Infant mortality at the Kigali University Teaching Hospital: Application of Aalen additive hazards model and comparison with other classical survival models.

Paul Gatabazi, Sileshi Fanta Melesse & Shaun Ramroop

School of Mathematics, Statistics and Computer Sciences,

University of KwaZulu-Natal, Pietermaritzburg Campus

South Africa

Email: gatabazi001@gmail.com

Abstract

Background: Beyond the effort provided on the population policy in Rwanda so far, extensive studies on factors that could prevent infant mortality (IM) should be done for more controlling the Infant mortality rate (IMR). This study presents an application of survival analysis to the infant mortality at the Kigali University Teaching Hospital (KUTH) in Rwanda.

Data and methods: The dataset of the KUTH was recorded. Aalen Additive Hazard Model (AAHM) is used for assessing the relationship between the IM and covariates. The Cox Proportional Hazard Model (CPHM) and the Cox-Aalen Hazard Model (CAHM) are also applied, the results of these three models are compared.

Findings: The AAHM distinguishes time dependent and fixed covariates, and this allows an easy interpretation of the results found in CPHM and CAHM.

Conclusion: Avoidance of pregnancy until after age 20 and clinically recommended nutrition for the mother during pregnancy would decrease IM.

Keywords: Survival analysis; counting processes; martingales; cumulative parameter function; Cox Proportional Hazard Model; Aalen additive hazards model.

Introduction

Infant mortality or mortality of children under their first birthday (Bourgeois, 1946; Reidpath and Allotey, 2003) attracts attention in several studies worldwide. Ester et al. (2011) pointed that half of the 10 million children who die every year in the world are in Sub-Saharan Africa (SSA). The study stress three factors that are inversely related to the IMR namely higher social security expenditure on health as a percentage of the general government expenditure on health, a higher per capita government expenditure on health and a higher number of children under 5 years of age with diarrhoea receiving oral rehydration therapy.

Benn Sartorius and Kurt Sartorius (2014) used data of the World Bank of Development from 192 countries from 1990 to 2011 and found that the average of the Infant Mortality rate (IMR) is 75/1000 in SSA versus 11/1000 in developed countries. Maternal mortality, lack of access to sanitation, female education, and lack of access to water are pointed as the most prominent risk factors for infant mortality. Other studies on infant mortality include Adetunji and Bos (2006) where the World Bank dataset from 1960 to 2005 is used and the study

suggests that low life expectancy at birth in SSA is relatively higher in Middle Africa as compared to other sub-regional disparities of SSA, Schell et al. (2007) who contributed in studying socio-economic determinants of infant mortality in 152 low, middle and high-income countries worldwide, Mturi and Curtis (1995) who studied determinants of infant and child mortality in Tanzania, Sartorius et al. (2011) who conducted an ecological spatial analysis on the infant mortality in South Africa, Akunga et al. (2014) who identified determinants of postnatal care use in Kenya and Rugiranka et al. who analysed factors associated with anemia among children in Lesotho.

The 2017 World Factbook includes 36 SSA countries in the top 40 countries by IMR, and 4 countries outside of SSA namely Afghanistan (IMR=110.60/1000), Laos (IMR=49.90/1000), Haiti (IMR=46.80/1000) and Yemen (IMR=46.00/1000). The top five SSA countries in IMR were Somalia (IMR=94.80/1000), Central African Republic (IMR=86.30/1000), Guinea Bissau (IMR=85.70/1000), Chad (IMR=85.40/1000) and

Niger (IMR=81.10/1000). The lowest IMR in SSA were found in Mauritius (9.80/1000).

The World Bank records of 2017 indicated that the IMR was 51.50/1000 in SSA, this confirms that the IMR remains a problem in SSA.

These various studies on the infant mortality leave a question on how factors are mathematically associated to the IMR. The incidence of relatively higher rate in SSA justify the need to identify and analyse mathematically the major factors of the infant mortality in SSA, for providing a help to the medical practitioners and policy makers to implement security measures for a better control of the infant mortality.

The IMR in Rwanda was 28.90/1000 in 2017 as evokes the 2017 World Factbook; this is a decrease from 107/1000 in 2000 to 56.90/1000 in 2007 and 37.50/1000 in 2012, as shows the 2017 Rwanda Statistics Portal on the IMR. A relatively decreasing trend of the IMR in Rwanda is due to the new population policy from 2003 that consists of slowing population growth, enhancing food safety, implementing access to primary and secondary education for all children, managing natural resources and reinforcing participation in development by both women and men (Rathavuth et al., 2009). However the IMR in Rwanda is still problematic and therefore, beyond the effort provided so far, extensive studies on factors that could prevent infant mortality should be done for more preventing the IMR. The present studies tackle a central hospital in Rwanda known as Kigali University Teaching Hospital (KUTH).

This article aims at using Kaplan-Meier estimation for presenting survival outcomes of infant mortality per covariate, and for measuring the PHA. AAHM will be conducted and fully interpreted for all covariates with event taken as the infant mortality. AAHM will indicate time dependent covariates and will allow a detection of fixed covariates that are adapted to the CPHM. AAHM will also give idea on covariates of multiplicative and additive parts of the CAHM. Significance will be measured for comparing performance of models. Comparison on the performance for different models will then follow.

Material and methods
Concept of survival analysis
Survival analysis known also as time to event analysis aims at making inferences on the time elapsed between the onset of observations, until the occurrence of some event of interest. In short,

regression model in survival analysis measures the dependence of time to event on predictor variables. Methods used in general statistical analysis, in particular in regression analysis, are not directly applicable to survival data due to censoring and truncation. Hosmer et al. (2008) describes three types of censoring: *Right censoring* arising when an individual is not subject to the event until the end of study due to either loss to follow up, or the event has not occurred at the end of the study, or the event has occurred from another cause not related to the cause of interest. *Left censoring* arises when an individual experienced an event before recruitment. *Interval censoring* refers to when the event occurs within some interval at the study termination, or the individual dropped out or observed the event before study termination for reasons unrelated to the study, or the individual was lost to follow-up. Klein and Moeschberger (2003) describe two types of truncation: *left truncation* occurs when subjects under a survival study have been at risk before the study time and *right truncation* when interest is only on individuals who have experienced the event by a specified future time before study termination. In this study, interest will be only on right censoring.

In survival analysis, a non-negative random variable representing the time to event is generally characterized by three fundamental functions: the probability density function (in continuous case) or probability mass function (in discrete case), the survival function and the hazard function known also as risk function or intensity rate.

Any of these three functions can be uniquely determined from at least one of the other two functions (Klein and Moeschberger, 2003; Hosmer et al., 2008; Collet, 2003).

Comparison of two or more groups of survival data
Two or more groups' survival time may be compared by using the plots of the survival functions in one system of axes. Log-rank and Wilcoxon tests are popular tests for comparing survival functions (Collet, 2003). The tests are based on the following hypotheses:

H_0: no difference in survival experiences of the individuals in groups,

H_1: there is difference in survival experiences of the individuals in groups.

The interpretation of tests is summarized in Table 1 suggested by Collet (2003).

Table 1: Evidence for or against H_0 based on comparing the p-value with the level of significance $\alpha = 0.05$.

p-value (P)	Interpretation
$P > 0.1$	No evidence to reject the null hypothesis
$0.05 < P \leq 0.1$	Slight evidence against the null hypothesis
$0.01 < P \leq 0.05$	Moderate evidence against the null hypothesis
$0.001 < P \leq 0.01$	Strong evidence against the null hypothesis
$P \leq 0.001$	Overwhelming evidence against the null hypothesis

The log-rank test is suitable if proportional hazards can be assumed. In such situation, the plots of survival functions do not cross one another. The Wilcoxon test is suitable when there is no proportional hazards assumption. Here, the plots cross one another.

Cox proportional hazards model (CPHM)

Assume p fixed covariates with values $x_i = (x_{i1}, x_{i2}, ..., x_{ip})'$ and $h_0(t)$ a hazard function when values of all covariates are zeros. The CPHM is given in [7] as

$$h(t|x_i) = h_0(t)\exp(\beta'x_i) \qquad (1)$$

where $\beta = (\beta_1, \beta_2, ..., \beta_p)'$ is a p-dimensional vector of model parameters. The quantity

$$\psi = e^{\beta_k} \qquad (2)$$

is called "*hazard ratio*" and is reported in applied studies as it is easier to interpret than the log-hazard ratio $\beta_k = \ln\psi$ (Collet, 2003).

Parameter estimation for model (1) with no tied events is done using partial likelihood introduced by Cox (1972). Three approaches of approximating the partial likelihood in presence of tied event are suggested by Breslow (1974), Efron (1977) and Cox (1972). In practice, the three approximations of the partial likelihood function lead to similar results (Collet, 2003). STATA provides options for using each of the above approximations with Breslow being the default.

Aalen additive hazards model (AAHM)

The AAHM expresses the hazard rate at time t of the ith of n individuals with vector of covariates $x_i(t) = (x_{i1}, x_{i2}, ..., x_{ip})'$. That is given by

$$h[t|x_i(t)] = \beta_0(t) + \beta_1(t)x_{i1}(t) + \beta_2(t)x_{i2}(t) + ... + \beta_p(t)x_{ip}(t) \qquad (3)$$

where $\beta(t) = (\beta_0(t), \beta_1(t), ..., \beta_p(t))'$ is the vector of parameter functions that may be estimated and $\beta_0(t)$ is the baseline hazard (Aalen, 1989).

