Ayoo Edwin Ochola
Department of Statistics and
Actuarial Science, Jomo
Kenyatta University of
Agriculture and Technology,
Nairobi, Kenya

Charity Wamwea
Department of Statistics and
Actuarial Science, Jomo
Kenyatta University of
Agriculture and Technology,
Nairobi, Kenya

Herbert Orang'o Imboga
Department of Statistics and
Actuarial Science, Jomo
Kenyatta University of
Agriculture and Technology,
Nairobi, Kenya

Joel Chelule
Department of Statistics and
Actuarial Science, Jomo
Kenyatta University of
Agriculture and Technology,
Nairobi, Kenya

An analysis of child mortality using survival regression models

Ayoo Edwin Ochola, Charity Wamwea, Herbert Orang'o Imboga and Joel Chelule

Abstract
World over Kenya and Uasin Gishu in particular, child mortality remains a challenge as children are expected to live up to adulthood. However, they often fail due to variety of diseases. This work considered child mortality as default and time default to be five years. The study analyses child mortality using survival regression models. Secondary data was obtained from Moi Teaching and Referral Hospital. Diseases (Risk Factors) that influenced under-five child mortality were considered as variables of the study. The study began by testing proportional hazard assumption on the data collected relating to child mortality data where it was found that proportional assumption is not violated. Cox Proportional Hazard model (CPHM) was fitted to determine the effects of risk factors on child mortality where factors such as gender, malformation, and dehydration increased child mortality risk while cancer decreased this risk. Factors such as tuberculosis, pneumonia and digestive system depicted an insignificant effect on child mortality when evaluated at 5% significance level. The overall goodness of the model was checked using concordance index and Wald test. Child mortality is a very important aspect of measuring health status of the current population and predicting the health of the future generation. The research project recommends modelling of child mortality using other survival models. The study is significant in higher institution of learning since it broadens the knowledge on the application of survival analysis in modelling child mortality. It provides the background for researchers in high institution of learning who are interested in doing studies on child mortality modelling. The study is also important to the health sector since by knowing the factors making an under-five child to be more prevalent to death will help the government achieve the MDG 4 goals faster by improving on these factors.

Keywords: Wald test, child mortality, cox proportional hazard, survival analysis, concordance index

Introduction
Background of the Study
Child mortality is defined as the risk that a child might not get to their 5th birthday. Globally, child mortality is a core indicator for child health and well-being. Both developed and developing nations considers child health as a central issue in the public agenda. There have been implementation of several policies over the years geared towards improving child health with different degrees of success. The Millennium Development Goal 4 (MDG 4) suggest that the rate of child mortality was reduced by two thirds in the years between 1990 and 2015. The SDGs proposed target for child mortality represent commitment renewed to the world's children to end child mortality by the year 2030. Hence, all countries aim to reduce neonatal deaths to as low as 12 deaths per 1000 live births and under-five deaths be reduced to as low as 25 deaths per 1000 live births. However, the target risk has not been achieved globally. The child mortality report as per the UN Inter-agency group of 2013 estimated that the child mortality rate has reduced globally by 47% between the year 1990 and 2012, from 90 per 1000 live births to 48 per 1000 live births respectively.

Despite the global improvement, deaths of children less than five years are highly concentrated in sub-Saharan African countries, where the rate of child mortality is 98 deaths per 1000 live births which is 15 times more than the average for the developed region, while for other nation that are less developed the rate of child mortality is 71/1000 live births (Moise *et al*., 2017) [9].

Corresponding Author:
Ayoo Edwin Ochola
Department of Statistics and
Actuarial Science, Jomo
Kenyatta University of
Agriculture and Technology,
Nairobi, Kenya

The findings by Frisbie *et al*. (2004) [4] show that over the years 2000s, the rate of child mortality in the sub-Saharan region remained the highest globally at 94 deaths per 1000 live births as compared to the rest of African countries at 88 deaths per 1000 live births and less developed countries at 61/1000. Omariba and Boyle (2007) [13] found that both social-economic and social-cultural factors have significant influence on the increasing rates of mortality in the sub-Saharan Africa. The (2008/2009) Kenya Demographic and Health Survey shows that the 74/1000 death rate is way above the national target of 32/1000 deaths. The 2012 achieved mortality rate of 48/1000 is also higher than the national average. Since the developing countries are still at risk, it is prudent to analyse the trend and level of infant mortality in these countries.

In addition, social cultural factors such as the vast ethnic, religious, beliefs and cultural practices that exist in the sub-Saharan Africa affect whether or not individual seek bio-medical forms of health care when their child is in ill health and during pregnancy (Omariba and Boyle, 2007) [13]. According to many cultures, the social status of women is increased by her ability to bear children and this also increases her marital chance. The sub-Saharan Africa women have high fertility hence have short birth intervals between pregnancies which compromised health outcome for both the child and the mother. Due to unemployment in the sub Saharan Africa region, parents are unable to seek medical assistance whenever their child is ill because of lack of finances. Therefore, the socio-economic and socio-cultural factors are important in understanding the high child mortality rate in the sub-Saharan Africa as compared to other regions globally. In Kenya, the under-five child mortality is becoming key issue and has shown an increasing trend in the country over the years. Kenya faces many challenges with regard to decreasing child mortality within its borders. These may include economic insecurity, poor health services and the worsening effects of the HIV/AIDS pandemic within its borders. Over the past years, Kenya's economy has suffered a major blow. The economic hazards experienced here can be traced to various factors which majorly impacts on child mortality as it affects every sector of the country, employment and the health sectors, which directly are associated with child health outcomes. Uasin Gishu being one of the county in Kenya is not an exemption hence faces challenges of child mortality.

Statement of the problem
Under-five mortality remains a challenge as children are expected to live up to adulthood. However, they often fail due to various factors varying from one place to another depending on the accessibility of the health facilities and how well these health facilities are. The availability of health care services has been reported to have an inverse relationship with child mortality (Kayode *et al*., 2016) [6]. Several studies' (Ezeh *et al*., 2015; Omolo, 2014; Murithii & Murithii, 2015) [16, 14, 10] on causes of child mortality in Kenya show the effects of environmental conditions and social characteristics. Researchers who have done research on child mortality using survival regression models have only used either of the models. Therefore, the interest is to test proportional hazard assumption on the child mortality data and assess the effect the risk factors have on child mortality using secondary data of all births and deaths of children for a five-year period (2015-2019) obtained from MTRH. This will help the government achieve the MDG 4 goals faster by improving on these factors to reduce child mortality. Time defaulting occurs within the five-year age period non-uniformly, meaning it varies from one child to the other. Age when death occurs is the event of interest while the analysis is based on the time from birth to age at death (of children below five years).

