex were not statistically significant with age, meaning they generated a p-value greater than the one given, which was 5%, however marital status and the current employment status yielded a p-value less than 5% as shown in the appendix. As a result, the Cox regression assumption is satisfied for maritaland current employment status, and their hazard ratios are proportionate to each other across time. Therefore, we say that the covariates (i.e., urban, residence, education level, wealth status, sex of the child, and the year of birth) strongly influence the hazard risk of under-five childmortality over time. The covariates (i.e., current work and marital status) do not highly affect the under-five child mortality hazard risk. Similarly, since these covariates are statistically significant with age, marital status is linkedto a high but later lower survival rate, based on the order of the various categories named in the table, which is explained by the coefficient value and current employment is linked to lower under-five mortality, in other words, a high survival rate for children under the age of five. As a result, marital status and current employment status had the highest impact onunder-five child mortality in Ghana.

	Coef	exp(coef)	se(coef)	z	Pr(>|z|)
Male	0.14851	1.16010	0.11879	1.250	0.21125
Working	-0.41207	0.66228	0.15101	-2.729	0.00636
2010	0.68147	1.97678	0.46468	1.467	0.14250
2011	0.79897	2.22325	0.46369	1.723	0.08488
2012	0.83964	2.31554	0.46507	1.805	0.07101
2013	0.45230	1.57193	0.47589	0.950	0.34189
2014	0.36363	1.43854	0.48380	0.752	0.45228
urban	-0.03685	0.96382	0.15943	-0.231	0.81719
Living together	-0.08837	0.91542	0.47557	-0.186	0.85259
married	-0.09200	0.91211	0.45565	-0.202	0.83999
never married	1.40844	4.08956	0.49404	2.851	0.00436
separated/not living together	0.25057	1.28475	0.55061	0.455	0.64906
Widowed	0.16419	1.17844	0.56118	0.293	0.76985
visitor	-0.35793	0.69912	0.58292	-0.614	0.53919
no education	-0.26671	0.76589	0.35104	-0.760	0.44739
primary	-0.17583	0.83876	0.35335	-0.498	0.61875
secondary	-0.07647	0.92638	0.32231	-0.237	0.81245
poorer	-0.06775	0.93449	0.16517	-0.410	0.68166

middle	-0.05692	0.94467	0.20300	-0.280	0.77917
richer	-0.07431	0.92838	0.24197	-0.307	0.75875
richest	-0.27856	0.75687	0.28757	-0.969	0.33270

Table 5. Cox proportional hazards regression analysis for under-five childmortality in Ghana.

	exp(coef)	exp(-coef)	lower 95%	Upper 95%
male	1.1601	0.8620	0.9191	1.4642
working	0.6623	1.5099	0.4926	0.8904
2010	1.9768	0.505	0.7951	4.9147
2011	2.2232	0.4498	0.8960	5.5168
2012	2.3155	0.4319	0.9306	5.7613
2013	1.5719	0.6362	0.6185	3.9949
2014	1.4385	0.6951	0.5573	3.7131
urban	0.9638	1.0375	0.7052	1.3173
living together	0.9154	1.0924	0.3604	2.3250
married	0.9121	1.0964	0.3734	2.2279
never married	4.0896	0.2445	1.5529	10.7696
separated/not living together	1.2848	0.7784	0.4367	3.7801
Widowed	1.178	0.8486	0.3923	3.5399
visitor	0.6991	1.4304	0.2230	2.1915
no education	0.765	1.3057	0.3849	1.5240
primary	0.8388	1.1922	0.4196	1.6765
secondary		1.0795	0.4925	1.7424
	0.9264			
poorer	0.9345	1.0701	0.6761	1.2917
middle	0.9447	1.0586	0.6346	1.4063
richer	0.9284	1.0771	0.5778	1.4917
richest	0.7569	1.3212	0.4308	1.3298

Table 6 Cox proportional hazards regression analysis for under-five childmortality in Ghana

Table 7: an overview of the various test of significance from the Coxproportional hazards regression analysis.