Aalen et al. (2008) argue that, for computation stability, estimation in model (3) should be based on the cumulative parameter functions

$$B_k(t) = \int_0^t \beta_k(v)dv, \qquad (4)$$

$k = 0, 1, 2, ..., p$. Clearly, if $\beta_k(t)$ is constant, say $\beta k(t) = \beta_k$, then $B_k(t) = \int_0^t \beta_k dv = \beta_k t$ which is represented by a straight line.

Let $Y_i(t) = \begin{cases} 1, & \text{if individual } i \text{ is at risk at time } t \\ 0, & \text{Otherwise.} \end{cases}$

It can be shown by using stochastic counting processes that model (3) leads to the form

$$dN_i(t) = \sum_{k=0}^{p} Y_i(t)x_{ik}(t)dB_k(t) + dM_i \qquad (5)$$

where $x_0 = 1$.

Model (5) has the form of a multiple linear regression model for the ith individual with response variable (observations) $dN_i(t)$, covariates $Y_i(t)x_{ik}$, random error terms $dM_i(t)$ and parameters $dB_k(t)$ for $k = 0, 1, 2, ..., p$ and $i = 0, 1, 2, ..., n$. Model (5) can be written in matrix form as

$$d\mathbf{N}(t) = \mathbf{X}(t)d\mathbf{B}(t) + d\mathbf{M}(t) \qquad (6)$$

where
$d\mathbf{N}(t)$ is the $n \times 1$ vector of observations $dN_i(t)$,
$\mathbf{X}(t)$ is the $n \times (p+1)$ design matrix with i^{th} row $Y_i(t)$, $Y_i(t)x_{i1}(t), ..., Y_i(t)x_{ip}(t)$,
$d\mathbf{B}(t) = (dB_0(t), dB_1(t), ..., dB_p(t))'$ is the $(p+1) \times 1$ vector of parameter functions,
$d\mathbf{M}(t)$ is the $n \times 1$ vector of martingales (error terms) each with mean zero.
It follows from (6) and from the theory of least square estimation that if $\mathbf{X}(t)$ is of full rank, that is $[\mathbf{X}(t)]'\mathbf{X}(t)$ is non-singular, then the ordinary least squares estimator of $d\mathbf{B}(t)$ is

$$d\widehat{\mathbf{B}}(t) = [(\mathbf{X}(t)'\mathbf{X}(t)]^{-1}(\mathbf{X}(t)'d\mathbf{N}(t). \qquad (7)$$

If $\mathbf{X}(t)$ is not of full rank, then $d\mathbf{B}(t)$ is not estimable unless some constraint is imposed.

However, most of current statistical packages have built-in routines to deal with matrices that are not of full rank and provide robust estimates of model parameters. Integrating both sides of equation (7) with respect to t yields

$$\hat{\mathbf{B}}(t) = \int_0^t [(\mathbf{X}(t))'\mathbf{X}(t)]^{-1}(\mathbf{X}(t))' d\mathbf{N}(t)$$
$$= \Sigma_{t_j \le t}[(\mathbf{X}(t_j))'\mathbf{X}(t_j)]^{-1}\left(\mathbf{X}(t_j)\right)' y_j \quad (8)$$

where y_j is $n \times 1$ vector of zeros except the jth component equals to unit if the jth individual observes an event at time t_j [2]. Furthermore, the variance-covariance matrix of $\hat{\mathbf{B}}(t)$ is

$\mathrm{Var}[\hat{\mathbf{B}}(t)] =$
$\Sigma_{t_j \le t}[(\mathbf{X}(t_j))'\mathbf{X}(t_j)]^{-1}\left(\mathbf{X}(t_j)\right)'\mathbf{D}(t_j)\mathbf{X}(t_j)[(\mathbf{X}(t_j))'\mathbf{X}(t_j)]^{-1}$
$$(9)$$

where $\mathbf{D}(t_j)$ is an $n \times n$ diagonal matrix with elements y_j on the main diagonal (Aalen et al., 2008; Hosmer and Royston, 2002). The derivation of results (9) from (8) is easy to understand. In fact, if two random vectors of variables $\mathbf{X}$ and $\mathbf{Y}$ are linked by $\mathbf{Y} = \mathbf{AX}$, where $\mathbf{A}$ is a matrix, then

$$\mathrm{Var}(\mathbf{Y}) = \mathbf{A}\mathrm{Var}(\mathbf{X})\mathbf{A}'.$$

Hosmer and Royston (2002) assumed that if the vector of cumulative parameter coefficients at time t is estimated by (8), and its variance-covariance matrix by (9), then the estimator of the model vector of parameter coefficients at time t_j is

$$\hat{\beta}(t_j) = [(\mathbf{X}(t_j))'\mathbf{X}(t_j)]^{-1}\left(\mathbf{X}(t_j)\right)' y_j \quad (10)$$

and

$\mathrm{Var}[\hat{\beta}(t_j)] =$
$[(\mathbf{X}(t_j))'\mathbf{X}(t_j)]^{-1}\left(\mathbf{X}(t_j)\right)'\mathbf{D}(t_j)\mathbf{X}(t_j)[(\mathbf{X}(t_j))'\mathbf{X}(t_j)]^{-1} \quad (11)$

Aalen et al. (2008) showed that the cumulative parameter function estimator has approximately a multivariate normal distribution around its true value $\mathbf{B}(t)$, with the variance-covariance matrix expressed in (9). Therefore, the $100(1-\alpha)$ % confidence interval for the kth cumulative parameter functions $\mathbf{B}_k(t)$ is expressed by

$$\hat{B}_k(t) = \pm z_{\frac{\alpha}{2}}\sqrt{\hat{\sigma}_{kk}(t)} \quad (12)$$

The expression in square root is the kth diagonal element of the variance-covariance matrix expressed in the equation (9). To test that a covariate X_k has no significant effect on the hazard function given in model (3), Aalen et al. Aalen et al. (2008) formulated

the null and alternative hypotheses in the usual way as follows

H_0: $\beta_k(t) = 0, \forall t \,\epsilon\, [0, t_0]$
versus
H_1: $\beta_k(t) > 0$ or $\beta_k(t) < 0$

where t_0 is a suitably chosen time point, but often t_0 is the upper limit of the study time interval. If H_0 is true, then the increment $\Delta\hat{B}_k(t)$ at time t_j tends to fluctuate around zero Aalen et al. (2008). Under the alternative hypothesis H_1: $\beta_k(t) > 0$, the increment $\Delta\hat{B}_k(t)$ tends to be positive while under H_1: $\beta_k(t) < 0$, they tends to be negative. Furthermore, if $\hat{B}_k(t)$ approximately follows a straight line, then $\beta_k(t)$ is constant, that is not time-varying. The test described above is helpful when the estimated cumulative parameter functions are plotted against time. However, a quantitative measure of significance may be needed to assess the magnitude of significance. Hosmer and Royston (2002) advised to proceed as follows. Consider model (3) and assume that there is a need to test the null hypothesis

H_0: $\beta_k(t_j) = 0$ for k with $k = 0, 1, ..., p$. \quad (13)

Hosmer and Royston (2002) stated that the $(p+1)$ statistics for the above hypothesis are obtained from the components of the vector

$$\hat{\mathbf{u}} = \Sigma_{t_j} \mathbf{K}_j \hat{\beta}(t_j) \quad (14)$$

where $\hat{\beta}(t_j)$ given in (10) is the vector of estimators of the parameter coefficients for model (3), and $\mathbf{K}_j$ is a $(p+1) \times (p+1)$ diagonal matrix of weights. Four types of weights can be used.

Weights 1: $\mathbf{K}_j$ = diag (1), that is $\mathbf{K}_j$ is a diagonal matrix with each element of the main diagonal equals to unit.

Weights 2: $\mathbf{K}_j$ = diag(n_j) where n_j is the number of individuals at risk at time t_j.

Weights 3: $\mathbf{K}_j$ = diag[$\hat{S}_{KM}(t_{j-1})$] where $\hat{S}_{KM}(t_{j-1})$ is the Kaplan-Meier estimate of the survival function at time t_{j-1} for $j = 2, 3, ...$ and $\mathbf{K}_1$ = diag[$\hat{S}_{KM}(t_0) = 1$].

Weights 4: $\mathbf{K}_j$ = diag[$\hat{S}_{KM}(t_{j-1})/se(\hat{\beta}_{kk}(tj))$] where $\hat{\beta}_{kk}(tj)$ is the kth diagonal element (i.e. a variance) of the variance-covariance matrix (11). Hence, $\mathbf{K}_j$ is a diagonal matrix whose main diagonal elements are the ratio of the Kaplan-Meier estimates of the survival function at time t_{j-1} and the standard error of the Aalen estimate of the parameter function of interest at time t_j. To completely define the test

statistic to use, the estimator of the variance-covariance matrix of $\hat{\mathbf{u}}$ in (14) is given by

$$\widehat{\mathrm{Var}}(\hat{\mathbf{u}}) = \sum_{t_j} \mathbf{K}_j \mathrm{Var}[\hat{\beta}(t_j)\mathbf{K}'_j]$$
$$= \Sigma_{t_j} \mathbf{K}_j \left[\left(\mathbf{X}(t_j)\right)' \mathbf{X}(t_j)\right]^{-1} \left(\mathbf{X}(t_j)\right)' \mathbf{D}(t_j)\mathbf{X}(t_j)\left[\left(\mathbf{X}(t_j)\right)' \mathbf{X}(t_j)\right]^{-1}\mathbf{K}'_j$$
(15)

Hence, the test statistic for H_0 given in (13) is

$$z_{u_k} = \frac{\hat{u}_k}{\mathrm{se}(\hat{u}_k)}$$
(16)

Where $\hat{u}_k$ is the kth element of $\hat{\mathbf{u}}$ given in (14) and $\mathrm{se}(\hat{u}_k)$ is the square root of the kth diagonal element $\widehat{\mathrm{Var}}(\hat{\mathbf{u}})$ given in (15). Hosmer and Royston (2002) pointed out that the statistic z_{u_k} in (16) approximately follows the standard normal distribution.