Under-five mortality is important to note in order to know the significant risk factors of U5M. The Research Project applied a statistical technique in the analysis of child mortality data using CPHM. Specifically, the study examined the effects of several explanatory variables (risk factors) on the survival of children. The findings of this study is significant in high institution of learning since it broadens the knowledge on the application of survival analysis in modelling child's mortality and also provide the background for researchers in high institution of learning who are interested in doing studies on child mortality modelling.

Objectives of the study
General Objective
To analyse child mortality using survival regression models.

Specific Objectives
- To test for the proportional hazard assumption on the data.
- To fit the appropriate survival regression model on child mortality (AFT and CPHM).
- To test the effect the risk factors have on child mortality.
- To test for the overall goodness of fit of the fitted model.

Chapter Two: Literature Review
Introduction
This chapter presents review of studies done in relation to child mortality. A relation of the topic is analysed to show what has been done. Survival analysis focuses on the distribution of survival times. There are various models used to estimate survival distributions, however, most survival analysis literature examine the relationship between survival rate and one or more covariates.

Review on Survival Analysis Models
The Cox proportional Hazard regression model is broadly applicable and the most used method of modelling survival analysis. The model simultaneously explores the effects of several variables on survival (Mani *et al*., 2012) [8]. The proportional hazard assumption refers to the hazard function to be multiplicatively related, that is, its ratio is assumed constant over survival time, thereby not allowing a temporal bias to become an influential player at the endpoint. In Cox's method the dependent variables are the risk function at any given time (Bradburn *et al*., 2003) [2].

Earlier study in child mortality conducted by Kayode *et al*. (2013) [17] on the risk factor for child mortality came up with predictive model (multivariable logistics regression model) that associated the explanatory variables (risk factors) and dependent

variable mortality in the study. The risk factors were maternal including current age, education, occupation parity, marital status, age of first marriage, family planning, preceding birth interval, breastfeeding and health seeking behaviour. Also, childhood factors were considered including sex, birth order and birth weight. Additionally, household factors were considered including family size, sanitation, number of wives, wealth index, fuel and water sources, and paternal factors such as age and occupation, and other factors including place of residence, ethnicity and geopolitical region. They then used invariable logistic regression to examine the association between the explanatory variables and the dependent outcome. Their study revealed that maternal, childhood, family and other factors were important risk to deaths among children under the age of five years in Nigeria.

Kembo and Van Ginneken (2009) [7] conducted a study to investigate the association between socio-economic, sanitation and maternal variables, and child mortality in Zimbabwe using multivariable proportional hazards regression model. Maternal and child health services were found to improve child survival rates. Hailemariam and Tesfaye (1997) [5] applied CPHM to examine factors impinging the survival of child's and children 1-3 years of age. The findings show that there is higher risk of child mortality for births with short previous birth intervals than for births that are of higher order (more than 5), for young women births with under 20 years of age, and for those women who are older with more than 35 years of age.

O'Leary et al. (2013) [12] carried out a study on maternal alcohol use, sudden child death syndrome and child mortality excluding SIDS. Analyses were conducted using CPHM and the results show that there is higher risk of SIDS when maternal alcohol diagnosis is recorded during pregnancy and the Research concluded that maternal alcohol-use disorder is a significant risk factor for SIDS and child mortality excluding SIDS.

Kayode et al. (2013) [17] conducted a research in modelling long-term graft survival with-varying covariates effects: An application to a single kidney transplant centre in Johannesburg, South Africa and the conclusion made about predictors of graft survival is that AFT model over more flexibility in understanding covariates with non-constant effects on graft survival.

Saroj et al. (2019) [15] examined factors affecting under-five child mortality using survival parametric models and the results show that various factors influenced child mortality. Ayele et al. (2019) [18] used Cox and frailty survival analysis model to determine mortality of children under the age of five years in Ethiopia. The findings show that children under the age of five years have less chances of dying if before the child attain five years of age the mother does not become pregnant for another child. Naz and Patel (2020) [11] concluded that birth spacing of 3 years and above is associated with reduced risk of child mortality contrasted with short birth interval.

Research Gap
Previous scholars have done Research on child mortality using survival regression models but none has specifically analyzed the child mortality data from MTRH in Uasin Gishu and tested it to ascertain if at all the proportional hazard assumption holds on the data.

Chapter Three: Research Methodology
Study Variables
The research analyzes child mortality which refers to mortality of children from birth to the age of 5 years. In this case, under-five mortality is the dependent variable of the study and is defined as the time to death of a child before the age of five years. Diseases and conditions (risk factors) that cause most of the deaths among children before the age of five years in Uasin Gishu were used in this study as independent variable (covariates). Hence, the aim of this project was to determine the effects of these covariates on child mortality rates. The dependent variable is a function of a series of the mentioned covariates.

Target Population
The study was conducted at Moi Teaching and Referral Hospital (MTRH) in Eldoret town in Uasin Gishu. MTRH falls under highly densely area in Eldoret hence receives more patients; hence wise to take it as a case study. The research project used secondary data that targeted population of all births and deaths of children from the year 2015 to 2019 in Moi teaching and referral hospital. The unit of analysis of the Research Proposal were the risk factors (diseases) that mostly influence child mortality in Uasin Gishu.

Sample Size
The study assumed that the sample of the under-five child who die will be half of the sample taken from the whole population of children, if they do not exist then we assume

P = 0:5 then

$$N = \frac{\left(z_{\frac{\alpha}{2}}\right)^2 pq}{d^2} \tag{1}$$

$$= \frac{(1.96)^2 * 0.5 * 0.5}{0.015^2} = 384.16 = 384 \ children$$

Since the total population for the under-five children were less than 10,000 for the Hospital records, the finite correction factor was employed to determine the correct sample size that was used in this research. When the sample represents a significant proportion, then the finite population correction factor can be applied. This reduced the sample size as required. The formula for this is:

$$n = \frac{n_\alpha N}{n_\alpha + (N-1)} \tag{2}$$

The list of a sample frame that the Research Project considered in collected secondary data from MTRH of all birth and death of children for a five-year period (2015-2019) are; Name (serial number), gender, age at death, and the risk factor/disease that accounted for the death.

