UNIVERSITY OF KWAZULU-NATAL

Some Statistical Methods in Analysis of Single and Multiple Events with Application to Infant Mortality Data

by

Paul Gatabazi

2020

Some Statistical Methods in Analysis of Single and Multiple Events with Application to Infant Mortality Data

by

Paul Gatabazi

Submitted in fulfillment of the academic requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Applied Statistics

in the

School of Mathematics, Statistics and Computer Science

University of KwaZulu-Natal, South Africa

Pietermaritzburg

2020

UNIVERSITY OF KWAZULU-NATAL

INYUVESI YAKWAZULU-NATALI

Dedication

This dissertation is dedicated to my beloved late father Gratien Gatabazi, my mother Flora Uwera, my brothers and sisters, my aunt Goretti Mukakamali, my wife Josiane and my sons Luc and Louis-Marie.

Declaration

The research work described in this thesis was carried out in the School of Mathematics, Statistics and Computer Sciences, University of KwaZulu-Natal, Pietermaritzburg, under the supervision of Dr. Sileshi Fanta Melesse and Prof. Shaun Ramroop.

I, Paul Gatabazi, declare that this thesis is my own, unaided work. It has not been submitted in any form for any degree or diploma to any other University. Where use has been made of the work of others, it is duly acknowledged.

Paul Gatabazi — 16 August 2020

Dr. Sileshi Fanta Melesse — 17/08/2020

Prof. Shaun Ramroop — 17/08/20

Date

List of Papers and Conferences

The following papers have been published from this thesis.

1. Gatabazi, P., Melesse, S.F and Ramroop S. (2018). Multiple Events Model for the Infant Mortality at Kigali University Teaching Hospital. *The Open Public Health Journal*, **11**, 464-473.

2. Gatabazi, P., Melesse, S.F and Ramroop S. (2019). Infant mortality at the Kigali University Teaching Hospital: Application of Aalen additive hazards model and comparison with other classical survival models. *African Population Studies*, **33** (2), 4834-4851.

3. Gatabazi, P., Melesse, S.F and Ramroop S. (2019). Resampled Cox Proportional Hazard Model for the Infant Mortality at the Kigali University Teaching Hospital. *The Open Public Health Journal*, **12**, 136-144.

4. Gatabazi, P., Melesse, S.F and Ramroop S. (2020). Comparison of three classes of Marginal Risk Set Model in predicting infant mortality among newborn babies at Kigali University Teaching Hospital, Rwanda, 2016. *BMC Pediatrics*, **20**, 62.

5. Gatabazi, P., Melesse, S.F and Ramroop S. (2019). Discussion on confidence interval length in Cox model for the analysis of the risk to infant mortality. *International Journal of Public Health*, **Under review**.

Parts of this thesis have been presented at the following international conferences:

1. Joint Conference of the Sub-Saharan Network (SUSAN) of the International Biometrics Society (IBS) and DELTAS Africa Sub-Saharan Africa Consortium for Advanced Bio-

statistics (SSACAB). Cape Town, South Africa, 8th -11th September 2019: Re-sampled survival models for the Infant Mortality Data from the Kigali University Teaching Hospital.

2. The 2019 Developing Excellence in Leadership, Training And Science in Africa (DELTAS Africa) scientific conference. Dakar, Senegal, 15th -18th July 2019: Single and Multiple Events Analysis with Application to Infant Mortality Data.

3. 2018 Post Graduate Research and Innovation Symposium. University of Kwazulu-Natal, South Africa, 25 October 2018: Multiple events Model for the infant mortality at Kigali University Teaching Hospital.

4. Biostatistics for Evidence-Based Policy Decisions in Health and Achievements of Sustainable Development Goals. Nairobi, Kenya, 10th -12th September 2018: Multiple events Model for the infant mortality at Kigali University Teaching Hospital.

5. The 2nd Developing Excellence in Leadership Training And Science (DELTAS- Africa) Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB) Annual general meeting, Windhoek, Namibia, 1st-3rd November, 2017: Single and Multiple events Model for the infant mortality at Kigali University Teaching Hospital.

6. 2017 Post Graduate Research and Innovation Symposium. University of Kwazulu-Natal, South Africa, 25 October 2017: Modeling infant mortality at Kigali University Teaching Hospital using Aalen Additive Hazards Model.

7. International Biometric Society Susan Conference 2017, Lilongwe-Malawi, 22-25 August 2017: Modeling infant mortality at Kigali University Teaching Hospital using Aalen Additive Hazards Model.

Acknowledgements

At the end of this PhD research program, I would like to thank all people who contributed to its successful outcome.

First, my greatest thanks go to my supervisors Dr. Sileshi Fanta Melesse and Prof. Shaun Ramroop for their tireless encourangement, advices, guidance and support. Your assistance, corrections and comments during the analysis of data and writing up phases of this thesis were very helpful. It was an honour to work with you.

Special thanks go to the School of Mathematics, Statistics and Computer Sciences of the University of KwaZulu-Natal, for support and encouragement. I also thank the Department of Maternity and Pediatrics in the Kigali University Teaching Hospital for providing data used in this dissertation. I would like to thank my family and friends for all the support and encouragement that they have provided to me.

This work was supported through the DELTAS Africa Initiative. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust, grant 107754/Z/15/Z- DELTAS Africa Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB) programme and the UK government. The views expressed in this publication are those of the authors and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

Abstract

The time to event analysis or survival analysis aims at making inferences on the time elapsed between the recruitment of subjects or the onset of observations, until the occurrence of some event of interest. Methods used in general statistical analysis, in particular in regression analysis, are not directly applicable to time to event data due to covariate correlation, censoring and truncation. While analysing time to event data, medical statistics adopts mainly non-parametric methods due to difficulty in finding the adequate distribution of the phenomenon under study.

This study reviews non-parametric classical methods of time to event analysis namely Aalen Additive Hazards Model (AAHM) trough counting and martingale processes, Cox Proportional Hazard Model (CPHM) and Cox-Aalen Hazards Model (CAHM) with application to the infant mortality at Kigali University Teaching Hospital (KUTH) in Rwanda. Proportional hazards assumption (PHA) was checked by assessing Kaplan-Meier estimates of survival functions per groups of covariates. Multiple events models were also reviewed and a model suitable to the dataset was selected. The dataset comprises 2117 newborns and socio-economic and clinical covariates for mothers and children. Two events per subject were modeled namely, the death and the occurrence of at least one of the conditions that may also cause long term death to infants.

To overcome the instability of models (also known as checking consistence of models) and potential small sample size, re-sampling was applied to both CPHM and appropriate multiple events model. The popular non-parametric re-sampling methods namely bootstrap and jack-knife for the available covariates were conducted and then re-sampled models were compared

to the non-re-sampled ones.

The results in different models reveal significant and non-significant covariates, the relative risk and related standard error and confidence intervals per covariate. Among the results, it was found that babies from under 20 years old mothers were at relatively higher risk and therefore, pregnancy of under 20 years old mothers should be avoided. It was also found that an infant's abnormality in weight and head increases the risk of infant mortality, clinically recommended ways of keeping pregnancy against any cause of infant abnormality were then recommended.

Contents

List of Figures

List of Tables

Abbreviations

AAHM: Aalen Additive Hazards Model

AGM: Andersen-Gill Model

APGAR: Appearance, Pulse, Grimace, Activity and Respiration

BCPHM: Bootstrap Cox Proportional Hazards Model

BMRSM: Bootstrap Marginal Risk Set Model

CAHM: Cox-Aalen Hazards Model

CI: Confidence Interval

CPHM: Cox Proportional Hazards Model

GEE: Generalised Estimating Equation

IMR: Infant Mortality Rate

JCPHM: Jackknife Cox Proportional Hazards Model

JMRSM: Jackknife Marginal Risk Set Model

KM: Kaplan-Meier

KUTH: Kigali University Teaching Hospital

MRSM: Marginal Risk Set Model

PHA: Proportional Hazard Assumption

PWPGTM: Prentice, Williams and Peterson Gap Time Model

PWPM: Prentice, Williams and Peterson Model

PWPTTM: Prentice, Williams and Peterson Total Time Model

SE: Standard Error

SSA: Sub-Saharan Africa

WLWM: Wei, Lin and Weissfeld Model

CHAPTER 1

INTRODUCTION

1.1 Background

Infant mortality refers to the death of an infant during the first year of life (Wasserman, 2013). The analysis of the anatomy of a disparity in infant mortality by Wise (2003) associates infant mortality with poor maternal health, poor quality and access to medical care and preventive services, and low socio-economic position. Ester et al. (2011) reported that a half of the 10 million infants who die annually in the world are from Sub-Saharan Africa (SSA). The report investigated different factors associated to such high infant mortality in SSA and pointed relatively poor health services due to economic and development indicators. Ester et al. (2011) suggested two requirements for decreasing infant mortality in SSA. These are extensive pushing to the economic growth and extensive studies on the main factors of the infant mortality. This thesis uses a dataset from one of the SSA countries and use different methods of survival analysis for understanding the factors and the level on which these factors are associated to the infant mortality.

The foundation of the theory of survival analysis started in 1975 with Aalen (1975). The theory on survival analysis has been consolidated by Flemming and Harrington (1991) with interest on using stochastic processes approach in survival analysis. Other researchers who

discussed on survival analysis include Andersen et al. (1993), Collet (2003), and Hosmer et al. (2008).

Aalen et al. (2008, p. 1) define survival analysis as a set of statistical methods for data where the outcome variable is the time until the occurrence of an event of interest.The event of interest can be for example death, occurrence of a disease or failure of a device. The survival analysis is complicated by censoring and truncation. Hosmer et al. (2008, pp. 3-9) give three types of censoring. Left censoring arising when an individual experienced an event before recruitment. Interval censoring refers to when the event occurs within some interval while right censoring arises 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. Two types of truncation as described by Klein and Moeschberger (2003, pp. 72-73) are left truncation occurring 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, due to the limited time allocated for PhD program and funds allocated for data collection, the time frame of the study will be one year and therefore interest will be only on right censoring.