To implement the theoretical results discussed in this section, Hosmer and Royston (2002) provided an ado STATA command, *stlh*.

• Plot of the function (8) and their $100(1-\alpha)$ % confidence limits (12):

stlh "list of variables", level(#)

where # indicates the confidence level.

• Test of significance for the model parameter functions:

stlh "list of variables", testwt(#) nograph

where # indicates one or some or all the weight types discussed above.

1: all weights equal 1.

2: weight with size of the risk set

3: weight is the Kaplan-Meier estimator at time t_{j-1}

4: weight is the Kaplan-Meier estimator divided by the standard error of the parameter coefficients.

Other options are: *nodots* to suppress dots in outputs and *tcent(#)* to specify the upper limit of the time axis for the plots. Hosmer and Royston [13] pointed out that type 3 and mainly type 4 weights should be recommended since type 1 weights are

sensitive to later effects of covariates on the time to event while type 2 weights are sensitive to earlier effects.

Cox-Aalen Hazards Regression Models (CAHM)

CAHM was proposed by Scheike and Zhang (2002). The model consists of partitioning co-variates into two parts, one part working additively as in AAHM and other part acting multiplicatively as in CPHM. Assume that $Y(t)$ is the risk indicator, $(X(t), Z(t))$ is a $(p+q)\times 1$ vector of covariates; $\beta(t)$ is a $(p\times 1)$ of time-varying regression coefficient and α is a $(q\times 1)$ vector of relative risk regression coefficients. Then the hazard function is given by

$$h(t\,|\,x) = Y(t)[X'(t)\beta(t)]\exp(Z'(t)\alpha)$$
(17)

Estimation is based on cumulative parameter functions $B(t) = \int_0^t \beta(v)dv$ and model parameters α. Approximate maximum likelihood estimators are derived from the score function developed in Scheike and Zhang (2002).

Dataset

At KUTH, all newborns are recorded in registries with all details of parents and clinical outcomes of newborn. The information in registry provide also references on card index that provide information on clinical behavior of babies after leaving hospital. KUTH as sites of interest in this study is a central Hospital where most of complicated childbirth countrywide are transferred. KUTH records relatively high incidence of stillborn cases (69 stillborn babies or 32.59/1000) and relatively high infant mortality rate (82 babies died over 2048 babies born alive or 40.04/1000).

Table 2 describes the variables of interest, the full dataset can be found via the authors of this article

Table 2: Description of variables in the dataset on newborns at Kigali University Teaching Hospital (KUTH) during the period 01-January-2016 to 31-December-2016.

Variable	Description	Codes/Values/Unit
Age	Age of parent	0=under 20, 1=20 years old to 34 years old, 2=35 years old and above
Residence	Indicator of the residential area of a parent	0=rural , 1=urban
Antecedents	Indicator on whether a new born is the first or not	0=Not the first new born, 1= first newborn,

Abortion	Indicator on whether a parent aborted previously	0=not aborted, 1=aborted once, 2= aborted more than once
Childbirth	Type of childbirth	0=born using ventouse, 1=born naturally, 2= born after surgery
Gender	Gender of a newborn	0=female, 1=male
Number	Indicator of the number of births at a time	0=singleton, 1=multiple
APGAR	Score of appearance, pulse, grimaces, activity and respiration of a newborn	0= APGAR less than 4/10, 1=APGAR from 4/10 to 6/10, 2=APGAR greater or equal to 7/10
Weight	Weight of a newborn	0 = under 2500 g, 1= 2500 g to 4500 g, 2= above 4500 g
Head	Head circumference of a newborn	0= below 32 cm, 1=32 cm to 36 cm, 2=above 36 cm
Height	Height of a new born	0=below 46 cm, 1=46 cm to 54 cm, 2=above 54 cm
Time	Time from recruitment to study termination	Days
Event	Indicator describing if death occurred during the study or not	0=censored, 1=dead

The primary dataset of 2117 newborns at KUTH is recorded from 1st January to 31st December 2016 and a complete case analysis is considered. Beside the event status and the time to event, eleven covariates are of interest. The demographic covariates include the age and the place of residence for parents; clinical covariates for parents include obstetric antecedents, type of childbirth and previous abortion. Clinical covariates for children include APGAR; gender, number of births at a time, weight, circumference of the head, and height. The variable age was considered by Gourbin (2005) while studying interaction of infant mortality and age of their parents. Variables residence and gender are found in several datasets of survival analysis such as for example Collet (2003), Klein and Moeschberger (2003) and Flemming and Harrington (2005). The

standard pediatric measurement of newborns is found for example in Jansen et al. (2007).

Results and interpretation
Kaplan-Meier estimates of the survival, cumulative hazards and hazard functions for the infant mortality at KUTH
The Kaplan-Meier estimates of the survival, cumulative hazards and hazard functions are calculated and plotted. The 95% confidence limits of S(t), H(t) and h(t) are also calculated for the data at hand. Portions of the Kaplan-Meier estimates of the survival, cumulative hazards and the hazard functions along with the 95% confidence limits are given in Table 3, Table 4, and Table 5.

The entire Kaplan-Meier estimates along with the 95% confidence limits for survival and cumulative hazard functions are plotted in
Figure 1, Figure 2 and Figure 3.

Table 3: Survival function estimate and 95% confidence interval.

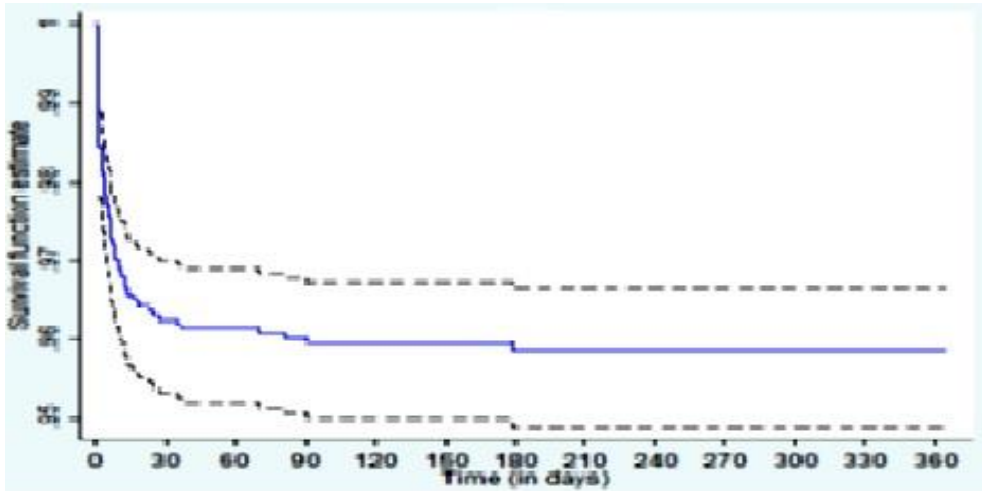| Time	Beg. Total	Fail	Net Lost	Surv. Function	Std. Error	95% Conf. Int.
1	2048	32	5	0.984	0.003	[0.978; 0.989]
2	2011	7	3	0.981	0.003	[0.974 ; 0.986]
3	2001	5	16	0.979	0.003	[0.971; 0.984]
4	1980	3	3	0.977	0.003	[0.970; 0.983]
5	1974	3	0	0.976	0.003	[0.968; 0.981]
⋮	⋮	⋮	⋮	⋮	⋮	⋮
361	28	0	6	0.959	0.005	[0.949; 0.967]
362	22	0	7	0.959	0.005	[0.949; 0.967]
363	15	0	2	0.959	0.005	[0.949; 0.967]
364	13	0	4	0.959	0.005	[0.949; 0.967]
365	9	0	9	0.959	0.005	[0.949; 0.967]

Figure 1: Survival function estimates and 95% confidence limits.

Table 4: Cumulative hazard function estimate and 95% limits

Time	Beg. Total	Fail	Net Lost	H	Std error	95% Conf. Int.
1	2048	32	5	0.0156	0.003	[0.011; 0.022]
2	2011	7	3	0.0191	0.003	[0.014; 0.026]
3	2001	5	16	0.0215	0.003	[0.016; 0.029]
4	1980	3	3	0.023	0.003	[0.017 ; 0.031]
5	1974	3	0	0.0245	0.003	[0.019; 0.032]
⋮	⋮	⋮	⋮	⋮	⋮	⋮
361	28	0	6	0.0415	0.005	[0.034; 0.051]
362	22	0	7	0.0415	0.005	[0.034; 0.051]
363	15	0	2	0.0415	0.005	[0.034; 0.051]
364	13	0	4	0.0415	0.005	[0.034; 0.051]
365	9	0	9	0.0415	0.005	[0.034 ; 0.051]