Concordance = 0.629 (se = 0.02)

Likelihood ratio test= 61.07 on 21 df, p=9e-06 Wald test = 78.48 on 21 df, p=1e-08

Score (log rank) test = 90.93 on 21 df, p=1e-10

Summary

This study examined the factors that determine under-five mortality in Ghana. We reviewedother works on the topic and concluded those that aided our research. Ghana, like other countries, seeks to attain reduced child mortality and improved child health as per the MDGs and SDGs by the United Nations, but this goal has not been achieved; thus, Ghana is nowhere close to achieving this goal, and such child mortality is a cause for worry. Other works showed factors that influence under-five mortality in Ghana, such as household income, twin birth, birth spacing, etc., depending on the covariates and the model chosen by the author.

The study was done using the 2014 DHS data obtained from IPUMS Global Health. The Cox regression model was applied to the data, and based on the covariates chosen, which were the child's sex, marital status, urban-rural status, residence status, current employment status, wealth status, and education level, certain deductions were made. The Cox regression model was used because it is a model which is commonlyused to analyze survival data. It is the model chosen for this paper because it is simple, easyto use, and very efficient in implementation. The model measures the relationship betweenan independent variable and survival time with censored data. The Cox regression is an excellent model because it provides a reasonable estimate for β, and the hazard ratio can alsobe used to determine the association of the risk factors between two individuals. From the model, we concluded that marital status and the current employment status are associated with decreasing the mortality for under-five, and much focus should be put on maternal education and female education to help reduce under-five mortality. We, therefore, say that the Cox regression model is used in the business and economics spheres as well as it measures the time to the event and that is a helpful model that can help predict a particular topic under study, mainly when using survival analysis.

Conclusion

Conclusions were drawn from the study, and these conclusions are as follows; The chosencovariates all readily influenced the under-five mortality rate. The study concludes that some covariates were statistically significant with age, while othersdid not consider each covariate individually and that each influenced under-five mortality one way or the other. We conclude that the geographical location of the child (i.e., the urban-rural status), the wealth status, the residence, the education level of the mother, the year of birth of the child, and the sex of the child were not statistically significant with ageand hence are associated with increasing under-five mortality. Female under-five childrentend to have a higher mortality rate than male children. Male under-fives had a lower hazardrisk of death as compared to female under-fives. Current working and marital status covariates are also associated with decreasing under-five mortality rates. Some covariates were not included in the study based on the Cox regression assumption, thereby making the covariates influence under-five mortality two: marital status and the current working status.

Recommendation

Identifying potential risk factors underlying under-five is very important and of much concern when policies and interventions are to be put in place to help reduce under-five mortality. Some of these interventions and actions that can be implemented are establishing excellent and robust health systems and maternal education; that is, mothers of under-fives should be well educated on caring for the child, birth spacing, and family planning. Also, female education should be of topmost priority in our various societies. Parents of under-fives should be encouraged to live together so the child receives love and affection from both the mother and father. To be able to reduce under-five child mortality to a particular as mentioned in the MDGs and SDGs of the UN, proper implementation and effective administration of these policies should be enforced, and good government policies must also be implemented for the poor and marginalized in the society to help them also contribute to the reduction of under-fivemortality in the country.

REFERENCES

Alotaibi, R. M., Rezk, H. R., & Guure, C. (2020). Bayesian frailty modeling of correlated survival data with application to under-five mortality. BMC Public Health, 20, 1-24.

Asare, H., Carboo, J., Nel, E. D., Dolman, R. C., Conradie, C., Lombard, M. J., & Ricci, C. (2021). Mortality concerning profiles of clinical features in Ghanaian severely undernourished children aged 0–59 months: an observational study. British Journal of Nutrition, 125(10), 1157-1165.

Ayele, D. G., Zewotir, T. T., & Mwambi, H. (2017). Survival analysis of under-five mortality using Cox and frailty models in Ethiopia. Journal of Health, Population, and Nutrition, 36, 1-9.