Data Collection Procedures
The researcher used secondary data of all births and deaths of all the children under-five years of age from the year 2015 to 2019 obtained from MTRH hospital.

Data Processing and Analysis
Data processing and analysis was conducted using R statistical package where the collected secondary data in relation to all births and deaths of under-five-year child mortality from the year 2015-2019 were coded to assume numerical values to perform CPHM. This computer aided software for research assisted the researcher to carry out the analysis. This applied descriptive statistics with the mean, standard deviation and frequency tally. Analysis carried out by performing regression analysis using CPHM. Several curves are displayed for analysis purpose. Each variable is used for comparison on how they affect child mortality in Uasin Gishu.

Survival Regression Models
In this study, CPH or AFT model was fitted depending on the characteristics of the data. The CPH model is fitted if the assumption of proportional hazard is achieved, however, the AFT model does not take such assumptions into consideration.

Accelerated Failure Time Model

$$S(t/x) = S_0(\exp(\beta'x)t) \; for \; t \geq 0 \tag{3}$$

Where $S(t/x)$ refers to the survival function at time t, $S_0(\exp(\beta'x)t)$ is the survival function baseline at the time t and the factor $\exp(\beta'x)$, is the acceleration factor.
If $\exp(\beta'x) > 1$, the effects of covariates are decelerated.
If $\exp(\beta'x) < 1$, the effects of covariates are accelerated.
If $\exp(\beta'x) = 0$, the effects of covariates are constant.

The hazard function of AFT model can be expressed by;

$$\lambda(t/x) = exp(\beta'x \; \lambda_0(exp(\beta'x) \; for \; all \; t \tag{4}$$

Log-normal AFT Model
If εi has a standard normal distribution, then *Ti* has log-normal distribution. The density function of normal distribution is;

$$f\varepsilon_i(\varepsilon) = \frac{1}{\surd(2\pi)} exp - \left(\frac{logt - \mu - \beta_1 x_1 \ldots \ldots \beta_p x_{2p}}{\sigma}\right)/2 \tag{5}$$

The survival function of normal distribution is:

$$S\varepsilon_i(\varepsilon) = 1 - \emptyset(\varepsilon) \tag{6}$$

The distribution function of normal distribution is:

$$\emptyset(\varepsilon) = \frac{(log - \mu - \beta_1 \beta_1 \ldots \ldots \beta_p \beta_p)}{\sigma} \tag{7}$$

The cumulative hazard function is:

$$H\varepsilon_i(\varepsilon) = -\log\left(1 - \emptyset(\varepsilon)\right) \tag{8}$$

And the hazard function is:

$$H\varepsilon_i(\varepsilon) = \frac{f\varepsilon_i(\varepsilon)}{S\varepsilon_i(\varepsilon)} \tag{9}$$

In this way the log-normal AFT form can be derived.

Cox Proportional Hazard Regression Model
The cox proportional hazard model is given by:

$$h(t/x) = h_0(t)exp(\beta_1 x_1 + \beta_2 x_2 + \cdots \beta_p x_p) = h_0(t)\exp(\beta' x) \tag{10}$$

Where $h_0(t)$ is the baseline hazard function, $x = x_1, x_2 \ldots \ldots \ldots x_p$ is the vector of explanatory variables that represents a particular individual and $\beta = \beta_1, \beta_2 \ldots \ldots \ldots \beta_p$ is a vector for regression coefficients. Hence, the resulting survival function is expressed as follows.

$$S(t/x) = S_0(t)exp\left(\sum_{i=1}^{p} \beta_i x_i\right) \tag{11}$$

The model is semi-parametric and this vagueness creates no problem for estimation. Even though the baseline is not specified, good estimates for regression coefficients β hazard ratio and adjusted hazard curves can be estimated. The hazard ratio for two individuals with different covariates x and x^* will be given by;

$$\text{HR} = \frac{h_0(t)\exp(\beta x)}{h_0(t)\exp(\beta x^*)} = \exp\{\sum \beta'(x - x^*)\} \tag{12}$$

This hazard ratio is time-independent, hence proportional hazard model.

Estimate for Cox Proportional Hazards Model
The Cox Proportional hazard model was fitted to estimate $h_o(t)$ and β. This approach maximizes the likelihood function simultaneously for the observed data with respect to $h_o(t)$ and β. Hence, the partial likelihood is a technique developed to make inferences about the regression parameters in the presence of nuisance parameters $h_o(t)$ in the Cox PH model
- P_i refers to possible censored failure time random variables.
- β_i is the failure/censored indicators (*1*=fail,*0*=censor).
- X_i represents a set of covariates.

Let $R(t) = \{i: Xi \geq t\}$ denote the set of individuals who are at risk of failure at time t, known as the risk set, then at each failure time X_J, the contribution to the likelihood is:

$$L_j(\beta) = P\{individual\ j\ fails/one\ failure\ from\ R\left(X_j\right)\} = \frac{P\left(individual\ j\ fails/one\ failure\ from\ R(X_j)\right)}{\sum_{I \varepsilon r}(X_j)P(individuals\ j\ fails/at\ risk\ at\ X_J)} \tag{13}$$

$$= \frac{h(X_j/Z_j)}{\sum_{i \varepsilon r}(X_j)h(X_j/Z_j)} \tag{3.14}$$

Under the PH assumption that $h(t/Z) = h_0(t)e^{\beta'Z}$, then the partial likelihood function is given by;

$$L(\beta) = \prod_{j=1}^{k} \frac{h_0(x_j)e^{\beta'z_j}}{\sum_{i \varepsilon r}(x_j)h_0(x_j)e^{\beta'z_j}} \tag{15}$$

Where $\prod_{j=1}^{k} \frac{e^{\beta'z_j}}{\sum_{i \varepsilon r}(x_j)e^{\beta'z_j}}$ \hfill (16)

Proportional Hazard Assumption Checking
The Cox Proportional Hazard Model assumes that the hazard function of one individual is proportional to that of the other individual which is also constant over time.