Abundant studies in survival analysis deal with only single event for different subjects under study. In real world, more than one event may occur on the same subject over time and data within and among subjects may be correlated. Some examples in biomedical studies are repeated lung infections with pseudomonas in children with cystic fibrosis, development of breast cancer in genetically predisposed families, repeated heart attacks per subject, recurrence of bladder cancer tumors or deteriorating episodes of visual acuity. However, classical survival analysis is unable to handle multiple events on the same subject. Relatively recent approach

uses stochastic processes for extending classical survival analysis to multiple events analysis (Castañeda and Gerritse, 2010). This study will investigate how concepts of classical survival analysis can be incorporated in stochastic processes to analyse multiple events over time.

The multiple events may be identical or not and may occur more than once per subject with or without a certain order (Cook and Lawless, 2007, p. 1). An example is the time to the hospitalisation and death per subject where a typical work has been done by Castañeda and Gerritse (2010).

The majority of authors are interested only on the time to the first event, however, with these analyses, biased results may probably occur (Sagara et al., 2014). The mathematical formulation of multiple events has been intensively studied by authors such as Cook and Lawless (2007) who provided the likelihood formulation and important properties of multiple events models, Louzada (2007) who discussed on the important quantity called "intensity" in likelihood formulation, Sankaran and Anisha (2011) who introduced frailties in multiple events and Sun et al. (2006) who adapted additive hazard regression models to the multiple events. Multiple events are analysed through different models such as the Andersen-Gill Model (AGM); Prentice, Williams and Peterson Total Time Model (PWPTTM); Prentice, Williams and Peterson Gap Time Model (PWPGTM); the Frailty Model; the Wei, Lin and Weissfeld Model (WLWM) known also as the Marginal Risk set Model (MRSM) and the Generalised Estimating Equation Model (GEEM) (Amorim and Cai, 2015; Sagara et al., 2014; Wei et al., 1989). The existing software packages for multiple events survival analysis include SAS, STATA and R. An example where interest is taken on STATA is the work of Cleves (2000) for both ordered and unordered failure events with application to different medical researches, namely the diabetic retinopathy study, ursodeoxycholic acid in the treatment of biliary cirrhosis presented and discussed by Lindor et al. (1994) and the bladder cancer data presented by

Wei et al. (1989). Multiple events have been the interest of Cai and Schaubel (2004) who used multiple events data, presented by Schaubel et al. (1996) for a cohort study aimed at determining the incidence of asthma on preschool children. Amorim and Cai (2015) used the AGM, PWPTTM and PWPGTM while analyzing the case of bladder cancer while Sagara et al. (2014) analysed multiple events for the case of malaria episodes with a use of the AGM, PWPM, the GEEM and the Frailty Model.

This study applied single and multiple events methods of survival analysis to the data on infant mortality obtained from Kigali University Teaching Hospital (KUTH) collected from January 2016 to December 2016.

1.2 Aims and objectives of study

The aim of this study is to review and summarise methods for analysing survival data and discuss multiple events analysis of survival analysis. The objectives of this study are:

- To review methods of analysis time to event data.
- To use counting processes for analysing survival data.
- To analyse survival data with multiple events.

1.3 Study methodology

In this study, the classical survival models will be firstly reviewed. The Aalen Additive Hazard Model (AAHM) for single event per individual which involves Stochastic processes will be applied to the data. The AAHM will be used for assessing the relationship between the infant mortality and covariates. The event will be the death at or after birth. The Cox Proportional Hazard Model (CPHM) and the Cox-Aalen Hazard Model (CAHM) will be also reviewed and the results of these three models will be compared by using the existing methods of adequacy checking in survival analysis. Proportional hazards assumption (PHA) will be checked by assessing the Kaplan-Meier estimates of the survival functions per groups of covariates.

The model parameters in AAHM are considered as time dependent and the interest is taken on cumulative parameters. The AAHM assesses time dependence of covariates and suggests the fixed covariates. The CAHM combines multiplicative and additive parts, multiplicative part includes covariates whose PHA is realised, while covariates showing violation of PHA are included in the additive part. The CPHM assumes the presence of fixed covariates and would be preferred if the time dependent covariates are dropped out.

The present study selected a suitable model for multiple events analysis of the infant mortality at the Kigali University Teaching Hospital. The two events are of interest, namely, death and the occurrence of at least one of the conditions that may also cause long term death to infants.

To overcome the problem of small sample size and instability in modelling, re-sampled models were conducted by applying the popular nonparametric re-sampling techniques namely bootstrap and jackknife.

1.4 Research content

The research work described in this thesis contains eight chapters. The first chapter gives a brief background on survival analysis and models of survival analysis and the aim and objectives of the study. The second chapter describes the dataset that was used throughout the study. The third chapter reviews basic concepts, nonparametric methods, and single event models used in survival analysis. The fourth chapter presents multiple events models of survival analysis. The fifth chapter presents the re-sampled CPHM. Chapter six discusses re-sampled model with multiple events. Chapter seven discusses on the length of the confidence intervals in the Cox models and Chapter eight gives conclusions. STATA code for analysing the data used in this dissertation is given in the Appendix.

CHAPTER 2

BACKGROUND ON THE DATASET

2.1 Dataset

The theoretical results of this thesis will be illustrated by applying the existing methods of survival analysis and analyse data on infant mortality from KUTH. The time to event primary dataset of 2117 newborns at KUTH was recorded from the 1st January to 31st December 2016.

At KUTH, all newborns are recorded in registries with all details of mothers and clinical outcomes of the newborn. The information in the registry provides also references on card index that provides information on the clinical behavior of babies after leaving the hospital. KUTH as a site of interest in this study is a central hospital where most of the complicated childbirth countrywide are transferred. Along 2016, KUTH recorded a 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).

2.2 Data analysis

A complete case analysis is considered where the event for single event survival analysis is the death of the infant. Eleven covariates of interest are demographic covariates that include the *age* of a mother and the place of *residence* for parents; clinical covariates for mothers 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 the interaction of infant mortality and age of their mothers. Variables *residence* and *gender* are included 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 are found for example in Janssen et al. (2007).

Table 2.1: 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 mother	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 mother	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 mother 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 time or not	0=censored, 1=dead
n_events	Indicator on the rank of records per subject	1=first record, 2=second record

The multiple events analysis considers the second event as one of the conditions that may also cause long term death to infants such as severe oliguria, severe prematurity, very low birth weight, macrosomia, severe respiratory distress, gastroparesis, hemolytic, trisomy, asphyxia and laparoschisis. Table 2.1 describes the variables of interest and Table 2.2 summarises the dataset. The full dataset can be found via the authors of this thesis.

Table 2.2: Summary on newborns under study.

Total observations	2117
Deaths during the study time	82 (3.873%)
Stillborn babies	69 (3.259 %)
Total events	151 (7.132 %)
Censored babies	1966 (92.867 %)

2.3 Minimum sample size

The minimum sample size according to Peduzzi et al. (1996) is $N = \frac{10\,k}{p}$ where k is the number of predictor variables and p is the proportion of the total events. This suggests the minimum sample size at KUTH as $N = \frac{10\times 11}{0.07132} \approx 1542$.

CHAPTER 3

CLASSICAL SURVIVAL ANALYSIS WITH APPLICATION TO INFANT MORTALITY AT KUTH

3.1 Introduction

Infant mortality or mortality of children under their first birthday (Reidpath and Allotey, 2003; Bourgeois, 1946), attracts attention in several studies worldwide.

Benn Sartorius and Kurt Sartorius (2014) used data of the World Bank from 192 countries from 1990 to 2011 and found that the average of the Infant Mortality Rate (IMR) was 75/1000 in SSA versus 11/1000 in developed countries. Adetunji and Bos (2006) used the World Bank dataset from 1960 to 2005 and suggested that low life expectancy at birth in SSA is relatively higher in Middle Africa as compared to other sub-regional disparities of SSA. Other studies on infant mortality include 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 and Sartorius et al. (2011) who conducted an ecological spatial analysis on the infant mortality in South Africa. The incidence of a relatively higher rate in SSA justifies the need to identify and analyse 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 better control of the infant

mortality.

This chapter aims at using Kaplan-Meir estimation for presenting survival outcomes of infant mortality per covariate at KUTH, and for measuring the PHA. AAHM is conducted and fully interpreted for all covariates with event taken as the infant mortality. AAHM indicates time dependent covariates and allows to see fixed covariates that are adapted to the CPHM. Also, AAHM gives an idea on the covariates of multiplicative and additive parts of the CAHM. Significance is measured for comparing the performance of models.

This chapter comprises five sections. The first section is an introduction to the study. The second reviews background in survival analysis. The third section discusses nonparametric methods used in survival analysis. The fourth section discusses the classical survival regression models and the fifth section gives conclusions.

3.2 Concept of survival analysis

3.2.1 Background

Survival analysis is known also as time to event analysis. Survival analysis aims at making inferences on the time elapsed between the onset of an initiating event, until the occurrence of some event of interest. In short, the survival model in time to event analysis measures the dependence of time to an 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, pp. 3-9) describe 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, due to the structure of the recorded dataset, interest will be only on the 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 (also known 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).