The results show that the percentage surviving throughout the study period is nowhere less than 95.85% (CI: 94.87%-96.65%). The cumulative hazard function (Figure 2) presents increasing slopes at about the first three months of the study time and keep relatively constant slope elsewhere. This suggests that the hazard of death of infants is constant along the study time except at the first three months. This is confirmed by the plot of the hazard function (Figure 3). The confidence intervals of the hazard function are too wide since they include negative values and therefore, they are meaningless.
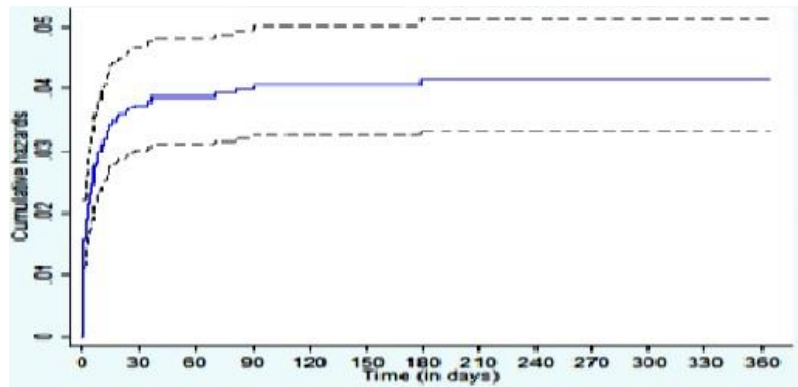
Figure 2: Cumulative hazard function estimate and 95% confidence interval

Table 5: Hazard function estimate and 95% confidence interval

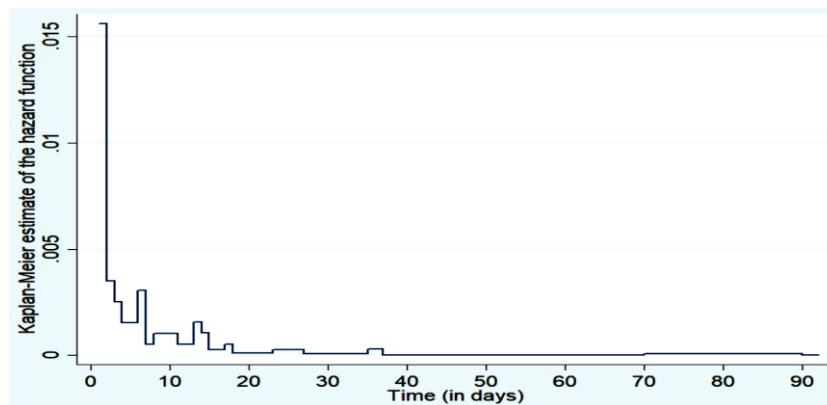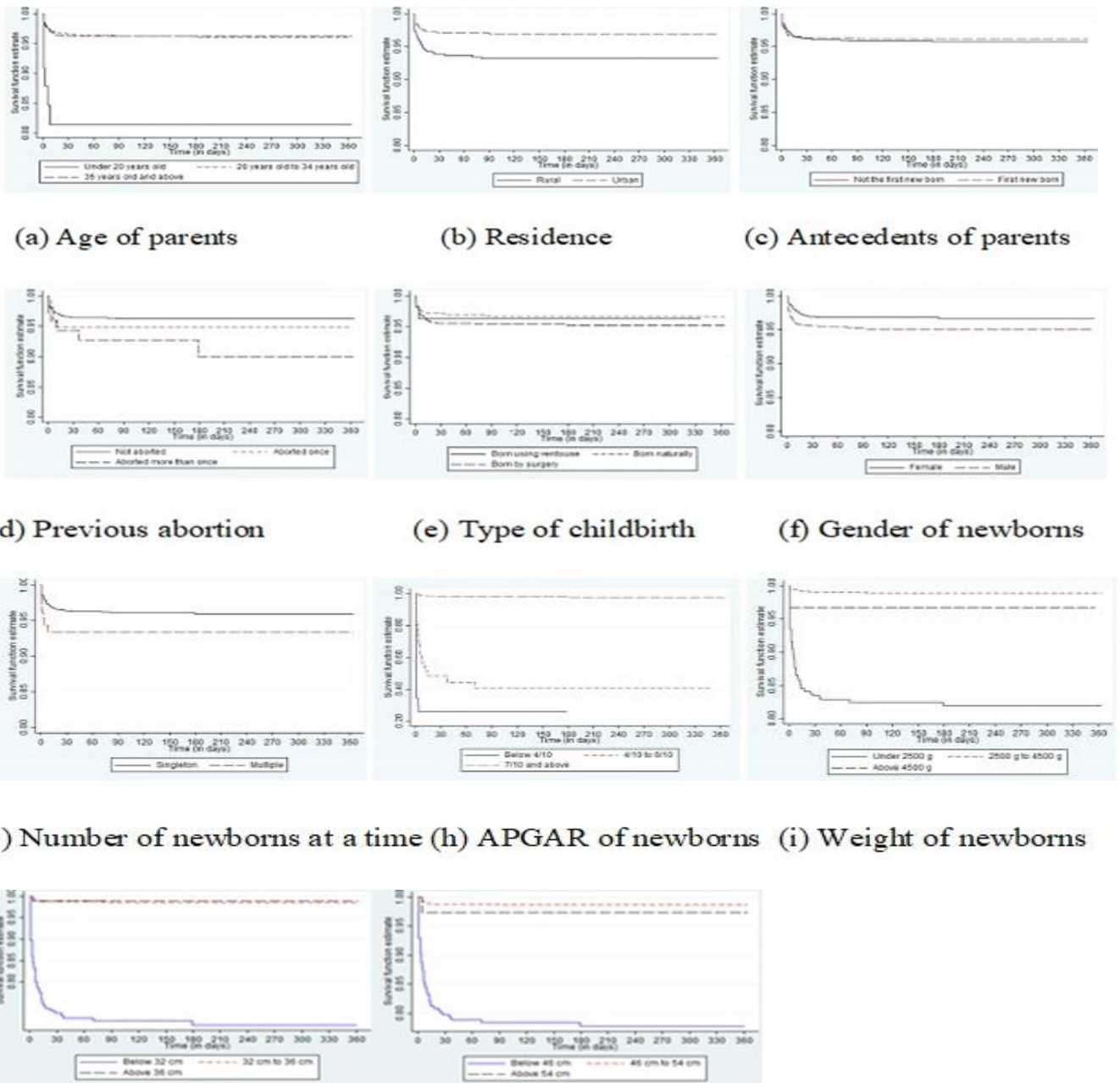| j	n_j	d_j	$\hat{h}(t)$	$se[\hat{h}(t)]$	95%CI
1	2048	32	0.021	0.003	[0.010; 0.016]
2	2011	7	0.006	0.001	[0.001; 0.003]
3	2001	5	0.005	0.001	[0.000; 0.002]
4	1980	3	0.003	0.001	N/A
⋮	⋮	⋮	⋮	⋮	⋮
21	1606	1	< 0.001	< 0.001	N/A
22	1529	1	< 0.001	< 0.001	N/A
23	1466	1	< 0.001	< 0.001	N/A
24	961	1	< 0.001	< 0.001	N/A

Figure 3: Hazard function estimate and 95% confidence limits.

Comparison of groups of survival data

The comparison among the levels of the variables is done graphically. The illustration is summarized in Figure 4. The results of the log-rank and Wilcoxon test statistics are summarized in Table 6. The log-rank test for comparison is suitable for comparing levels of variables *residence*, *gender*, *number*, *APGAR* and, *weight* where the plots do not cross. Wilcoxon test is suitable in comparing the levels of the rest of variables since theirs plots cross.

(a) Age of parents   (b) Residence   (c) Antecedents of parents

(d) Previous abortion   (e) Type of childbirth   (f) Gender of newborns

(g) Number of newborns at a time (h) APGAR of newborns (i) Weight of newborns

(j) Heah circumference of newborns   (k) Height of newborns

Figure 4: Plots of the Kaplan-Meier estimates of the survival function for variables (a) *age,* (b) *residence,* (c) *antecedents,* (d) *abortion,* (e) *childbirth,* (f) *gender,* (g) *number,* (h) *APGAR,* (i) *weight* (j) *head* and (k) *height* for dataset on newborns at KUTH, year 2016.

Figure 4 (a) shows that babies whose parents are 20 years old to 34 years and above 34 years survive better than babies whose parents are under 20 years old. The Wilcoxon test of no difference shows strong evidence against the no difference between these categories of age.

Figure 4(j) shows that babies whose circumference of head is 32cm and above survive better than those with circumference of head below 32cm, with

overwhelming evidence against the non-difference as shows to the Wilcoxon test. Figure 4 (k) shows that babies with normal height (46-54cm) survive better than stunted and over-height babies. The Wilcoxon test of no difference shows overwhelming evidence against the no difference between the levels of height. Wilcoxon test shows that there is no evidence of any difference between levels of variables antecedents, abortion and childbirth.

Figure 4 (b) shows that urban babies survive better than rural babies; the log-rank test for no difference confirms overwhelming no difference. Figure 4 (f) shows that the female babies survive better that males but the log-rank test confirm slight evidence against no difference. Singleton survive better than multiple births (Figure 4 (g)) but the log-rank test shows that there is no evidence of no difference in levels number of newborn at time. Figure 4 (h) shows that babies with APGAR greater or equal to 7/10 survive better than babies whose APGAR is from 4/10 to 6/10 and much better than babies whose APGAR is less than 4/10. Log-rank test show overwhelming evidence against the no difference between the levels of APGAR.

Table 6: Log-rank and Wilcoxon test statistics.

Variable	Log-rank χ^2 test statistic (p-value)	Wilcoxon χ^2 test statistic (p-value)
Age	11.84 (0.003)	12.44 (0.002)
Residence	13.74 ($p < 0.001$)	13.79 ($p < 0.001$)
Antecedents	0.10 (0.752)	0.06 (0.812)
Abortion	4.48 (0.107)	3.39 (0.183)
Childbirth	2.14 (0.343)	2.07 (0.355)
Gender	3.45 (0.063)	3.69 (0.055)
Number	1.93 (0.165)	2.23 (0.135)
APGAR	912.49 ($p < 0.001$)	919.37 ($p < 0.001$)
Weight	219.90 ($p < 0.001$)	219.32 ($p < 0.001$)
Head	382.38 ($p < 0.001$)	376.31 ($p < 0.001$)
Height	262.69 ($p < 0.001$)	259.03 ($p < 0.001$)

CPHM for the infant mortality at KUTH

Table 7 presents the estimates of the hazard ratios using the Cox proportional hazard model (1). For handling ties, Breslow, Efron and Cox approaches give similar results and thus those presented in Table 7 are from the default (Breslow).