Graphical Method
Graphical method to check proportional hazard assumption was performed based on the Arjas plot which shows the estimated cumulative hazard versus number of failures and comparison was made to the other procedures including non-proportional hazards such as crossing a non-monotonic hazards. The smoothed plot of the difference between log cumulative baseline hazard rates and time, and the smoothed plot of scaled Schoenfeld residuals versus time are compared for the increasing hazards. Hence, the survival function for the Cox PH is found y determining the relationship between survival function and hazard function as follows;

$$S(t, x_1) = S_0(t)^{exp}\left(\sum_{i=1}^{p} \beta_i x_i\right) \tag{17}$$

Taking logarithm twice we get:

$$ln[-lnS(t, x)] = \sum_{i=1}^{p} \beta_i x_i + ln[-lnS_0(t)] \tag{18}$$

Therefore, the difference in the log-log curves that corresponds to two different individuals with variables, $x_1 = (x_{11}, x_{12} \ldots \ldots \ldots x_{1p})$ and $x_2 = (x_{21}, x_{22} \ldots \ldots \ldots x_{2p})$ is given by;

$$ln[-lnS(t, x_1)] - lnS(t, x_2) = \sum_{i=1}^{p} \beta_i(x_{1i} - x_{2i}) \tag{19}$$

Test based on the schoenfeld residuals
If Schoenfeld $H_o < 0.05$, then the model or feature does not meet the PH assumption and when $H_1 > 0.05$ the proportional hazard assumption is met. The test is accomplished by finding the relationship between the Schoenfeld residuals for a particular covariate and the ranking of individual survival time. The null hypothesis is that the relationship between the Schoenfeld residuals and the ranked survival time is zero. Rejection of null hypothesis concludes that PH assumption is violated.

Cox proportional hazard model analysis using cox-snell and deviance residuals
The Cox-Snell residuals for the ith individual with observed survival time t_i is defined as;

$$r_{ei} = exp(\beta' x_i) H_0(t_i) = -log S_0(t_i) \tag{20}$$

Where $H_0(t_i)$ is the estimate of the baseline cumulative hazard function at time t_i.
That the residual is based on the following findings. Let T have a continuous survival distribution $S(t)$ with the cumulative hazard $H(t) = -log(S(t))$. Then, $ST(t) = exp(-H(t))$. Let $H(T)$ be the transformation of T based on the cumulative hazard function, then, the survival function for Y is:

$$ST(y) = P(Y > y) = P(H(t) > y) \tag{21}$$

$$P(T < HT(y)) = ST(HT(y)) \tag{22}$$

$$exp(-HT(HT(y)) = exp(-y) \tag{23}$$

Thus, regardless of the distribution of T, the new variable, Y=H (T), has an exponential distribution with unit mean. If the model is well fitted, the value $S_i(t_i)$ will have a similar property with that of $ST(y)$ and $r_{ci} = -log S_i(t_i)$ will have a unit exponential distribution with $R(r) = exp(-r)$.

Let $S_R(r)$ denote the survival function for Cox-Snell residual r_{ci}, then

$$S_R(r) = \int_r^\infty f_R(x)\,dx = \int_r^\infty exp(-x)\,dx = exp(-r) H_R(r) \tag{24}$$

$$H_R(r) = -log S_R(r) = -log(exp(exp(-r)) = r \tag{25}$$

And

$$H_R(r) = -log S_R(r) = -log(exp(-r)) = r \tag{26}$$

Schoenfeld Residuals
Schoenfeld residuals was originally called partial residuals because for i^{th} individual on the jth explanatory variable, X_j is an estimate of the ith component of the first derivative of the logarithm of the partial likelihood function with respect to β_j. The partial likelihood function is given by;

$$\frac{\sigma log L(\beta)}{\sigma \beta_j} = \sum_{i=1}^p \sigma_i (x_{ij} - a_{ij}) \tag{27}$$

Where x_{ij} is the value of the jth explanatory variable j=1, 2., p for the ith individual and;

$$a_{ij} = \frac{\sum_{i \in R(t_j)} x_{ij} exp(\beta' x_1)}{\sum_{i \in R(t_j)} exp(\beta' x_i)} \tag{28}$$

The Schoenfeld residual for i^{th} individual on X_j is given by $r_{pji} = \delta_i [x_{ji} - a_{ji}]$.

Testing for Goodness of Fitted model
Akaike information criterion (AIC) and Cox-Snell residual concept were applied in this research project to test the goodness of the fitted model. AIC trades off the complexity of an estimated model against how well the model fits the data. The AIC is given by;

$$AIC = -2 * log(likelihood) + 2(p + k)) \tag{29}$$

Where, the number of parameter/covariates is defined by p, exponential model denoted by k=1 and log-logistic and Weibull model denoted by k=2 (Klein *et al.*, 1997) [19]. Better likelihood is indicated by lower AIC. The Cox-Snell residual for the ith individual with observed time t is defined as:

$$R_{ci} = S_i(t) = (S_{\varepsilon i}(log t - \mu - \beta_1 x_1 - \cdots \beta_p \beta_p)/\sigma) \tag{30}$$

With respect to distribution, competing models are judged by approximation of cumulative Cox-Snell residuals to (-log), Kaplan-Meier estimates and AIC minimization. The researcher applied the optimal model identified in predicting child mortality.

Testing for the statistical significant of the variables

The Wald test (Wald Chi-Squared Test) is estimated by dividing the maximum likelihood estimate of the slope parameter by the estimates of its standard error as follows.

$$W = \frac{(\hat{\theta} - \theta_0)^2}{var(\hat{\theta})} \tag{31}$$

The Null hypothesis is $W_0 < 0.05$ and the Alterative Hypothesis is $W_1 > 0.05$. The null hypothesis is rejected when Wald test value is greater than 0.05 and made the conclusion that the particular independent variable is statistically insignificant and when W_0 is less than 0.05 we conclude that the independent variable is statistically significant.

Checking the effect of the Ratio factor

The assumption of the Cox proportional hazards model suggests a multiplicative effect of the predictors on the hazard and that this effect is constant over time. This is given by;

$$h\left(\frac{t}{x}\right) = h_0(t)e\beta_1 x_1 + \cdots . + \beta_p x_p \tag{32}$$

Therefore, the interpretation of the Cox model is done using hazard ratios (HR). The HR has also been defined as the ratio of (risk of outcome in one group), (risk of outcome in another group), occurring at a given interval of time.

$$h = \frac{d_A / e_A}{d_B / e_B} \tag{33}$$

A hazard ratio greater than 1 means the event is more likely to occur, and a ratio less than one means an event is less likely to occur. A hazard ratio of 1 means the predictor has no effect on the hazard of the event.