3.2.2 Basic functions in survival analysis

Survival function

Assume that T is a random variable representing the time until the occurrence of an event of interest. Let $f(t)$ be the probability density function (pdf) of T. The cumulative distribution

function of T is given by

$$F(t) = Pr(T \leq t) = \int_{-\infty}^{t} f(z)\, dz.$$

The survival function is defined as

$$S(t) = Pr(T > t) = \int_{t}^{\infty} f(z)\, dz = 1 - F(t) \tag{3.1}$$

(Collet, 2003, p. 11). Equation (3.1) yields the following differential equation:

$$\frac{dS(t)}{dt} = -f(t). \tag{3.2}$$

Clearly,

1. $S(t)$ is decreasing on $[0, \infty)$ since $F(t)$ is increasing on $[0, \infty)$.
2. $S(0) = 1$ and $S(\infty) = 0$ since $F(0) = 0$ and $F(\infty) = 1$.

From the survival function, the other useful function can defined:

1. **Quantile** life time denoted by t_q: the smallest value t_q for which $S(t_q) \leq 1 - q$ where $0 \leq q \leq 1$.
2. **Median** lifetime: the 50^{th} percentile and thus corresponds to $q = 0.5$.

Hazard function

The hazard function also known as risk function or intensity rate is denoted by $h(t)$. The hazard function is the rate at which an individual is subject to the event along a small interval of time Δt given that the individual has not observed the event up to time t (Macdonald, 1996). That is

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \tag{3.3}$$

The hazard function (3.3) can theoretically take any value from zero to infinity. Equation (3.3) gives

$$h(t) = \frac{f(t)}{S(t)}. \tag{3.4}$$

Equations (3.2) and (3.4) yield $S(t)$ as

$$S(t) = e^{-H(t)} \tag{3.5}$$

or

$$H(t) = -\ln S(t) \tag{3.6}$$

where $H(t)$ is the cumulative hazard function given by

$$H(t) = \int_0^t h(z)\, dz. \tag{3.7}$$

3.3 Nonparametric estimation of the basic functions of survival analysis

The non-parametric approaches for estimating the three main functions of survival analysis include the life-table approach, the Kaplan-Meier approach and the Nelson-Aalen approach. The details of these approaches, the confidence interval of each function at a given time point and the related hypothesis test are developed in Collet (2003, pp. 17-36). In practice, the estimations from these approaches are close. This thesis adopts Kaplan-Meier approach which is relatively popular in medical statistical studies.

3.3.1 Kaplan-Meier estimate of survival function

Kaplan and Meier (1958) introduced an approach of estimating the survival function for the event times as described below: Assume that n is the the number of individuals whose survival times are $t_1, t_2, \ldots, t_n$.

Let r be the number of event times arranged in ascending order as $t_{(1)}, t_{(2)}, \ldots, t_{(r)}$ with $t_{(1)} < t_{(2)} < \ldots < t_{(r)}$. Let n_j denotes the number of individuals at risk before time $t_{(j)}$, $j = 1, 2, \ldots, r$. Let d_j be the number of individuals observing the event before $t_{(j)}$. The probability that an individual observes an event before time $t_{(j)}$ is estimated by the ratio $\frac{d_j}{n_j}$ and therefore, the probability of surviving at $t_{(j)}$ is $1 - \frac{d_j}{n_j}$. The estimated value of the survival function at any time t in the time interval from $t_{(k)}$ to $t_{(k+1)}$ is given in Collet (2003, p. 20) as

$$\widehat{S}(t) = \prod_{j=1}^{k} \left(1 - \frac{d_j}{n_j}\right) = \prod_{j=1}^{k} \left(\frac{n_j - d_j}{n_j}\right) \tag{3.8}$$

for $t_k \leq t < t_{k+1},\ k = 1, 2, \ldots, r$ with $\widehat{S}(0) = 1$ and $\widehat{S}(\infty) = 0$. The expression (3.8) is the Kaplan-Meier estimate of survival function at time t known also as the *Product Limit* estimate of the survival function at time t (Kaplan and Meier, 1958).

Taking the number of individuals at risk at time $t_{(j)}$ as

$$n_j - d_j = n_{j+1},\ j = 1, 2, \ldots, k,$$

(Collet, 2003, pp. 20-21) shows that the Kaplan-Meier estimate (3.8) is given by

$$\widehat{S}(t) = \frac{n_2}{n_1}\frac{n_3}{n_2} \cdots \frac{n_{k+1}}{n_k} = \frac{n_{k+1}}{n_1}$$

for $k = 1, 2, \ldots, r - 1$.

It can be shown by using the Greenwood formula (Greenwood, 1926), that the approximation of the variance of $\widehat{S}(t)$ is given by

$$\text{Var}\left[\widehat{S}(t)\right] \approx \widehat{S}^2(t) \sum_{j=1}^{k} \frac{d_j}{n_j(n_j - d_j)}.$$

Hence, the standard error of $\widehat{S}(t)$ is given by

$$\text{se}\left[\widehat{S}(t)\right] \approx \widehat{S}(t) \sqrt{\sum_{j=1}^{k} \frac{d_j}{n_j(n_j - d_j)}}, \tag{3.9}$$

with $t \in \left[t_{(k)},\, t_{(k+1)}\right)$. If there is no censored survival time,

$$n_j - d_j = n_{j+1}. \tag{3.10}$$

Expanding the sum in (3.9) gives

$$se\left[\widehat{S}(t)\right] \approx \widehat{S}(t) \sqrt{\frac{1 - \widehat{S}(t)}{n_1 \widehat{S}(t)}} \tag{3.11}$$

(Collet, 2003, p. 25). The $100(1-\alpha)\%$ confidence interval for the survival function, $S(t)$ is

$$\left[\widehat{S}(t) - z_{\frac{\alpha}{2}} \text{se}\left(\widehat{S}(t)\right);\, \widehat{S}(t) + z_{\frac{\alpha}{2}} \text{se}\left(\widehat{S}(t)\right)\right] \tag{3.12}$$

where se $\left[\hat{S}(t)\right]$ is given by the formula (3.9) or (3.11), and $z_{\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}$-point of the standard normal distribution.

3.3.2 Kaplan-Meier estimate of the cumulative hazard function

Equation (3.6) can be used to calculate the estimated Kaplan-Meier cumulative hazard function $\widehat{H}(t)$. It follows from Equation (3.6) and Equation (3.8) that the estimate of the cumulative hazard function is given by

$$\widehat{H}(t) = -\ln \widehat{S}(t) = -\sum_{j=1}^{k} \ln\left(1 - \frac{d_j}{n_j}\right). \tag{3.13}$$

The $100(1-\alpha)\%$ confidence limits of $H(t)$ can be calculated using a log-transformation on limits of results (3.12). That is

$$\left[-\ln\left(\widehat{S}(t)-z_{\frac{\alpha}{2}}\text{se}\left(\widehat{S}(t)\right)\right);-\ln\left(\widehat{S}(t)+z_{\frac{\alpha}{2}}\text{se}\left(\widehat{S}(t)\right)\right)\right]. \tag{3.14}$$

3.3.3 Kaplan-Meier estimate of the hazard function

Consider d_j, the number of events at the j^{th} event time denoted by $t_{(j)}$; n_j, the number of individuals at risk at time $t_{(j)}$ and $h(t)$, the hazard function at time t in the interval $[t_{(j)}, t_{(j+1)})$. The Kaplan-Meier estimate of $h(t)$ is

$$\hat{h}(t)=\frac{d_j}{n_j\,\tau_j} \tag{3.15}$$

where $\tau_j = t_{(j+1)} - t_{(j)}$ for $j = 1, 2, \ldots, r$ (Collet, 2003, pp. 30-31). Taking $p_j = \dfrac{d_j}{n_j}$ and assuming that d_j has a binomial distribution with parameters n_j and p_j, it follows that

$$\text{Var}(d_j) = n_j p_j(1-p_j) = n_j\,\frac{d_j}{n_j}\left(1-\frac{d_j}{n_j}\right) = d_j\frac{n_j-d_j}{n_j}. \tag{3.16}$$

It follows from (3.4) and (3.16) that

$$\text{Var}\left[\hat{h}(t)\right] = \frac{\text{Var}(d_j)}{n_j^2\,\tau_j^2} = \frac{d_j(n_j-d_j)}{n_j^3\,\tau_j^2} = \frac{d_j^2(n_j-d_j)}{n_j^2\,\tau_j^2\,n_j d_j} = \left[\hat{h}(t)\right]^2\frac{n_j-d_j}{n_j d_j}$$

and hence

$$\text{se}\left[\hat{h}(t)\right] = \hat{h}(t)\sqrt{\frac{n_j-d_j}{n_j d_j}}. \tag{3.17}$$

Thus, the $100(1-\alpha)\%$ confidence interval of the Kaplan-Meier of the hazard function at time t is

$$\left[\hat{h}(t)-z_{\frac{\alpha}{2}}\text{se}(\hat{h}(t));\ \hat{h}(t)+z_{\frac{\alpha}{2}}\text{se}(\hat{h}(t))\right]. \tag{3.18}$$

Collet (2003, p. 31) suggested that this confidence interval is meaningless for small values of d_j because it may be too wide.

3.3.4 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, p. 37-53). 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.

Collet (2003, p. 40-50) gives an argument on suitability of the log-rank and Wilcoxon tests. The log-rank test is suitable if proportional hazards can be assumed Collet (2003). In such a 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. The interpretation of tests is summarised in Table 3.1 suggested by Collet (2003).

Table 3.1: 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

3.3.5 Application

Consider the data described in Table 2.1 regarding the survival times of 2117 newborns from the Kigali University Teaching Hospital (KUTH). The aim of this study is to apply Kaplan-Meier for estimating the survival function, the cumulative hazard function and the hazard function as well as their 95% confidence intervals. The length of each time interval is taken as 30 days and thus 12 intervals are generated. Kaplan-Meier plots of survival function per covariate are used for making comparison of survival per groups of each covariate. The test of difference uses either log-rank or Wilcoxon test.