Table 7: CPHM for all covariates

Covariate (reference)	Level	Haz. Ratio	Std. Err.	z	$P > z$	95% Conf. Int
Age (Under 20 years old)	20 to 34 years old	0.216	0.104	-3.190	0.001	[0.084; 0.554]
	35 years old and above	0.279	0.147	-2.420	0.015	[0.099; 0.784]
Residence (Rural)	Urban	1.026	0.246	0.110	0.914	[0.642; 1.640]
Antecedents (Not 1st newborn)	1st new born	0.841	0.236	-0.620	0.536	[0.485; 1.457]
Abortion (Not aborted)	Aborted once	1.670	0.659	1.300	0.194	[0.771; 3.619]
	Aborted more than once	1.171	0.531	0.350	0.728	[0.481; 2.850]
Childbirth (Ventouse)	Natural	0.621	0.471	-0.630	0.530	[0.141; 2.745]
	Surgery	0.779	0.584	-0.330	0.739	[0.180; 3.383]
Gender (Female)	Male	1.852	0.443	2.580	0.010	[1.159; 2.960]
Number (Singleton)	Multiple	0.324	0.143	-2.550	0.011	[0.137; 0.770]
APGAR (Below 4/10)	4/10 to 6/10	0.387	0.149	-2.470	0.014	[0.182; 0.822]
	7/10 and above	0.056	0.020	-8.050	$p < 0.001$	[0.028; 0.113]
Weight (Under 2500 g)	2500 g to 4500 g	0.219	0.087	-3.810	$p < 0.001$	[0.101; 0.479]
	Above 4500 g	0.390	0.418	-0.880	0.379	[0.048; 3.187]
Head (Below 32 cm)	32 cm to 36 cm	0.287	0.111	-3.230	0.001	[0.134; 0.611]
	Above 36 cm	0.125	0.132	-1.980	0.048	[0.016; 0.980]
Height (Below 46 cm)	46 cm to 54 cm	0.559	0.234	-1.390	0.165	[0.246; 1.270]
	Above 54 cm	1.033	1.114	0.030	0.976	[0.125; 8.550]

The results in Table 7 indicate significant differences in levels of covariates age, gender, number, APGAR, weight and head where p-values are less or equal to 0.05. The model suggests that the hazard of death of babies whose parents are from 20 years and 34 years old is 0.216 (95% CI:0.084−0.554, p = 0.001) times that of babies whose parents are under 20 years old. The hazard of death of babies whose parents are 35 years old and above is 0.279 (95% CI:0.099−0.784, p = 0.015) times that of babies whose parents are under 20 years old. The hazard of death for male babies is 1.852 (95% CI:1.159−2.960, p = 0.010) times that of female babies. The hazard of death of multiple babies is 0.324 (95% CI:0.137−0.770, p = 0.011) times that of singleton babies. The hazard of death for babies whose APGAR range from 4/10 to

6/10 is 0.387 (95% CI:0.182−0.822, p = 0.014) times that of babies whose APGAR is below 4/10. The hazard of death for babies whose APGAR range from 7/10 to 10/10 is 0.056 (95% CI:0.028−0.113, p < 0.001) times that of babies whose APGAR is below 4/10. The hazard of death for babies whose weight range from 2500 g to 4500 g is 0.219 (95% CI:0.101−0.479, p < 0.001) times that of babies whose weight is below 2500 g. The hazard of death for babies whose circumference of head range from 32 cm to 36 cm is 0.287 (95% CI:0.134−0.611, p = 0.001) times that of babies whose circumference of head is below 32 cm. The hazard of death for babies whose circumference of head is above 36 cm is 0.125 (95% CI:0.016−0.980, p = 0.048) times that of babies whose circumference of head is below 32 cm.

Aalen additive hazards model

Unlike the CPHM based on quantitative measurement of hazard ratio, the cumulative parameter functions express the hazard by considering the slopes of the plots of cumulative parameter functions. The plots of the estimates of the cumulative parameter functions $B_k(t)$ and associated confidence intervals are presented, giving information about significance of the parameters $\beta_k(t)$. The analysis is done in STATA using the command *stlh* suggested by Hosmer and Royston (2003).

Figure 5 gives the plots of the cumulative parameter functions and their 95% confidence limits for the variables residence, gender, antecedents, APGAR and number. Figure 5 (a) represents the estimated cumulative parameter function with its 95% confidence limits for the urban parents compared to rural parents. The plot is approximately horizontal and negatives everywhere but the upper and lower limits of the confidence interval are on either side of the zero line.
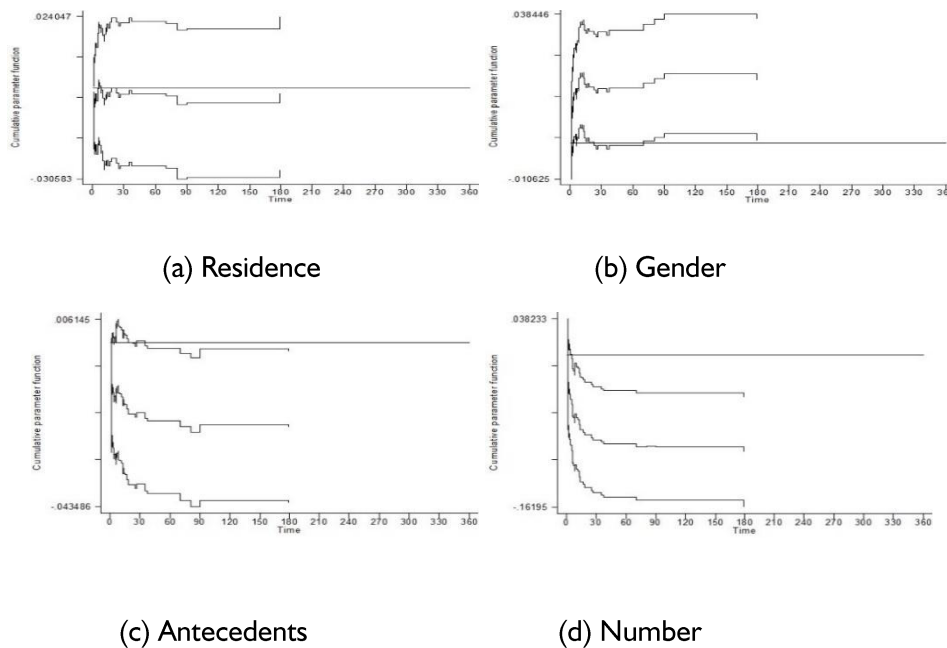

(a) Residence

(b) Gender

(c) Antecedents

(d) Number

Figure 5: Cumulative parameter function plots for the KUTH data for variables residence, gender, antecedents and number

This indicates that the hazard of death for urban babies may be slightly higher than that of rural babies, but the difference may be not significant.

Figure 5 (b) represents the estimated cumulative parameter function with its 95% confidence limits for the male compared to female babies. The plot is approximately horizontal and positive everywhere with the upper and lower limits of the confidence interval situated approximately above the zero line. This indicates that the hazard of death for male babies may be slightly higher than that of female babies.

Figure 5 (c) represents the estimated cumulative parameter function with its 95% confidence limits for the first new born compared to babies that are not. The plot decreases below the zero line and becomes horizontal towards the end of study time, with the confidence limits at either sides of the zero line. This indicates that the hazard of death for babies that are not first newborn may be higher than that of first newborn.

Figure 5 (d) represents the estimated cumulative parameter function with its 95% confidence limits for the multiple newborns compared to the singletons.

The plot decreases below the zero line and becomes horizontal towards the end of study time, with the confidence limits approximately below the zero line.

This indicates that the hazard of death for singletons may be higher than that of multiple newborns

.

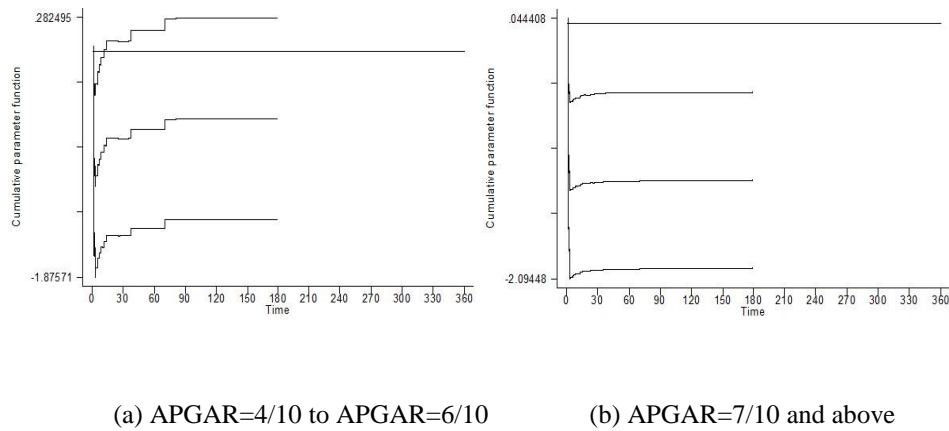(a) APGAR=4/10 to APGAR=6/10   (b) APGAR=7/10 and above

Figure 6: Cumulative parameter function plots for variable APGAR.