Test for overall statistically significant of the fitted model.

When T_1 and T_2 are continuous independent random variables with cumulative distribution functions F_1 and F_2, then the concordance index is:

$$C = P(T_1 > T_2) = \int \{1 - F_1(u)\} dF_2(u) \tag{34}$$

If T_1 and T_2 place positive mass at the same point, then we count half for ties and define C as;

$$C = \int \{1 - F(u) + \tfrac{1}{2} P(T_2 = u)\} dF_2(u) \tag{35}$$

If the two distributions are the same or have ties, then C=0.5. The concordance index can be estimated from the normalized Wilcoxon rank sum (Mann–Whitney) statistic as follows.

$$\hat{C} = (nm) - 1 \sum i = 1n \sum j = 1mI(T1i > T2j) + 12I(T1i = T2j) \tag{36}$$

Where T1i (i = 1,…, n) and T2j (j = 1, …, m) are independent samples from F1 and F2 respectively, and I denotes the indicator function. The hypothesis will be;
H_o: Concordance index is less than 0.05
H_1: Concordance index is greater than 0.05

This can also be tested using the Wald statistics. Confidence interval tells the actual coefficient value can lie within that range. If that interval includes 0, that means the actual coefficient value can be zero and that means that the predictor has no relationship with the response variable or it is insignificant in terms of its influence on response variable.
The likelihood ratio test statistic ($-2 \log L$) to test nested GLM models is;

$$-2logl0l1 = -2logl0 - logl1 = -2l0 - l1 \tag{37}$$

Where l_0 is the maximized value of the likelihood function for the simpler model and l_1 is the maximized likelihood function for the full model. L_0 and L_1 represent the maximized log-likelihood functions and are the transformations of l_0 and l_1. This transformation yields a chi-squared statistic with degrees of freedom (df) equal to the number of parameters in the null hypothesis. If the likelihood ratio test is less than 0.05 then we conclude that the overall goodness of the model is met and if it is less than 0.05 then we conclude that the overall goodness of the model is not met.

Chapter four: Data analysis, presentation and interpretation
**Descriptive statistics**
This study involved a sample of 3979 at Moi Teaching and Referral Hospital over five years' period between the year 2015 and 2019. The findings suggest that majority of the children who died during this period had an average age of 2.252 (mean=2.252) with maximum age being five years and minimum age of one year. Majority of the deaths occurred in the year 2016 with a total number of 1055 of children (F=1055). In addition, the child mortality ratio of males is higher compared to that of female. Furthermore, the findings show that most of the deaths were caused by birth asphyxia followed by congenital malformations. Tuberculosis was found to be the least possible cause of child mortality.

Table 1: Causes of child mortality

Variables	Frequency (censored)	Frequency (death)
Gender		
Female	1099	862
Male	1123	895
Year		
2015	290	235
2016	1437	1055
2017	273	259
2018	155	149
2019	57	69
Cause		
Birth Asphyxia	1459	833
Cancer	80	68
Congenital Malformations	145	151
Dehydration	21 42	36
Diarrhea	39	74
Digestive system	26	39
Heart	17	34
Kidney	21	23
HIV	9	22
Malaria	38	10
Malnutrition	35	50
Meningitis	120	51
Neonatal Sepsis	44	112
Pneumonia	35	47
Poisoning	60	59
Respiratory	4	99
Tuberculosis	27	8
Others		43

Testing for the proportional hazard assumption
Schoenfeld Residuals
The Schoenfeld residuals plots depict a random pattern that is fairly flat hence showing that there is no violations of the proportional hazard assumption for the variables (gender and cause) including the global variable thus suggesting that proportion hazard assumption holds for this model (Figure 1).
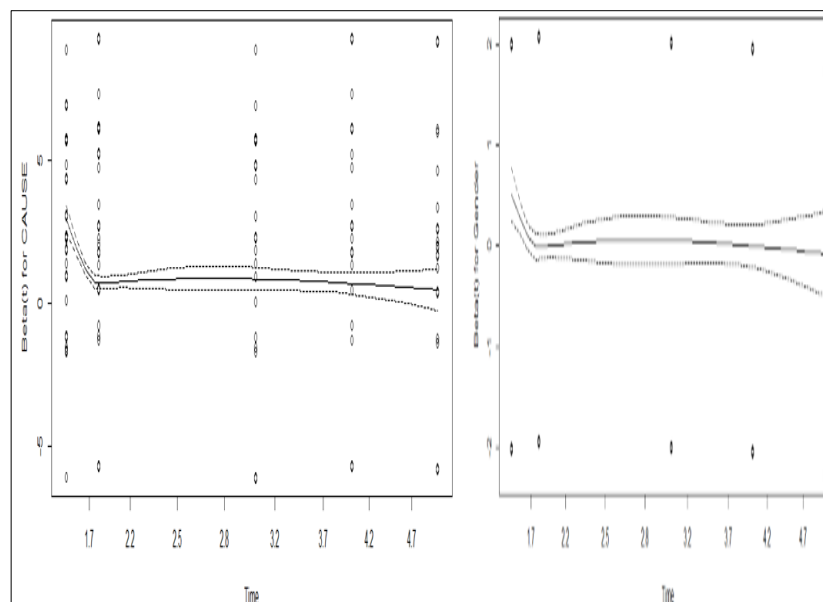
Fig 1: Schoenfeld Residuals

Test based on stitatistical test
In addition, test of assumption for proportional hazard was also performed with statistical test and graphical diagnostics using Schoenfeld residuals. The results as depicted in Table 2 below show statistical tests that suggest that both the two variables (gender and cause) were not statistically significant. Also, it was found out that the global test was also not statistically significant, hence, implying that the proportional hazards assumption was not violated in this study datasets.