Kaplan-Meier estimates of the survival, cumulative hazard and hazard functions for the infant mortality at KUTH

The Kaplan-Meier estimates of the survival, cumulative hazard and hazard functions were calculated and plotted. The 95% confidence limits of $S(t)$, $H(t)$ and $h(t)$ were also calculated for the data at hand. Portions of the Kaplan-Meier estimates of the survival, cumulative hazard and hazard functions along with the 95% confidence limits are given in Table 3.2, 3.3 and 3.4, respectively.

The entire Kaplan-Meier estimates along with the 95% confidence limits for the survival, cumulative hazard and hazard functions are plotted in Figure 3.1, 3.2 and 3.3, respectively.

Table 3.2: Survival function estimate and 95% confidence intervals.

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]
⋮	⋮	⋮	⋮	⋮	⋮	⋮
362	28	0	6	0.959	0.005	[0.949; 0.967]
363	22	0	7	0.959	0.005	[0.949; 0.967]
364	15	0	2	0.959	0.005	[0.949; 0.967]
365	13	0	4	0.959	0.005	[0.949; 0.967]
366	9	0	9	0.959	0.005	[0.949; 0.967]

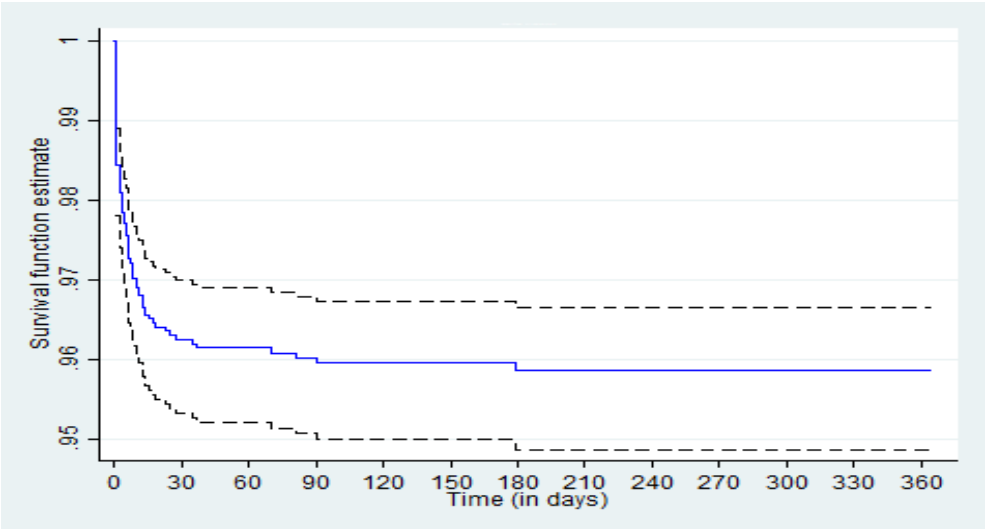


Figure 3.1: Survival function estimates and 95% confidence limits.

Table 3.3: Cumulative hazard function estimates and 95% confidence intervals.

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]
⋮	⋮	⋮	⋮	⋮	⋮	⋮
362	28	0	6	0.0415	0.005	[0.034; 0.051]
363	22	0	7	0.0415	0.005	[0.034; 0.051]
364	15	0	2	0.0415	0.005	[0.034; 0.051]
365	13	0	4	0.0415	0.005	[0.034; 0.051]
366	9	0	9	0.0415	0.005	[0.034 ; 0.051]

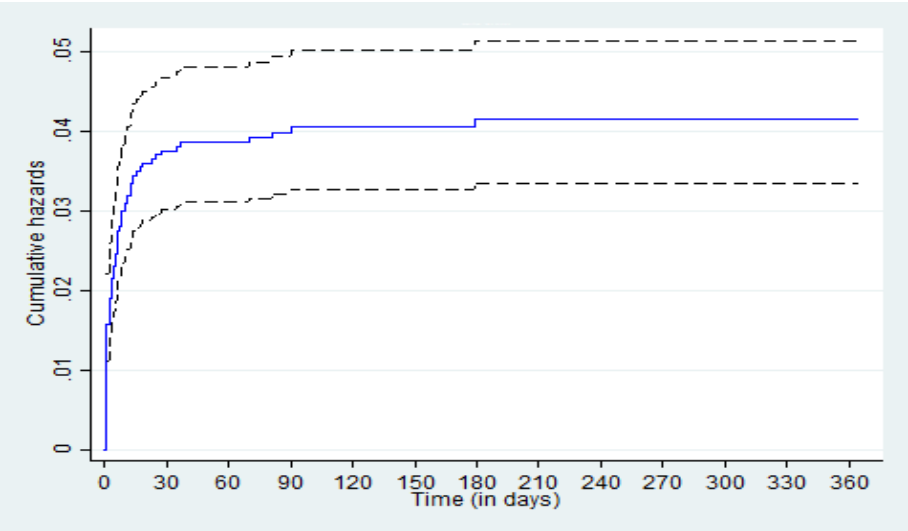


Figure 3.2: Cumulative hazard function estimates and 95% confidence limits.

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 3.2) 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 3.3). The confidence intervals of the hazard function are too wide since they include negative values and therefore, they are meaningless.

Table 3.4: Hazard function estimates.

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

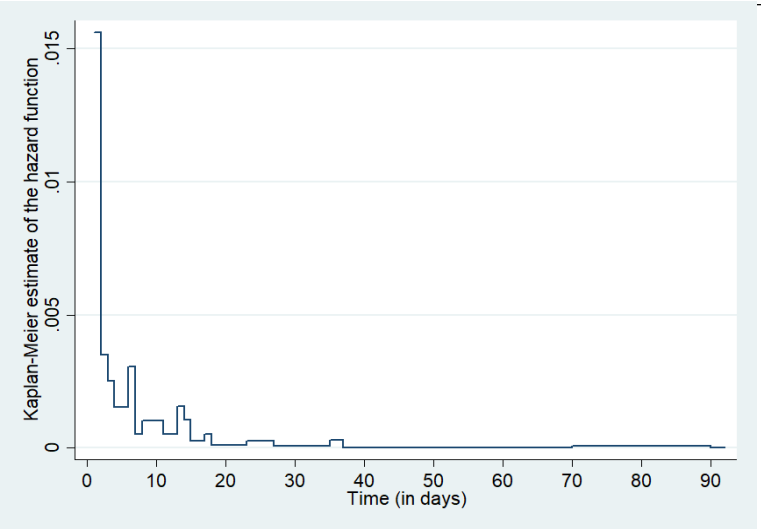


Figure 3.3: Hazard function estimates.

Comparison of groups of survival data

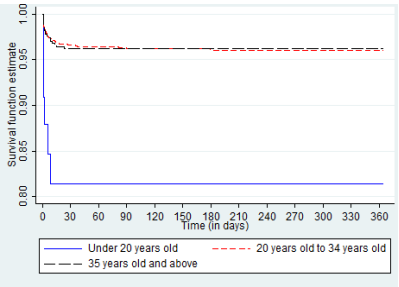
The comparison among the levels of the variables was done graphically. The illustration is summarised in Figure 3.4. The results of the log-rank and Wilcoxon test statistics are summarised in Table 3.5. The log-rank test for comparison is suitable for comparing levels of variables *residence*, *gender*, *number*, *APGAR* and *weight* where the plots do not cross. Wilcoxon test is suitable in comparing the levels of the rest of variables since their plots cross.

Figure 3.4 (a) suggests that babies whose mothers are 20 years old to 34 years and above 34 years survive better than babies whose mothers are under 20 years old. The Wilcoxon test strongly support the difference between these categories of age.

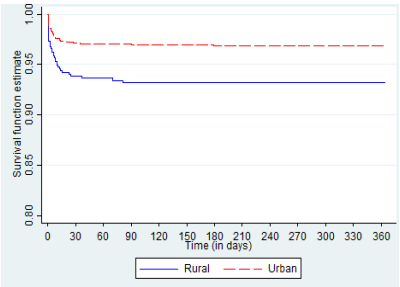
Figure 3.4 (j) suggests that babies whose circumference of the head is 32cm and above survive better than those with the circumference of the head below 32cm, with overwhelming evidence against the non-difference as shows the Wilcoxon test. Figure 3.4 (k) shows that babies with normal height (46-54cm) survive better than stunted (under-height babies or babies whose

height is less than 46cm) and over-height babies (babies whose height exceeds 54cm). The Wilcoxon test of no-difference is overwhelmingly against the no-difference between the levels of *height*. Wilcoxon test shows that there is no evidence of no-difference between levels of variables *antecedents*, *abortion* and *childbirth*.

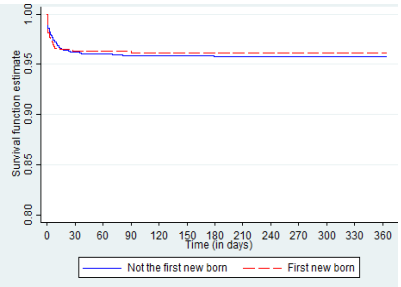
Figure 3.4 (b) suggests that urban babies survive better than rural babies, the log-rank test for no-difference confirms no-difference. Figure 3.4 (f) shows that the female babies survive better than males but the log-rank test confirms slight evidence against no-difference.