Figure 6 gives the plots of the cumulative parameter functions and their 95% confidence limits for levels of the variable APGAR Figure 6 (a) represents the estimated cumulative parameter function with its 95% confidence limits for the newborn's APGAR from 4/10 to 6/10 compared to the APGAR < 4/10. The plot increases below the zero line for the first 75 days and then become horizontal below the zero line with the major part of confidence intervals situated below the zero line. This indicates that the hazard of death for newborns with APGAR below 4/10 may be higher than that of newborns with APGAR from 4/10 to 6/10. Figure 6 (b) represents the estimated cumulative parameter function with its 95% confidence limits for the newborn's APGAR that is 7/10 and above compared to the newborn's APGAR below 4/10. The plot is horizontal below the zero line, with confidence limits below the zero line. This indicates that the hazard of death for newborns with

APGAR below 4/10 may be constant and higher than that of newborns whose APGAR is 7/10 and above.

Figure 7 gives the plots of the cumulative parameter functions and their 95% confidence limits for levels of the variable age. Figure 7 (a) displays the cumulative parameter function for parents with 20 years old to 34 years old with its 95% confidence limits for the variable Age with reference taken on under 20 years old. The plot is below the horizontal line zero and the slope decreases only during the first month of study time but a large portion of the lower limit of its 95% confidence interval is below the zero line. This is an indication that the hazard of death for newborns whose parent is under 20 years old is higher than that of newborns from parents with 20 years old to 34 years old. Similar situation is observed for newborns whose parents are 35 years old and above as indicates Figure 7 (b).
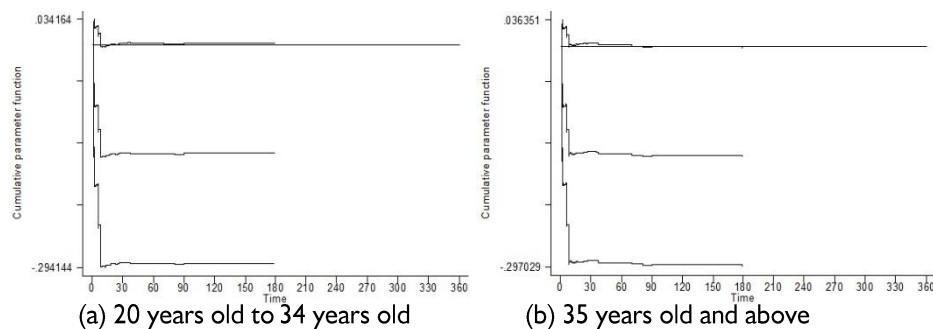

(a) 20 years old to 34 years old   (b) 35 years old and above

Figure 7: Cumulative parameter function plots for variable age.

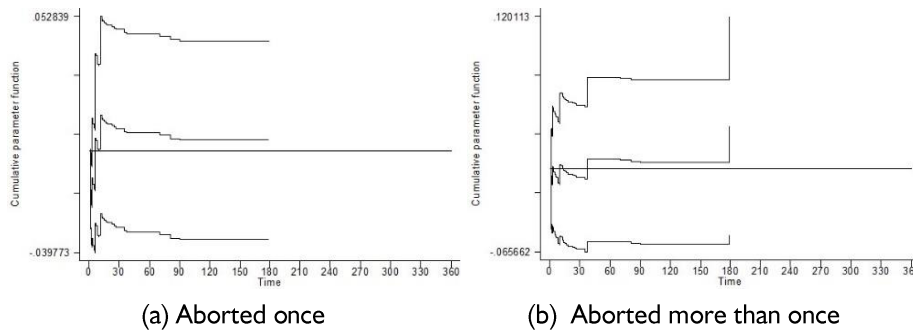(a) Aborted once　　　　　　(b) Aborted more than once

Figure 8: Cumulative parameter function plots for variable abortion.

Figure 8 gives the plots of the cumulative parameter functions and their 95% confidence limits for levels of the variable abortion. Figure 8 (a) displays the cumulative parameter function with its 95% confidence limits for newborns whose parents aborted once. The plot Figure 8 (b) fluctuates also around the zero line. This indicates that the differences among the levels of abortion are not significant.

Figure 9 gives the plots of the cumulative parameter functions and their 95% confidence limits

for levels of the variable childbirth. Both Figure 9 (a) and 9 (b) behave similarly: the plots are approximately horizontal and negative everywhere, but the upper and lower limits of the confidence interval are on either sides of the zero line. This indicates that the hazard of death of newborns by ventouse is higher than that of newborns naturally and by surgery.
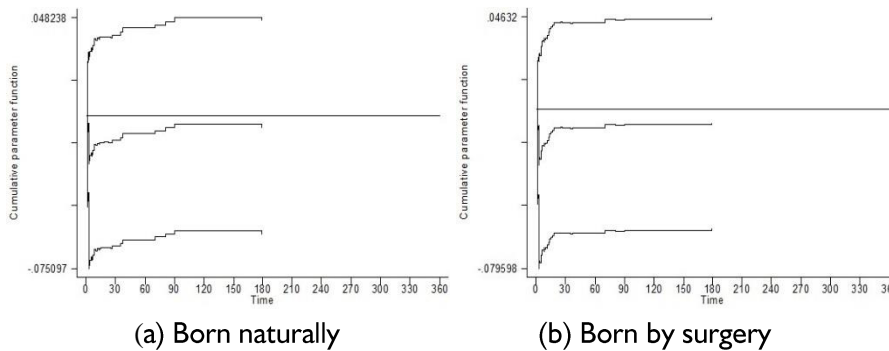

(a) Born naturally　　　　　　(b) Born by surgery

Figure 9: Cumulative parameter function plots for variable childbirth.

Figure 10 gives the plots of the cumulative parameter functions and their 95% confidence limits for levels of the variable head. Both Figure 10 (a) and 10 (b) behave similarly: the plots are decreasing below the zero line with the upper and lower limits of the confidence interval below the zero line. This indicates

that the hazard of death of newborns with circumference of head less than 32 cm is higher than that of newborns with normal circumference of the head and that of newborns with extra-normal circumference of the head.
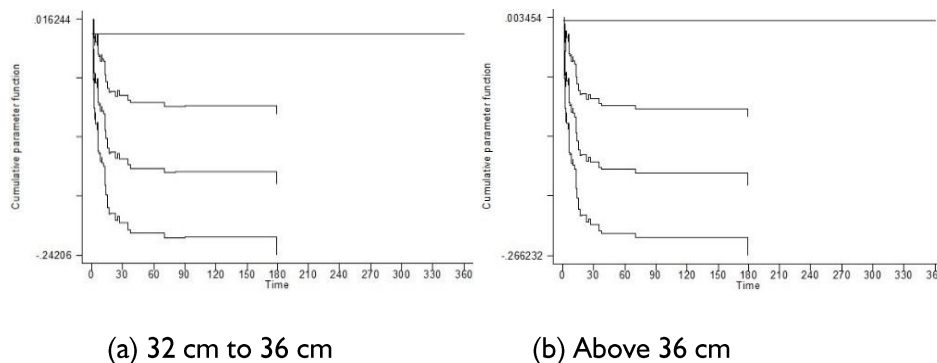

(a) 32 cm to 36 cm　　　　　　(b) Above 36 cm

Figure 10: Cumulative parameter for variable head.

Figure 11 gives the plots of the cumulative parameter functions and their 95% confidence limits for levels of the variable height. Both Figure 11 (a) and 11 (b) behave similarly apart from the upper limit of the confidence interval of at the plot 11 (b) situated above the zero line: the plots are decreasing below the zero line. This indicates that the hazard of death of newborns with under-height is higher than that of newborns with normal height and newborns with over-height.
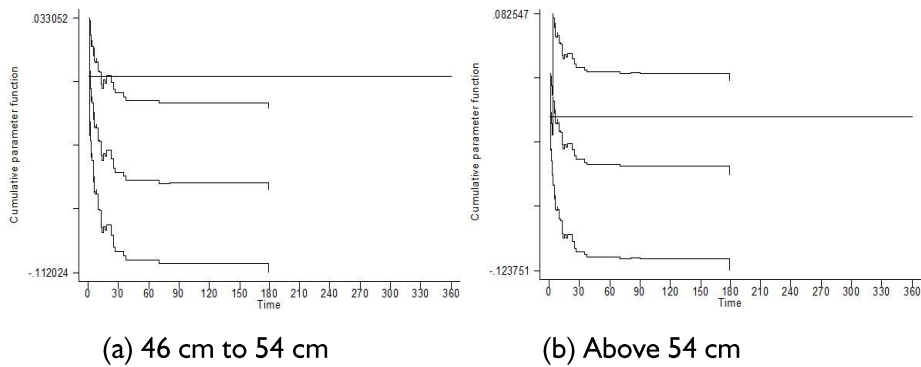

(a) 46 cm to 54 cm　　　　(b) Above 54 cm

Figure 11: Cumulative parameter function for variable height

Figure 12 gives the plots of the cumulative parameter functions and their 95% confidence limits for levels of the variable weight. Both Figure 12 (a) and 12 (b) behave similarly: the plots are decreasing below the zero line with the major part of the confidence interval below the zero line. This indicates that the hazard of death of newborns with underweight is higher than that of both newborns with normal weight and newborns with overweight.
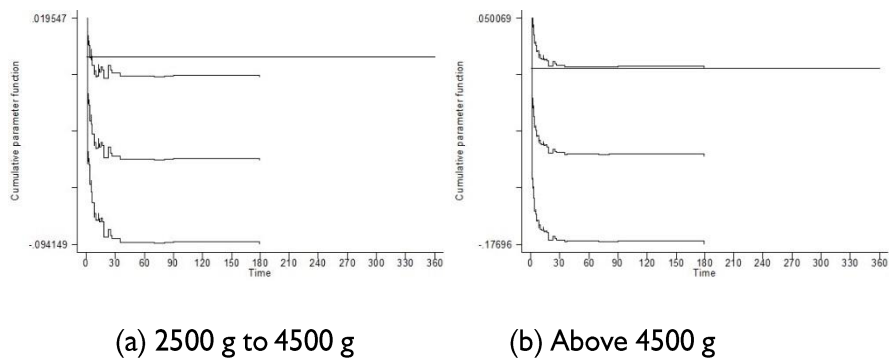

(a) 2500 g to 4500 g　　　　(b) Above 4500 g

Figure 12: Cumulative parameter function plots for variable weight.