Table 2: Test of proportional hazard assumptions

	Chisq	DF	P
Gender	1.23	1	0.27
Cause	24.64	17	0.10
Global	25.72	18	0.11

Comparison between the log-normal AFT and CPH Models
Cox-Snell residual and Akaike information criterion (AIC) were used to compare the results of Cox PH model with AFT model. Cox-Snell residuals were plotted for both the Log-Normal AFT and CPH models and they are as shown in Figure 2 below. Basing on the analysis of the Cox-Snell residual plots as shown in Figure 4.2 below, the results suggest that Cox PH model and Accelerated Failure Time Model had approximately equal fitness.
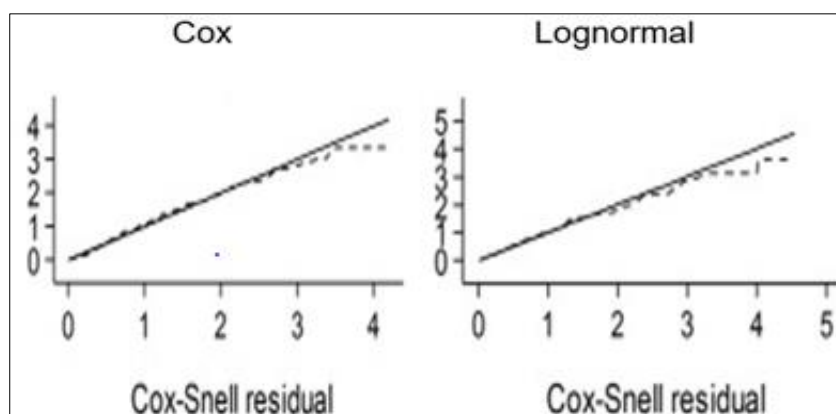
Fig 2: Cox-snell residual plots

According to AIC results appearing in Table 3 below, it can be seen that Cox PH model outperforms Accelerated Failure Time model in the analyzing child mortality for MTRH hospital located in Uasin Gishu County. Therefore, in this study Cox PH model was found to be optimal as compared to Accelerated Failure Time Model based on Log-normal model.

Table 3: AIC Results of Lognormal AFT and CPH Models

Model	AIC Value
Cox PHM	6364.75
Lognormal	6908.157

Cox Proportional hazard model results
Cox Proportional Hazard Model was fitted in order to predict the child mortality hazard. The results are presented in Table 4 below.

Table 4: Cox Proportional Hazard Model

Variables	COEF	EXP (COEF)	Z	PR(>|z|)
Gender	0.06144	1.06337	1.279	0.201
Cancer	-0.3662	0.69336	-2.83	0.0034 **
Malformations	0.29412	1.34194	3.316	0.001 ***
Dehydration	0.61654	1.85251	3.615	0.000 ***
Diarrhea	0.61045	1.84127	5.014	$5.3e^{-07}$ **
Digestive	0.14731	1.15871	0.890	0.373
System				
Heart	0.53958	1.71529	3.081	0.002 **
Kidney	0.25425	1.28949	1.201	0.230
HIV	0.04267	1.04360	0.197	0.844
Malaria	0.01206	1.01214	0.038	0.970
Malnutrition	0.49907	1.64718	3.417	0.001 ***
Meningitis	0.39234	1.48044	2.700	0.007 **
Neonatal	0.32838	1.38871	3.259	0.001 **
Sepsis				
Pneumonia	0.71293	2.03996	4.751	2e-06 ***
Poisoning	0.33313	1.39533	2.464	0.0137 *
Respiratory	0.87403	2.39656	8.212	< 2e-16 **
Tuberculosis	0.27841	1.32103	0.782	0.434
Others	0.21787	1.24343	1.388	0.165

Test of statistical significance of the variables
H$_o$: If p-value for Wald test is less than.05, then the variable is statistically significant.
H$_1$: If p-value for Wald test is greater than.05, then the variable is statistically significant.

From Table 5 below, it can be seen that the gender coefficient is positive but not significant. This implies that the male child mortality hazard is seen to increase with respect to the female child though the difference is not significant. Cancer with a negative beta coefficient indicates that Child mortality hazard for Children suffering from cancer is decreased as compared to those suffering from Birth Asphyxia though highly statistically significant. For the children suffering from Malformation, Dehydration, Diarrhea, Heart problem, Malnutrition, Meningitis, Neonatal Sepsis, Pneumonia, Poisoning and Respiratory problem has a positive beta coefficients which indicates that children mortality hazard is increased as compared to those suffering from Birth Asphyxia and are statistically significant. Moreover, the table further indicates that those Children suffering from Digestive problem, Kidney problem, HIV, Malaria, Tuberculosis, and Other factors not examined has positive beta coefficients which indicates that child mortality hazard is increased as compared to those suffering from Birth Asphyxia though not statistically significant.

Table 5: Statistical Tests as per the Fitted CPHM on the child mortality data

| Concordance	0.6 (se = 0.009)	
Likelihood ratio test	138.2 with 18	df, p=<2e-16
Wald test	147.5 with 18	df, p=<2e-16
Score (log rank) test	153.6 with 18	df, p=<2e-16

The Hazard Ratio
H$_o$: HR=1
H$_1$: HR ≠1

We can see that the variable Gender, Malformation, Dehydration. Diarrhea, Digestive system, Heart, Kidney, Malaria, HIV, Malnutrition, Meningitis, Neonatal Sepsis, Poisoning, Tuberculosis, pneumonia, Respiratory and other factors have Hazard ratio greater than 1. This show that these variables increases Mortality Hazard while variable Cancer with Hazard Ratio less than 1 shows that the effect of cancer in mortality hazard is reduced. As shown in Table 6 below.