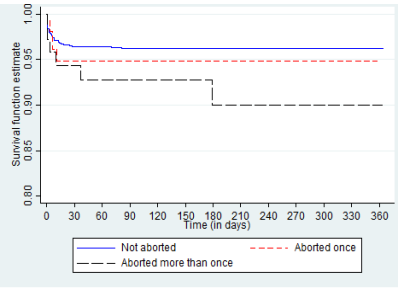
(a) Age of mothers



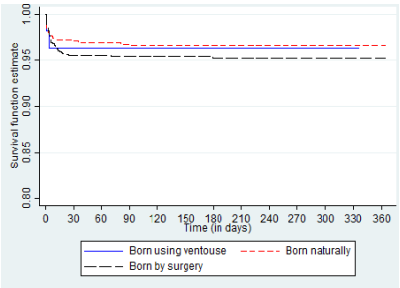
(b) Residence



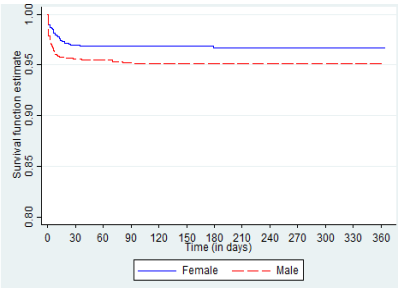
(c) Antecedents of parents



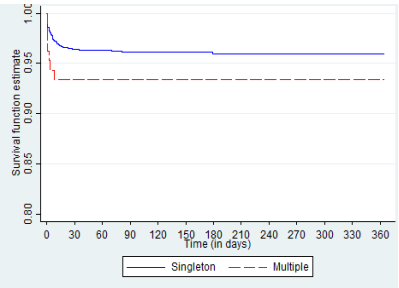
(d) Previous abortion



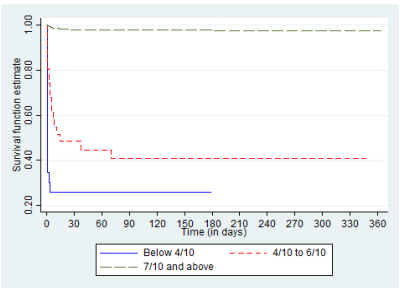
(e) Type of childbirth



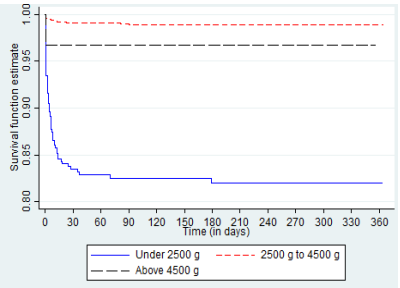
(f) Gender of newborns



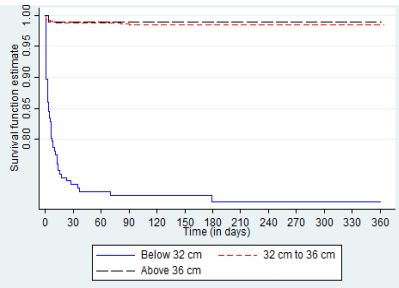
(g) Number of newborns at a time



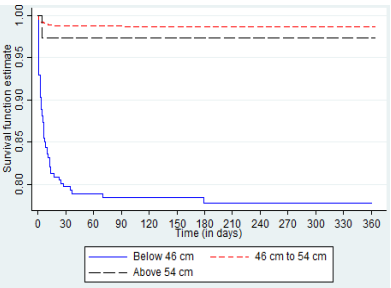
(h) APGAR of newborns



(i) Weight of newborns



(j) Head circumference of newborns



(k) Height of newborns

Figure 3.4: 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.

Singleton survive better than multiple births as suggests Figure 3.4 (g) but the log-rank test shows that there is no evidence of difference in levels of the number of newborns at a time. Figure 3.4 (h) suggests 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. The log-rank test shows an overwhelming evidence against the no-difference between the levels of APGAR.

Table 3.5: 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 $(p < 0.001)$

3.4 Regression models

3.4.1 Cox Proportional Hazards Model (CPHM)

Assume p fixed covariates with values $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})'$ and $h_0(t)$ a hazard function when values of all covariates are zeros. The CPHM is given by

$$h(t|\mathbf{x}_i) = h_0(t) \exp(\boldsymbol{\beta}'\mathbf{x}_i) \tag{3.19}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$ is a p-dimensional vector of model parameters (Collet, 2003, p. 58). The quantity

$$\psi = e^{\beta_k} \tag{3.20}$$

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, p. 90).

3.4.1.1 Parameter estimation in CPHM with no tied events

Parameter estimation for the model (3.19) with no tied events is done using partial likelihood introduced by Cox (1972) as introduced below. Consider time-to-event data with no tied events with ordered times to events $t_{(j)}$ for $j = 1, 2, \ldots, r$. Let $\mathbf{x}_{(j)}$ be the covariate vector of the individual whose time to event is $t_{(j)}$. Let $R(t_{(j)})$ be the set of individuals at risk at time $t_{(j)}$. The probability for an individual with the vector of covariate $\mathbf{x}_{(j)}$ to observe the event at

$t_{(j)}$ if one of the individuals in $R(t_{(j)})$ observes the event at $t_{(j)}$ is given by

$$\begin{aligned} P_j &= \frac{h[t_{(j)}|\mathbf{x}_{(j)}]}{\sum_{k\in R(t_{(j)})} h(t_{(j)}|\mathbf{x}_{(j)})} \\ &= \frac{h_0(t_{(j)})e^{\boldsymbol{\beta}'\mathbf{x}_{(j)}}}{\sum_{l\in R(t_{(j)})} h_0(t_{(j)})e^{\boldsymbol{\beta}'\mathbf{x}_l}} \\ &= \frac{e^{\boldsymbol{\beta}'\mathbf{x}_{(j)}}}{\sum_{l\in R(t_{(j)})} e^{\boldsymbol{\beta}'\mathbf{x}_l}}. \end{aligned}$$

The partial likelihood function is then given by

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{r} P_j = \prod_{j=1}^{r} \frac{e^{\boldsymbol{\beta}'\mathbf{x}_{(j)}}}{\sum_{l\in R(t_{(j)})} e^{\boldsymbol{\beta}'\mathbf{x}_l}} \tag{3.21}$$

(Collet, 2003, p. 66). The estimates of parameters $\beta_1, \beta_2, \ldots, \beta_p$ are obtained by maximizing $L(\boldsymbol{\beta})$ or, equivalently, by maximizing $\ln[L(\boldsymbol{\beta})]$ using the numerical methods such as for example the Newton-Raphson method (Autar, 2009). The usual Wald test for large samples tests the null hypothesis as $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ by assuming that

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' I(\hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \sim \chi^2_p. \tag{3.22}$$

In the expression 3.22, $I(\hat{\boldsymbol{\beta}})$ is the $p\times p$ observed information matrix $I(\boldsymbol{\beta}) = \left\{-\frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial\beta_j\partial\beta_k}\right\}_{(j,k)\in[1,p]\times[1,p]}$ evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ (Collet, 2003, p. 69). The variance-covariance matrix of the parameter estimates $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p)$ is $I^{-1}(\hat{\boldsymbol{\beta}})$. The variance of the parameter estimates $\hat{\beta}_k$ are the diagonal elements of $I^{-1}(\hat{\boldsymbol{\beta}})$ for $k = 1, 2, \ldots, p$. The likelihood ratio test is an alternative test for H_0: $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ versus $H_1 : \beta_k \neq 0$, for all $k \in [1, p]$. The test assumes that

$$\ln\left[\frac{L(\hat{\boldsymbol{\beta}})}{L(\boldsymbol{\beta}_0)}\right]^2 \sim \chi^2_p$$

(Klein and Moeschberger, 2003, p. 254). The $100(1-\alpha)\%$ confidence interval for β_k is given by

$$\left[\hat{\beta}_k - z_{\frac{\alpha}{2}}\mathsf{se}(\hat{\beta}_k); \hat{\beta}_k + z_{\frac{\alpha}{2}}\mathsf{se}(\hat{\beta}_k)\right] \tag{3.23}$$

where $\text{se}(\hat{\beta}_k) = \sqrt{\widehat{\text{Var}}(\hat{\beta}_k)}$ for $k = 1, 2, \ldots, p$ and $z_{\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}$-percentile of the standard normal distribution. It follows from result (3.23) that the $100(1-\alpha)\%$ confidence interval of a hazard ratio $\psi = e^{\beta_k}$ is given by

$$\left[e^{\hat{\beta}_k - z_{\frac{\alpha}{2}}\text{se}(\hat{\beta}_k)}; e^{\hat{\beta}_k + z_{\frac{\alpha}{2}}\text{se}(\hat{\beta}_k)}\right]$$

where $\hat{se}(\hat{\beta}_k) = \sqrt{\left[I(\hat{\beta}_k)\right]^{-1}}$.

3.4.1.2 Parameter estimation in CPHM with tied events

For tied events, the approximation of the partial likelihood function of the CPHM is based on the following procedure.