Table 8: Tests for significance of covariates.

(a) Test 1: weights equal to 1.0.

Covariate (reference)	Level	z	P
Age (Under 20 years old)	20 years old to 34 years old	-1.939	0.053
	35 years old and above	-1.992	0.046
Residence (Rural)	Urban	-0.137	0.891
Antecedendents (Not first new born)	first newborn	-2.190	0.028
Abortion (not aborted)	aborted once	0.224	0.823
	Aborted more than once	0.771	0.440
Childbirth (born using entouse)	Born naturally	-0.226	0.821
	Born by surgery	-0.246	0.805
Gender (Female)	Male	2.037	0.042
Number (Single)	Multiple	-3.488	$p < 0.001$
APGAR (under 4/10)	4/10 to 6/10	-1.299	0.194
	7/10 and above	-3.452	0.001
Weight (Under 2500 g)	2500 g to 4500g	-2.438	0.015
	Above 4500 g	-1.981	0.048
Head (below 32 cm)	32 cm to 36 cm	-4.192	$p < 0.001$
	Above 36 cm	-4.686	$p < 0.001$
Height (Below 46 cm)	46 cm to 54 cm	-2.752	0.006
	Above 54 cm	-1.227	0.220

(b) Test 2: weights equal to the size of the risk set.

Covariate (reference)	Level	z	P
Age (Under 20 years old)	20 years old to 34 years old	-1.943	0.052
	35 years old and above	-1.971	0.049
Residence (Rural)	Urban	-0.233	0.816
Antecedendents (Not first new born)	first newborn	-2.174	0.030
Abortion (not aborted)	aborted once	0.230	0.818
	Aborted more than once	0.510	0.610
Childbirth (born using entouse)	Born naturally	-0.256	0.798
	Born by surgery	-0.298	0.766
Gender (Female)	Male	2.122	0.034
Number (Single)	Multiple	-3.365	0.001
APGAR (under 4/10)	4/10 to 6/10	-1.437	0.151
	7/10 and above	-3.500	$p < 0.001$
Weight (Under 2500 g)	2500 g to 4500g	-2.418	0.016
	Above 4500 g	-1.937	0.053
Head (below 32 cm)	32 cm to 36 cm	-4.199	$p < 0.001$
	Above 36 cm	-4.730	$p < 0.001$
Height (Below 46 cm)	46 cm to 54 cm	-2.638	0.008
	Above 54 cm	-1.098	0.272

(c) Test 3: weights equal to Kaplan-Meier (KM) estimator.

Covariate (reference)	Level	z	P
Age (Under 20 years old)	20 years old to 34 years old	-1.941	0.052
	35 years old and above	-1.992	0.046
Residence (Rural)	Urban	-0.142	0.887
Antecedendents (Not first new born)	First newborn	-2.189	0.029
Abortion (Not aborted)	Aborted once	0.215	0.829
	Aborted more than once	0.762	0.466
Childbirth (Born using ventouse)	Born naturally	-0.234	0.815
	Born by surgery	-0.254	0.799
Gender (Female)	Male	2.046	0.041
Number (Single)	Multiple	-3.472	0.001
APGAR (under 4/10)	4/10 to 6/10	-1.325	0.185
	7/10 and above	-3.459	0.001
Weight (Under 2500 g)	2500 g to 4500g	-2.431	0.015
	Above 4500 g	-1.970	0.049
Head (below 32 cm)	32 cm to 36 cm	-4.191	$p < 0.001$
	Above 36 cm	-4.688	$p < 0.001$
Height (Below 46 cm)	46 cm to 54 cm	-2.738	0.006
	Above 54 cm	-1.213	0.225

(d) Test 4: weights equal to $KM/\mathrm{se}\left(\hat{\beta}(t)\right)$.

Covariate (reference)	Level	z	P
Age (Under 20 years old)	20 years old to 34 years old	1.988	0.047
	35 years old and above	-1.768	0.077
Residence (Rural)	Urban	0.884	0.376
Antecedendents (Not first new born)	first newborn	-2.855	0.004
Abortion (not aborted)	aborted once	-5.086	$p < 0.001$
	Aborted more than once	-6.188	$p < 0.001$
Childbirth (born using entouse)	Born naturally	3.089	0.002
	Born by surgery	3.079	0.002
Gender (Female)	Male	2.003	0.045
Number (Single)	Multiple	-6.177	$p < 0.001$
APGAR (under 4/10)	4/10 to 6/10	-1.790	0.073
	7/10 and above	1.508	0.131
Weight (Under 2500 g)	2500 g to 4500g	-2.643	0.008
	Above 4500 g	-5.309	$p < 0.001$
Head (below 32 cm)	32 cm to 36 cm	-3.754	$p < 0.001$
	Above 36 cm	-4.855	$p < 0.001$
Height (Below 46 cm)	46 cm to 54 cm	-3.750	$p < 0.001$
	Above 54 cm	-3.964	$p < 0.001$

Cox-Aalen hazards model (CAHM)

Multiplicative part of the CAHM (Table 8) shows significance on covariates age number and weight where the results are not far from that found for the CPHM, and covariates APGAR where the CAHM present a huge difference in levels. The CAHM suggests that the hazards of death of babies with APGAR less than 4/10 is 16.39 times that of babies with APGAR 4/10 to 6/10 (p-value< 0.001) and 166.7 times that of babies with APGAR greater than 6/10 (p-value< 0.001). Figure 13 summarizes the additive part of the CAHM. The interpretation is not far from that of AAHM (Section 3.4).

Table 9: Multiplicative part of the CAHM

| Covariate (Reference)	Level	Coef	Se	95% CI of Coef.	HR	z	$P > z$	95% CI of HR
Age (Under 20 years old)	20 to 34 years old	-1.910	0.411	[-2.720; -1.100]	0.148	-6.250	$p < 0.001$	[0.066; 0.333]
	35 years old and above	-1.630	0.436	[-2.480; -0.775]	0.196	-4.570	$p < 0.001$	[0.084; 0.461]
Residence (Rural)	Urban	-0.231	0.195	[-0.613; 0.151]	0.794	-1.210	0.228	[0.542; 1.163]
Abortion (Not aborted)	Aborted once	0.185	0.367	[-0.534; 0.904]	1.203	0.589	0.556	[0.586; 2.469]
	Aborted more than once	0.155	0.403	[-0.635; 0.945]	1.168	0.281	0.778	[0.530; 2.573]
Gender (Female)	Male	0.110	0.195	[-0.272; 0.492]	1.116	0.580	0.562	[0.762; 1.636]
Number (Singleton)	Multiple	-1.340	0.363	[-2.050, -0.629]	0.262	-4.150	$p < 0.001$	[0.129; 0.533]
APGAR (Below 4/10)	4/10 to 6/10	-2.800	0.325	[-3.440; -2.160]	0.061	-8.400	$p < 0.001$	[0.032; 0.115]
	7/10 and above	-5.120	0.357	[-5.820; -4.420]	0.006	-15.800	$p < 0.001$	[0.003; 0.012]
Weight (Under 2500 g)	2500 g to 4500 g	-1.320	0.301	[-1.910; -0.730]	0.267	-5.020	$p < 0.001$	[0.148; 0.482]
	Above 4500 g	-1.300	1.130	[-3.510; 0.915]	0.273	-1.080	0.281	[0.030; 2.497]
Head (Below 32 cm)	32 cm to 36 cm	0.077	0.356	[-0.621; 0.774]	1.080	0.241	0.809	[0.537; 2.168]
	Above 36 cm	-0.264	0.638	[-1.500; 1.000]	0.768	-0.421	0.674	[0.223; 2.718]
Height (Below 46 cm)	46 cm to 54 cm	-0.300	0.302	[-0.892; 0.292]	0.741	-1.080	0.279	[0.410; 1.339]
	Above 54 cm	0.364	0.652	[-0.914; 1.640]	1.439	0.587	0.557	[0.401; 5.155]
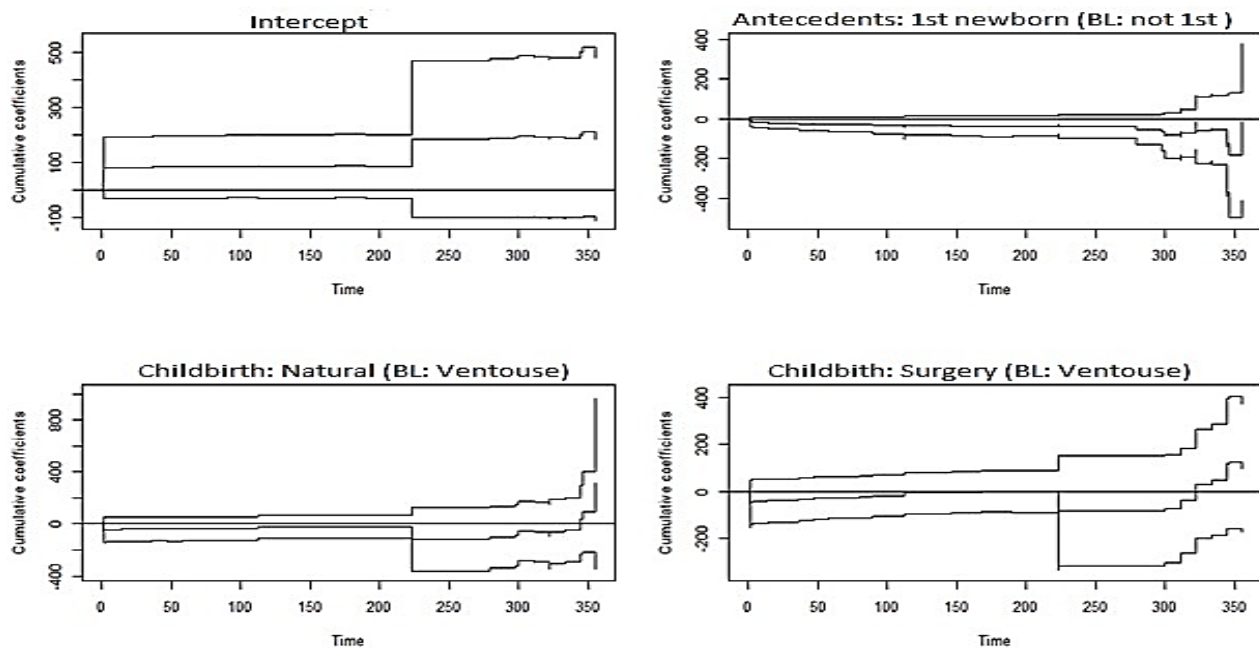
Figure 13: Additive part of the CAHM.