Table 6: Hazard Ratio

Variables	HR
Gender	1.06337
Cancer	0.69336
Malformations	1.34194
Dehydration	1.85251
Diarrhea	1.84127
Digestive system	1.15871
Heart	1.71529
Kidney	1.28949
HIV	1.04360
Malaria	1.01214
Malnutrition	1.64718
Meningitis	1.48044
Neonatal Sepsis	1.38871
Pneumonia	2.03996
Poisoning	1.39533
Respiratory	2.39656
Tuberculosis	1.32103
Others	1.24343

The Hazard Ratio Graph
The findings as shown in Figure 3 below show that the hazard ratios were all greater than one over the 2015-2019 period which increased mortality hazard. This clearly show that there is higher probability of deaths caused by these risk factors on the child except for cancer as a risk factor that had Hazard ratio less than 1. This implies that children suffering from cancer, there mortality hazard was decreased as compared to other children suffering from other risk factors which had their Hazard Ratio greater than 1.
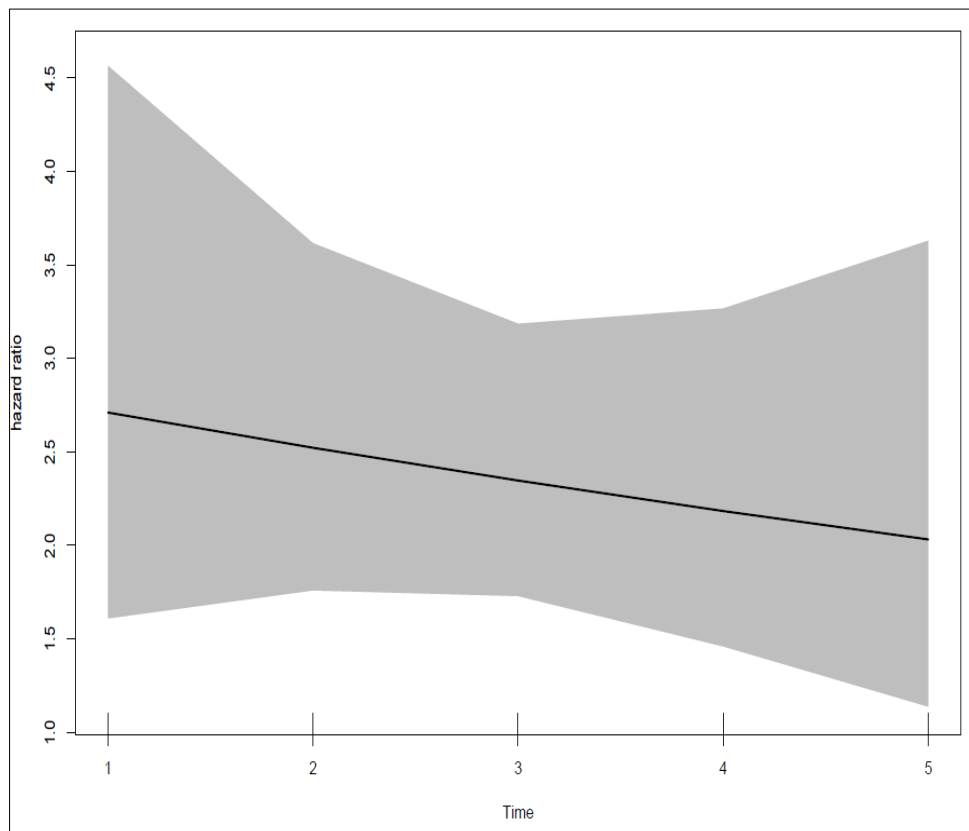
Fig 3: Hazard Ratio Graph

Test for statistically significant of the overall statistics
Ho: $W_O < 0.05$
H1: $W_1 > 0.05$

The three alternative tests including Wald test, the likelihood-ratio test and Score log-rank statistic for the overall significant of the model indicates that the model is statistically significant (P-Values<0.05).
The confidence interval was 95% (0.95) which is less than 1, this shows that parameters are good. However, if the confidence interval crosses 1, then the parameter is not good and this implies that there is no difference between arms of the study. Also, the likelihood ratio test shows that overall, the model is good since the likelihood ratio test is less than 0.05.

Predictions of child mortality
The findings as shown in Table 2 above indicates that the Wald Statistic value for gender variable is equal to 1.279(z=1.279) and is not statistically significant (P-Value=0.200879). Therefore, the gender variable does not have statistically significant coefficients. Also, the beta coefficient for gender=0.06144 (positive). This indicates that male child has higher risk of death than female child. The hazard ratio given by the exponential coefficients exp (coef) = 1.06337 suggest that being male increases the hazard of child mortality by a factor of 1.06337 times (6.3%) than female counterparts.
In addition, the results also show that the variable of cause has some factors that have statistically significant Wald Statistic values (p-values<0.05). Therefore, suggesting that cause has some factors with highly statistically significant coefficients. These factors included those with Cancer, Congenital Malformations, Dehydration, Diarrhea, Heart problems, Malnutrition, Meningitis, Neonatal Sepsis, Pneumonia, Poisoning and Respiratory problems. The findings further indicate that children with cancer had negative beta coefficients. This implies that children with cancer as the causative factor of child mortality were at lower risk of death as compared to children with Birth Asphyxia. Furthermore, the hazard ratio for children with Cancer was less than 1. This indicates that children with Cancer disease reduces the hazard of child mortality by a factor of 0.69 (31%) times respectively as compared to those having Birth Asphyxia. However, the beta coefficients for children with Congenital Malformations, Dehydration, Diarrhea, Heart problems, Malnutrition, Meningitis, Neonatal Sepsis, Pneumonia, Poisoning and Respiratory problems were positive. This indicates that children with these conditions as the causative factors of child mortality were at higher risk of death as compared to children with Birth Asphyxia. Moreover, the hazard ratio for children with Congenital Malformations, Dehydration, and Diarrhea, Heart problems, Malnutrition, Meningitis, Neonatal Sepsis, Pneumonia, Poisoning and Respiratory problems were greater than 1. This shows that children with these diseases increases the hazard of child mortality by a factor of 1.34, 1.85, 1.84, 1.71, 1.65, 1.48, 1.39, 2.04, 1.40 and 2.40 times respectively as compared to children with Birth Asphyxia.
On the other hand, the results also show that the variable of cause has some factors that were found to have insignificant Wald Statistic values (P-Values > 0.05). This shows that the variable of cause does not have highly statistically significant coefficients based on these factors including Digestive system problems, HIV, Kidney problems, Malaria, Tuberculosis and other factors not examined. The beta coefficients for those with these conditions as the causative factors of child mortality were found to be positive (β=0.14731, 0.25425, 0.04267, 0.01206, 0.27841 and 0.21787 respectively) though not statistically significant. This

indicates that children with Digestive system problems, HIV, Kidney problems, Malaria, Tuberculosis and other factors not examined are at higher risk of death than children with Birth Asphyxia. Moreover, the hazard ratio for children with Digestive system problems, HIV, Kidney problems, Malaria, Tuberculosis and other factors not examined were found to be 1.16, 1.04, 1.29, 1.01, 1.32 and 1.24 respectively. This suggests that children with Digestive system problems, HIV, Kidney problems, Malaria, Tuberculosis and other factors not examined increases the hazard of child mortality by a factor of 1.16 (16%), 1.29 (29%), 1.04 (4%), 1.01 (1%), 1.24 (24%) and 1.32 (32%) times respectively as compared to those having Birth Asphyxia though not statistically significant.