First, construct a vector $\mathbf{s}_j = (s_{1j}, s_{2j}, \ldots, s_{pj})'$ of sums of each of the p covariates for all individuals with event at the j^{th} ordered event time, $t_{(j)}$, for $j = 1, 2, \ldots, r$. More succinctly, if there are d_j events at time $t_{(j)}$, the h^{th} element of $\mathbf{s}_j$ is

$$s_{hj} = \sum_{k=1}^{d_j} x_{hjk}$$

where x_{hjk} is the value of the h^{th} explanatory variable, $h = 1, 2, \ldots, p$ for the k^{th} individual, $k = 1, 2, \ldots, d_j$, who observes the event at time $t_{(j)}$, $j = 1, 2, \ldots, r$. Second, construct the sets $\mathcal{D}(t_{(j)})$ and $\mathcal{R}(t_{(j)})$ of individuals who observe the event and who are at risk at time $t_{(j)}$, respectively. Third, express the observed vector of covariates for the i^{th} individual as $\mathbf{x}_i$, $i = 1, 2, \ldots, n$. The Breslow approximation of the partial likelihood function is given by

$$L_B(\boldsymbol{\beta}) = \prod_{j=1}^{r} \frac{e^{\boldsymbol{\beta}'\mathbf{s}_j}}{\left[\sum_{i \in \mathcal{R}(t_{(j)})} e^{\boldsymbol{\beta}'\mathbf{x}_i}\right]^{d_j}}. \tag{3.24}$$

The Breslow estimation performs well when $\frac{d_j}{n_j}$ is relatively small. Alternatively, the Efron approximation of the partial likelihood function is given by

$$L_E(\boldsymbol{\beta}) = \prod_{j=1}^{r} \frac{e^{\boldsymbol{\beta}'\mathbf{s}_j}}{\prod_{k=1}^{d_j}\left[\sum_{i\in\mathcal{R}(t_{(j)})} e^{\boldsymbol{\beta}'\mathbf{x}_i} - \frac{k-1}{d_j}\sum_{i\in\mathcal{D}t_{(j)}} e^{\boldsymbol{\beta}'\mathbf{x}_i}\right]} \tag{3.25}$$

and the Cox approximation of the partial likelihood function is given by

$$L_C(\boldsymbol{\beta}) = \prod_{j=1}^{r} \frac{e^{\boldsymbol{\beta}'\mathbf{s}_j}}{\sum_{i\in\mathcal{R}(t_{(j)},d_j)} e^{\boldsymbol{\beta}'\mathbf{s}_i}} \tag{3.26}$$

where $\mathcal{R}(t_{(j)}, d_j)$ is the set of d_j individuals drawn from the risk set $\mathcal{R}(t_{(j)})$ at time $t_{(j)}$. In practice, the three approximations of the partial likelihood function lead to similar results (Collet, 2003, p. 68). STATA provides options for using each of the above approximations with Breslow being the default.

Maximum likelihood estimation of model parameters and asymptotic variance-covariance matrix of parameter estimates, as well as confidence interval of parameters and test of significance are conducted in a similar fashion as for the model with no tied events described in Section 3.4.1.1.

3.4.2 Aalen Additive Hazards Model (AAHM)

The hazard function described by the CPHM in the previous section is expressed as the product of the baseline hazard function and a function of the covariates of interest, and thus proportional hazards may be assumed and covariates may be fixed. Aalen et al. (2008, p. 155-156) suggest that the additive hazards model may be adequate when there is no assumption of proportionality. Another advantage of the AAHM is that both fixed and time dependent covariates

are integrated in the model.

The AAHM expresses the hazard rate at time t of the i^{th} of n individuals with vector of covariates $\mathbf{x}_i(t) = (x_{i1}(t), x_{i2}(t), \ldots, x_{ip}(t))'$. That is given by

$$h[t|\mathbf{x}_i(t)] = \beta_0(t) + \beta_1(t)x_{i1}(t) + \beta_2(t)x_{i2}(t) + \ldots + \beta_p(t)x_{ip}(t) \tag{3.27}$$

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

Aalen et al. (2008, p. 157) argue that, for computation stability, estimation in model (3.27) should be based on the cumulative parameter functions

$$B_k(t) = \int_0^t \beta_k(v)dv, \tag{3.28}$$

$k = 0, 1, 2, \ldots, p$. Clearly, if $\beta_k(t)$ is constant, say $\beta_k(t) = \beta_k$, then

$$B_k(t) = \int_0^t \beta_k dv = \beta_k\, t \quad \text{which is represented by a straight line.}$$

Proposition 3.4.1. Let

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

Model (3.27) leads to the form

$$dN_i(t) = \sum_{k=0}^{p} Y_i(t)x_{ik}(t)dB_k(t) + dM_i(t) \tag{3.29}$$

where $x_{i0} = 1$.

Proof. Using Aalen et al.'s expression of the intensity process $\lambda_i(t)$ of the counting process $\{N_i(t), t \geq 0\}$ of the i^{th} individual at risk at time t as

$$\lambda_i(t) = Y_i(t)h\left(t|\mathbf{x}_i(t)\right) \tag{3.30}$$

where $h\left(t|\mathbf{x}_i(t)\right)$ is given by Equation (3.27), the intensity process of the counting process $N_i(t)$ of the i^{th} individual at risk at time t can be written as

$$\lambda_i(t) = Y_i(t)h\left(t|\mathbf{x}_i(t)\right) = Y_i(t)\left[\beta_0(t) + \sum_{k=1}^{p}\beta_k(t)x_{ik}(t)\right]. \tag{3.31}$$

Using (3.28), the intensity process (3.31) can be written as

$$\lambda_i(t) = Y_i(t)\left[\frac{d}{dt}B_0(t) + \sum_{k=1}^{p}x_{ik}(t)\frac{dB_k(t)}{dt}\right]$$

or, equivalently,

$$\lambda_i(t)dt = Y_i(t)\left[dB_0(t) + \sum_{k=1}^{p}x_{ik}(t)dB_k(t)\right]. \tag{3.32}$$

But the signal-noise representation of an observed counting process for the i^{th} individual from the Doob-Meyer decomposition theorem (Doob, 1953) is

$$dN_i(t) = \lambda_i(t)dt + dM_i(t) \tag{3.33}$$

where $M_i(t)$ is the martingale component of the counting process $N_i(t)$. Then, it follows from (3.32) and (3.33) that

$$dN_i(t) = \sum_{k=0}^{p}Y_i(t)x_{ik}(t)dB_k(t) + dM_i(t) \tag{3.34}$$

where $x_{i0} = 1$. □

Equation (3.34) has the form of a multiple linear regression model for the i^{th} 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 $k = 0, 1, 2, \ldots, p$ and $i = 0, 1, 2, \ldots, n$. Model (3.34) can be written in matrix form as

$$d\mathbf{N}(t) = \mathbf{X}(t)d\mathbf{B}(t) + d\mathbf{M}(t) \tag{3.35}$$

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 $Y_i(t), Y_i(t)x_{i1}(t), \ldots, Y_i(t)x_{ip}(t)$

$d\mathbf{B}(t) = (dB_0(t), dB_1(t), \ldots, dB_p(t))'$ 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.

Assuming that $d\mathbf{N}(t)$ is a Poisson process,

$$Var\left[d\mathbf{N}(t)\right] = E\left[d\mathbf{N}(t)\right] = \lambda(t)dt \text{ (Andersen et al., 1993, p. 52).}$$

It follows from (3.35) 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

$$d\widehat{\mathbf{B}}(t) = \left[(\mathbf{X}(t))'\mathbf{X}(t)\right]^{-1}(\mathbf{X}(t))'\,d\mathbf{N}(t). \tag{3.36}$$

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. The estimator $\widehat{\mathbf{B}}(t)$ is unbiased (Hosmer and Royston, 2002) and obtained by integrating both sides of equation (3.36) with respect to t, that is

$$\begin{aligned}\widehat{\mathbf{B}}(t) &= \int_0^t \left[(\mathbf{X}(t))'\mathbf{X}(t)\right]^{-1}(\mathbf{X}(t))'\,d\mathbf{N}(t) \\ &= \sum_{t_j \le t} \left[(\mathbf{X}(t_j))'\mathbf{X}(t_j)\right]^{-1}(\mathbf{X}(t_j))'\,\mathbf{y}_j \end{aligned} \tag{3.37}$$

where $\mathbf{y}_j$ is $n \times 1$ vector of zeros except the j^{th} component equals to unit if the j^{th} individual observes an event at time t_j (Hosmer et al., 2008, p. 319 and Hosmer and Royston, 2002). Furthermore, the variance-covariance matrix of $\widehat{\mathbf{B}}(t)$ is

$$\text{Var}\left[\widehat{\mathbf{B}}(t)\right] = \sum_{t_j \le t} \left[(\mathbf{X}(t_j))'\mathbf{X}(t_j)\right]^{-1}(\mathbf{X}(t_j))'\,\mathbf{D}(t_j)\mathbf{X}(t_j)\left[(\mathbf{X}(t_j))'\mathbf{X}(t_j)\right]^{-1} \tag{3.38}$$

where $\mathbf{D}(t_j)$ is an $n \times n$ diagonal matrix with elements y_j on the main diagonal where y_j are zeros except the j^{th} component equals to unit if the j^{th} individual observes an event at time t_j(Aalen et al., 2008, p. 158, and Hosmer and Royston, 2002). The derivation of results (3.38) from (3.37) 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

$$\text{Var}(\mathbf{Y}) = \mathbf{A}\,\text{Var}(\mathbf{X})\,\mathbf{A}'$$

(Mulaik, 2009).