Discussion

The Aalen Additive Hazard Model (AAHM) for the infant mortality at KUTH by type 1 weights test, type 2 weights test and type 3 weights test show significant difference between the lowest and highest levels of variables APGAR ($p \leq 0.001$), while all types of weight test show significant difference between levels of variables head ($p < 0.001$), gender and weight. The significant difference for all levels of the variable height is observed in type 4 weight test ($p < 0.001$). The difference between under-height and normal height is significant by type 1 weight test ($p = 0.006$), type 2 weight test ($p = 0.008$), and type 3 weight test ($p = 0.006$). Type 4 weight test suggests significant difference between age of under 20 years old and age ranging from 20 years old and 34 years old ($p < 0.047$) and all levels of variable height ($p < 0.001$). Type 4 weight test suggests also significance difference between all levels of variable abortion ($p < 0.001$) and antecedents ($p = 0.004$) and suggest significant difference between all levels of childbirth ($p = 0.002$). By combining all the four AAHM tests, all the covariates are significantly included in the model, unlike the CPHM in which covariates *residence, antecedents, abortion, childbirth,* and *height* are excluded while the CAHM excludes covariates *residence, abortion, gender, head* and *height.* The overall study at KUTH emphasizes a relatively higher risk for babies whose parents are under 20 years old, babies with relatively lower APGAR score and stunted babies mainly observed in low-income countries especially in SSA. This results from different problematic determinants suggested

by many researchers such as maternal mortality, lack of sanitation infrastructures, limited female education, and lack of access to the household basic energy and water in SSA.

Conclusion

This paper reviewed non-parametric methods of the survival analysis, namely the Kaplan- Meier method for estimating and graphing survival and the hazard function, the Cox proportional hazards model (CPHM) the Aalen additive hazards model (AAHM) and the Cox-Aalen hazards model. These methods are used to analyse the dataset collected at Kigali University Teaching hospital for 2117 newborns during 365 days of the year 2016.

The results revealed that the hazard of death of infant is higher in male babies as compared to female babies; it is higher for babies whose parents are under 20 years' old parents as compared to older parents. Babies born with APGAR greater or equal to 7/10 were found to have a better survival outcome than those born with APGAR less than 7/10. Babies with normal weight and overweight were found to have a lower hazard of death compared to underweight babies. Babies with normal circumference of head were found to survive better than those with relatively big head and relatively small head. Under-height babies were found to have a higher hazard of death, as compared to babies born with normal height and over-height newborns. Finally, babies born naturally were found to survive better than those born using ventouse or those born after surgery. For the CPHM, the results were

significant only for variables age, gender, number, APGAR, weight and head. The results of AAHM are significant for all variables except variable residence while for CAHM, the significance is found on covariates age, number, APGAR and weight. The results on variable height are surprisingly not significant by CPHM and by CAHM unlike expected results. Significance on variable height was rather observed in AAHM in accordance with related tests, especially test 4.

The study pointed relatively higher risk to death of infants whose female parents are under 20 years old; pregnancy of such parents should be avoided. Also stunting or abnormality of infant lead to relatively higher risk to mortality, clinically recommended nutrition during pregnancy would decrease abnormality and stunting of newborns which contribute to rise of infant mortality rate.

Analysis was limited to only 11 variables. Unavailable variables concerning parents that could improve models are for example, demographic variables such as education level, employment and income, behavioral variables such as smoking habit, alcohol consumption and dietary and physio-therapeutic variables such as sport activity level. These variables are not recorded in registry at KUTH.

The future work will consist of running multiple events model which would provide a much more informative model where two events will be death and incidence of chronic disease. Also, the suitability of the dataset to parametric survival model would be checked and then the suitable parametric model could be fitted.

Acknowledgements

References

Aalen, O. O. (1989). A linear regression model for the analysis of lifetimes.Statistics in Medecine, 8, 907-925.

Aalen, O. O., Borgan, Ø., and Gjessing, H. K. (2008). Survival and event history analysis. Springer, New York.

Adetunji, J., and Bos, E. (2006). Disease and Mortality in Sub-Saharan Africa, 2nd edition. The International Bank for Reconstruction and Development/ The World Bank. URL http://www.ncbi.nlm.nih.gov/books/NBK2292/

Akunga, D., Menya, D. and Kabue, M. (2014). Determinants of postnatal care use in Kenya. African Population Studies, 28 (30, 1447-1459.

Benn Sartorius, K. D., and Kurt Sartorius (2014). Global infant mortality trends and attributable determinants, an ecological study using data from 192 countries for the period 1990-2011. Population Health Metrics 2014, 12, 29. URL http://www.pophealthmetrics.com/content/12/1/29.

Bourgeois, J. (1946). De la mesure de la mortalité infantile. Population, 1 (1), 53-68. URL http://www.persee.fr/doc/pop_0032-4663_1946_num_1_1_1646.

Breslow, N. E. (1974). Covariance analysis of censored survival data. Biometrics, 30, 89- 99.

Collet, D. (2003). Modeling survival data in medical research, 2nd edition. Chapman & Hall, London.

Cox, D. R. (1972). Regression models and life-tables (with discussion). Journal of the Royal Statistical Society, Series B, 34 (2), 187-220.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. Journal of the American Statistical Association, 72, 557-565.

Ester, P. V., Torres, A., Freire, J. M., Hernández, V., and Gil, A. (2011). Factors associated to infant mortality in Sub-Saharan Africa. Journal of Public Health in Africa, 2 (e27), 112-116.

Flemming, T. R., and Harrington, D. P. (2005). Counting process and survival analysis. John Wiley & Sons, Inc., Haboken.

Gaston, R. T., Ramroop, S. and Habyarimana, F. (2018). Determinants of factors associated with anemia among children under five years in Lesotho. African Population Studies, 32 (1), 3893-3902.

Gourbin, C. (2005). Foetal mortality, infant mortality, and age of parents. Rev Epidemiol Sante Publique, 53 (2), 81-86.

Hosmer, D., and Royston, P. (2002). Using Aalen's linear hazards model to investigate time-varying effects in the proportional hazards regression model. The STATA Journal, 2 (4), 331-350.

Hosmer, D. W., Lemeshow, S., and May, S. (2008). Regression modeling of time-to- event data, 2nd edition. John Wiley & Sons Inc., Haboken.

Janssen, P., Thiessen, P., Klein, M., Whitfield, M., Ying, C., and Cullis-Kuhl, S. (2007). Standards for

the measurement of birth weight, length and head circumference at term in neonates of European, Chinese and south Asian ancestry. Open Med, 1 (2), 74-88.

Klein, J. P., and Moeschberger, M. L. (2003). Techniques for censored and truncated data, 2nd edition. Springer-Verlag, New York.

Mturi, A., and Curtis, S. (1995). The determinants of infants and child mortality in Tanzania. Health Policy Plan, 10 (4), 384-394.

Mulaik, S. A. (2009). Foundations of factor analysis, 2nd edition. Chapman Hall/ CRC, Boca Raton.

Rathavuth, H., Mohamed, A., Shea, R., and Ruilin, R. (2009). Childhood Mortality in Rwanda: Levels, trends, and differentials, ICF Macro Calverton, Maryland, USA.

Reidpath, D. D., and Allotey, P. (2003). Infant mortality rate as an indicator of population health. J Epidemiol Community Health, 57, 344-346.

Sartorius, K., Sartorius, K., Chirwa, T., and Fonn, S. (2011). Infant mortality in South Africa-distribution, associations and policy implications, 2007: an ecological spatial analysis. Int J Health Geogr 2011, 10, 61. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC32 50938/

Scheike, T., and Zhang, M. (2002). An additive-multiplicative Cox-Aalen regression model. Scandinavian Journal of Statistics, 29, 75-88.

Schell, C., Reilly, M., Rosling, H., Peterson, S., and Ekstrom, A. (2007). Socioeconomic determinants of infant mortality: a worldwide study of 152 low, middle and high-income countries. Scandinavian Journal of Public Health, 35 (3), 288-297.