Musa, M. C., Asiribo, O. E., Dikko, H. G., Usman, M., & Sani, S. S. (2020). Modeling the Determinants of Under-five Child Mortality Rates Using Cox Proportional Hazards Regression Model. FUDMA JOURNAL OF SCIENCES, 4(4), 401-408.

Nasejje, J. B., & Mwambi, H. (2017). Application of random survival forests in understanding the determinants of under-five child mortality in Uganda in the presence of covariates that satisfy the proportional and non-proportional hazards assumption. BMC research notes, 10(1), 1-18.

Ogbo, F. A., Ezeh, O. K., Awosemo, A. O., Ifegwu, I. K., Tan, L., Jessa, E., ... & Agho, K. E. (2019). Determinants of trends in neonatal, postneonatal, infant, child, and under-five mortalities in Tanzania from 2004 to 2016. BMC public health, 19, 1-12.

Al-Thalabi, S. H. (2021). Survival Analysis by using the Cox regression model for cancer patients in Basra (Model of the educational hospital). Journal of Statistics and Management Systems, 24(4), 779-789. doi:10.1080/09720510.2020.1859805.

Dwomoh, D., Amuasi, S., Agyabeng, K., Incoom, G., Alhassan, Y., & Yawson, A. E. (2019). Understanding the determinants of infants and under-five mortality rates: a multivariate decomposition analysis of demographic and health surveys in Ghana, 2003,2008, and 2014. (S. Yaya, Ed.) BMJ Global Health, 4:e001658. doi:10.1136/bmjgh-2019-001658.

Ngwa, J. S., Cabral, H. J., Cheng, D. M., Pencina, M. J., Gagnon, D. R., Lavalley, M. P., & Cupples, L. A. (2016). A comparison of time-dependent Cox regression pooled logistic regression and cross-sectional pooling with simulations and an application to the Framingham Heart Study. BMC Medical Research Methodology,16(148). doi:10.1186/s12874-016-0248-6.

Nurmalasari, M., Ifada, L., & Pramana, S. (2019). Survival analysis of under-five mortality in west Sulawesi Indonesia using Cox regression. International Conference on Health (ICOH 20190, (pp. 92-96). doi:10.5220/0009566900920096.

Ofori-Amanfo, E. K. (2019). Forecasting mortality rates and modeling longevity risk for SSNIT pensioners. Retrieved from the University of Ghana http://ugspace.ug.edu.gh.

Ogbo, F. A., Ezeh, O. K., Awosemo, A. O., Ifegwu, I. K., Tan, L., Jessa, E., . . . Agho, K. E. (2019). Determinants of trends in neonatal, postneonatal, infant, child, and under-five mortalities in Tanzania from 2004 to 2016. BMC Public Health, 19(1243). Retrieved from http://doi.org/10.1186/ s12889-019-7547-x

Stel, V. S., Dekker, F. W., Tripepi, G., Zoccalli, G., & Jager, K. J. (2011). Survival Analysis II: Cox Regression. Nephron Clinical Practice, 119, 255-260. doi:10.1159/000328916

Tolosie, k., & Sharma, M. K. (2014). Application of Cox proportional hazards model in case of Tuberculosis Patients in selected Addis Ababa Health Centres. Retrieved September 2021, from Hindawi Limited:http://www.hindawi.com/journals/trt/2014/536976.

Azaare, J., Akweongo, P., Aryeteey, G. C., & Dwomoh, D. (2022). Evaluating the impact of maternal health care policy on stillbirth and perinatal mortality in Ghana; a mixed method approach using two rounds of Ghana demographic and health survey data sets and qualitative design technique. PloS one, 17(9), e0274573.

Buor, D. (2003). Mothers' education and childhood mortality in Ghana. Health policy, 64(3), 297-309.

Musa, M. C., Asiribo, O. E., Dikko, H. G., Usman, M., & Sani, S. S. (2020). Modelling the Determinants of Under-five Child Mortality Rates Using Cox Proportional Hazards Regression Model. FUDMA JOURNAL OF SCIENCES, 4(4), 401-408.