Hosmer and Royston (2002) assumed that if the vector of cumulative parameter coefficients at time t is estimated by (3.37), and its variance-covariance matrix by (3.38), then the estimator of the model vector of parameter coefficients at time t_j is

$$\hat{\boldsymbol{\beta}}(t_j) = \left[(\mathbf{X}(t_j))'\,\mathbf{X}(t_j)\right]^{-1}(\mathbf{X}(t_j))'\,\mathbf{y}_j \tag{3.39}$$

and

$$\text{Var}\left[\widehat{\boldsymbol{\beta}}(t_j)\right] = \left[(\mathbf{X}(t_j))'\,\mathbf{X}(t_j)\right]^{-1}(\mathbf{X}(t_j))'\,\mathbf{D}(t_j)\mathbf{X}(t_j)\left[(\mathbf{X}(t_j))'\,\mathbf{X}(t_j)\right]^{-1}. \tag{3.40}$$

Aalen et al. (2008, p. 159) showed that the cumulative parameter function estimator $\hat{\mathbf{B}}(t)$ has approximately a multivariate normal distribution around its true value $\mathbf{B}(t)$, with the variance-covariance matrix expressed in (3.38). Therefore, the $100(1-\alpha)\%$ confidence interval for the k^{th} cumulative parameter functions $B_k(t)$ is expressed by

$$\widehat{B}_k(t) \pm z_{\frac{\alpha}{2}}\sqrt{\hat{\sigma}_{kk}(t)} \tag{3.41}$$

with $\hat{\sigma}_{kk}(t)$ the k^{th} diagonal element of the variance-covariance matrix expressed in the equation (3.38). To test that a covariate X_k has no significant effect on the hazard function given

in model (3.27), Aalen et al. (2008, p. 164) formulated the null and alternative hypotheses in the usual way as follows

$$H_0 : \beta_k(t) = 0, \ \forall t \in [0, t_0]$$

versus

$$H_1 : \beta_k(t) > 0 \text{ or } \beta_k(t) < 0$$

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 $\Delta\hat{B}_k(t_j)$ at time t_j of the cumulative parameter function given in (3.37) tends to fluctuate around zero (Aalen et al., 2008, p. 164). Under the alternative hypothesis $H_1 : \beta_k(t) > 0$, the increment $\Delta\hat{B}_k(t_j)$ tends to be positive while under $H_1 : \beta_k(t) < 0$, the increment $\Delta\hat{B}_k(t_j)$ tends to be negative. Furthermore if $\widehat{B}_k(t)$ 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 (3.27) and assume that there is a need to test the null hypothesis

$$H_0 : \beta_k(t_j) = 0 \text{ for all } \ k \text{ with } k = 0, 1, \ldots, p. \tag{3.42}$$

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

$$\hat{\mathbf{u}} = \sum_{t_j} \mathbf{K}_j \hat{\boldsymbol{\beta}}(t_j) \tag{3.43}$$

where $\hat{\boldsymbol{\beta}}(t_j)$ given by (3.39) is the vector of estimators of the parameter coefficients for model (3.27), and $\mathbf{K}_j$ is a $(p + 1) \times (p + 1)$ diagonal matrix of weights. Four types of weights can be used (Hosmer and Royston, 2002) .

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

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

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

Weights 4: $\mathbf{K}_j = \text{diag}\left[\widehat{S}_{KM}(t_{j-1})/\text{se}(\hat{\beta}_{kk}(t_j)\right]$ where $\hat{\beta}_{kk}(t_j)$ is the k^{th} diagonal element (i.e. a variance) of the variance-covariance matrix (3.40). 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 $\hat{\mathbf{u}}$ given in (3.43) is obtained from the symmetric matrix $\mathbf{K}_j$ and the variance of $\hat{\beta}(t_j)$ given by (3.40). Hence,

$$\begin{aligned}\widehat{\text{Var}}(\hat{\mathbf{u}}) &= \sum_{t_j} \mathbf{K}_j \text{Var}\left[\hat{\boldsymbol{\beta}}(t_j)\right] \mathbf{K}_j' \\ &= \sum_{t_j} \mathbf{K}_j \left[(\mathbf{X}(t_j))' \mathbf{X}(t_j)\right]^{-1} (\mathbf{X}(t_j))' \mathbf{D}(t_j) \mathbf{X}(t_j) \left[(\mathbf{X}(t_j))' \mathbf{X}(t_j)\right]^{-1} \mathbf{K}_j'. \quad (3.44)\end{aligned}$$

Hence, the test statistic for H_0 given in (3.42) is

$$z_{u_k} = \frac{\hat{u}_k}{\text{se}(\hat{u}_k)} \quad (3.45)$$

where $\hat{u}_k$ is the k^{th} element of $\hat{\mathbf{u}}$ given in (3.43) and $\text{se}(\hat{u}_k)$ is the square root of the k^{th} diagonal element of $\widehat{\text{Var}}(\hat{\mathbf{u}})$ given in (3.44). Hosmer and Royston (2002) pointed out that the statistic z_{u_k} in (3.45) 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*. The plot of parameter functions and related test of

significance using the command *stlh* are presented in the appendix of this thesis.

Hosmer and Royston (2002) 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.

3.4.3 Cox-Aalen Hazards regression Models (CAHM)

CAHM was proposed by Scheike and Zhang (2002). The model consists of partitioning covariates 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 $(p \times 1)$ vector 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

$$h(t|x) = Y(t)[X'(t)\beta(t)] \exp(Z'(t)\alpha). \tag{3.46}$$

The 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). The score function assumes Aalen additive hazards model as its covariate dependent baseline. The baseline intensity functions and the relative risk parameters of the Cox model are suggested by solving the score equations. Numerical methods such as Newton-Raphson may be used for approximating model parameters.

3.4.4 Application

Consider the data described in Table 2.1 regarding the survival times of 2117 newborns from the Kigali University Teaching Hospital (KUTH). The aim of this study is to apply the CPHM, AAHM and CAHM for estimating the relative risk in each covariate. These three models were compared to select the relatively better model to the dataset of interest.

CPHM for the infant mortality at KUTH

Table 3.6 presents the estimates of the hazard ratios using the Cox proportional hazard model (3.19). For handling ties, Breslow, Efron and Cox approaches as defined in Section 3.4.1, give similar results and thus those presented in Table 3.6 are from the default (Breslow).

Table 3.6: 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 3.6 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 mothers 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 mothers are under 20 years old. The hazard of death of babies whose mothers are 35 years old and above is 0.279 (95% CI: 0.099-0.784, p=0.015) times that of babies whose mothers are under 20 years old. The argument of Olausson et al. (1999) confirms a relatively higher risk for teenage pregnancies due to biological immaturity. As for the advanced maternal age, Lampinen et al. (2009) suggested that relatively poorer outcomes to pregnancies are due to the observed higher

incidence of chronic medical conditions among older women.

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 usual better survival outcome of the females is also reported in several manuscripts such as Pongou (2013) and Zarulli et al. (2018).

The hazard of death of multiple babies is 0.324 (95% CI: 0.137-0.770, p=0.011) times that of singleton babies. This is however against the results from studies conducted in Sub-Saharan Africa by Monden and Smits (2017) and Pongou et al. (2019). This may be due to the small number of multiple newborns recorded at KUTH along the year 2016.

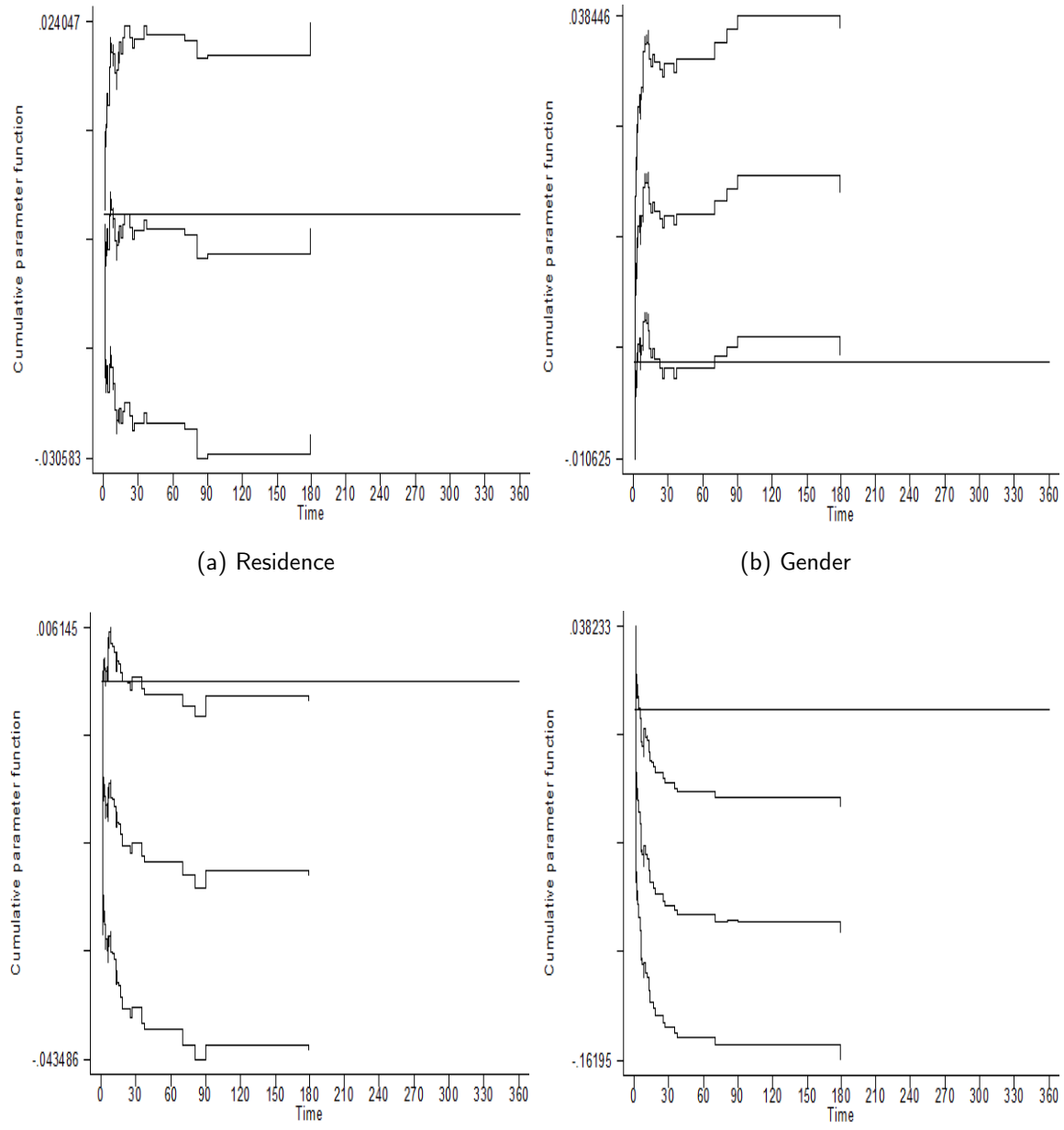
The hazard of death for babies whose APGAR range from 4/10 to 6/10 is 0.387 (95% CI: 0.182-0.822, p=0.014) times that of babies whose APGAR is below 4/10. The hazard of death for babies whose APGAR range from 7/10 to 10/10 is 0.056 (95% CI: 0.028-0.113, p<0.001) times that of babies whose APGAR is below 4/10. The hazard of death for babies whose weight range from 2500g to 4500g is 0.219 (95% CI: 0.101-0.479, p<0.001) times that of babies whose weight is below 2500g. The hazard of death for babies whose circumference of the head range from 32cm to 36cm is 0.287 (95% CI: 0.134-0.611, p=0.001) times that of babies whose circumference of the head is below 32cm. The hazard of death for babies whose circumference of the head is above 36cm is 0.125 (95% CI: 0.016-0.980, p=0.048) times that of babies whose circumference of the head is below 32cm. The results of APGAR, weight and circumference of the head comply with the recommendations of the clinical medicine as suggested by Janssen et al. (2007).