Chapter Five: Summary, Conclusion and Recommendations
Summary of the findings
In this study, data analysis was performed using the Cox PH model. The results suggest that majority of the children deaths occurred in the year 2016 with average age of 2 years. A large number of children deaths were of male gender and Birth Asphyxia was identified as the main causative factor of under-five child mortality. Others main causes that were identified included Congenital Malformations, Neonatal Sepsis, Respiratory problems, Diarrhea and Cancer. Tuberculosis, Malaria, kidney problems, Heart problems and HIV were among the least factors causing child mortality.

Children suffering from congenital malformation were at higher risk of death, this support the finding of Collen *et al*. (2013) who did a research on child mortality and their finding suggest that SIDS was the risk factor for child mortality. Male children were at higher risk of death than female children. This supports the findings by Kayode *et al*. (2013) [17] who examined the risk factors for child mortality in Nigeria using multivariate logistic regression analysis.

Conclusion
Child mortality has become a very important aspect of measuring the health status of the current population and predicting the health of the future generation. This study aimed at testing the proportional hazard assumption on child mortality data based on children under the age of five years that was from secondary data at Moi Teaching and Referral Hospital. Data was performed using Cox PH model since the proportionality assumption was met. Therefore, the analysis of Cox PH model suggests that gender was not statistically significant in this study for predicting child mortality. On the other hand, the variable of cause was found to influence child mortality though some of its factors were found to be statistically insignificant. Among the factors that were found to be significant included those children suffering from Cancer, Congenital Malformations, Dehydration, Diarrhea, Heart problems, Malnutrition, Meningitis, Neonatal Sepsis, Pneumonia, Poisoning and Respiratory problems. Hence, these factors were the most important significant causative factors influencing child mortality in this study.

Recommendations
This study concluded that Cox PH model was considered appropriated for predicting child mortality. However, one disadvantage of using Cox PH model is that it requires assumption of proportional hazard to hold where in some cases this is impossible. Further Research based on this study should be conducted using Cox Proportional Time Dependent Hazard model which does not requires the assumption of proportional hazard. Also, further studies should be conducted to include other aspects of practical cases including larger censoring.

References
1. Ayele DG, Zewotir TT, Mwambi H. Survival analysis of under-five mortality using Cox and frailty models in Ethiopia. Journal of Health, Population and Nutrition. 2017;36:1-9.
2. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis part III: multivariate data analysis–choosing a model and assessing its adequacy and fit. British Journal of Cancer. 2003;89(4):605-611.
3. Ezeh OK, Agho KE, Dibley MJ, Hall J, Page AN. Determinants of neonatal mortality in Nigeria: evidence from the 2008 demographic and health survey. BMC Public Health. 2014;14(1):1-10.
4. Frisbie WP, Song SE, Powers DA, Street JA. The increasing racial disparity in infant mortality: Respiratory distress syndrome and other causes. Demography. 2004;41:773-800.
5. Hailemariam A, Tesfaye M. Determinants of infant and early childhood mortality in a small urban community of Ethiopia: a hazard model analysis. Ethiopian Journal of Health Development. 1997;11:3.
6. Kayode GA, Grobbee DE, Koduah A, Amoakoh-Coleman M, Agyepong IA, Ansah E, *et al*. Temporal trends in childhood mortality in Ghana: impacts and challenges of health policies and programs. Global Health Action. 2016;9(1):31907.
7. Kembo J, Van Ginneken JK. Determinants of infant and child mortality in Zimbabwe: Results of multivariate hazard analysis. Demographic Research. 2009;21:367-384.
8. Mani K, Dwivedi SN, Pandey RM. Determinants of under-five mortality in rural empowered action group states in India: an application of cox frailty model. International Journal of MCH and AIDS. 2012;1(1):60.
9. Moise IK, Kalipeni E, Jusrut P, Iwelunmor JI. Assessing the reduction in infant mortality rates in Malawi over the 1990-2010 decades. Global Public Health. 2017;12(6):757-779.
10. Muriithi DM, Muriithi DK. Determination of infant and child mortality in Kenya using cox-proportional hazard model. American Journal of Theoretical and Applied Statistics. 2015;4(5):404-413.
11. Naz L, Patel KK. Determinants of infant mortality in Sierra Leone: applying Cox proportional hazards model. International Journal of Social Economics. 2020;47(6):711-726.
12. O'Leary CM, Jacoby PJ, Bartu A, D'Antoine H, Bower C. Maternal alcohol use and sudden infant death syndrome and infant mortality excluding SIDS. Pediatrics. 2013;131(3):e770-e778.
13. Omariba DWR, Boyle MH. Family structure and child mortality in Sub-Saharan Africa: Cross-national effects of polygyny. Journal of Marriage and Family. 2007;69(2):528-543.

14. Omolo S. Factors militating against Early Girl-Child Education: A Case Study of Asego Division, Homa-Bay District, Homa-Bay County; c2014.

15. Saroj RK, Murthy KHN, Kumar M, Singh R, Kumar A. Survival parametric models to estimate the factors of under-five child mortality. Journal of Health Research and Reviews. 2019;6(2):82.

16. Ezeh OK, Agho KE, Dibley MJ, Hall JJ, Page AN. Risk factors for postneonatal, infant, child and under-5 mortality in Nigeria: a pooled cross-sectional analysis. BMJ open. 2015 Mar 1;5(3):e006779.

17. Adekanmbi VT, Kayode GA, Uthman OA. Individual and contextual factors associated with childhood stunting in Nigeria: a multilevel analysis. Maternal & child nutrition. 2013 Apr;9(2):244-59.

18. Habte AT, Ayele DW. Synthesis and characterization of reduced graphene oxide (rGO) started from grapheme oxide (GO) using the tour method with different parameters. Advances in Materials Science and Engineering. 2019 Aug 15.

19. Klein G. The recognition-primed decision (RPD) model: Looking back, looking forward. Naturalistic Decision Making. 1997;285:292.