Aalen additive hazards model

Unlike the CPHM based on quantitative measurement of the 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 the significance of the parameters $\beta_k(t)$. The analysis was done in STATA using the command *stlh* suggested by Hosmer and Royston (2002).

Figure 3.5 gives the plots of the cumulative parameter functions and their 95% confidence limits for the variables *residence*, *gender*, *antecedents* and *number*.



(a) Residence (b) Gender

(c) Antecedents (d) Number

Figure 3.5: Cumulative parameter function plots for variables *residence*, *gender*, *antecedents and number*.

Figure 3.5 (a) represents the estimated cumulative parameter function with its 95% confidence

limits for urban parents compared to rural parents. The plot is 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 for urban babies may be slightly higher than that of rural babies, but the difference may be not significant.

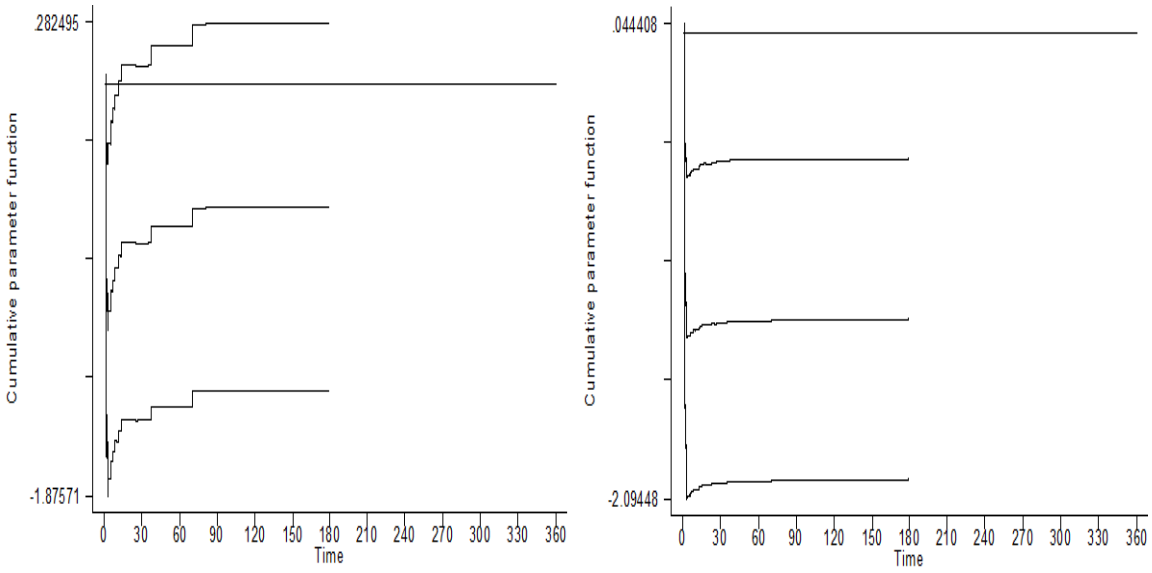
Figure 3.5 (b) represents the estimated cumulative parameter function with its 95% confidence limits for the male compared to female babies. The plot is approximately horizontal and positive everywhere with the upper and lower limits of the confidence interval situated approximately above the zero line. This indicates that the hazard of death for male babies may be higher than that of female babies.

Figure 3.5 (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 slightly higher than that of first newborn.

Figure 3.5 (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.

Figure 3.6 gives the plots of the cumulative parameter functions and their 95% confidence limits for levels of the variable *APGAR* Figure 3.6 (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

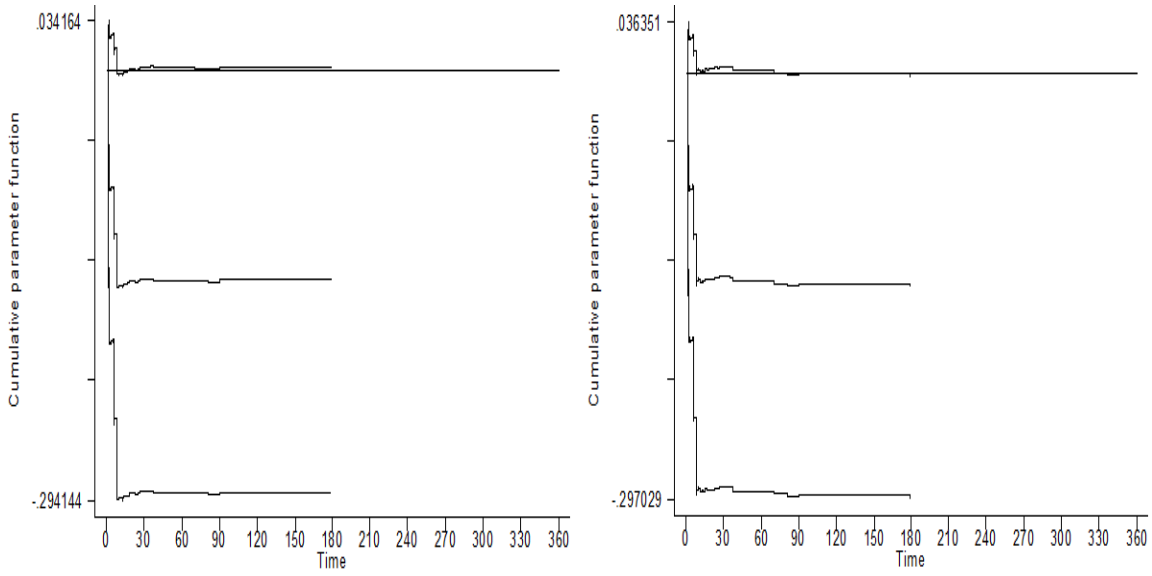
intervals situated below the zero line. This indicates that the hazard of death for newborns with APGAR below $4/10$ may be higher than that of newborns with APGAR from $4/10$ to $6/10$. Figure 3.6 (b) represents the estimated cumulative parameter function with its 95% confidence limits for the newborn's 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. The results of APGAR by AAHM comply with that of the CPHM and the recommendations of the clinical medicine found in Janssen et al. (2007).



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

Figure 3.6: Cumulative parameter function plots for variable *APGAR*.

Figure 3.7 gives the plots of the cumulative parameter functions and their 95% confidence limits for levels of the variable *age*.



(a) 20 years old to 34 years old

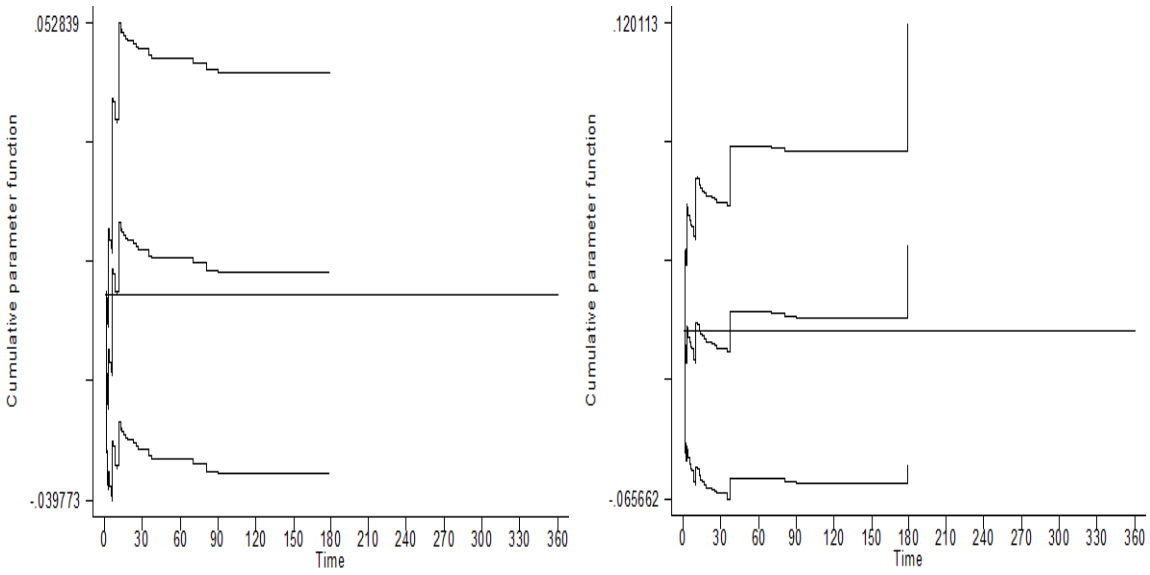
(b) 35 years old and above

Figure 3.7: Cumulative parameter function plots for variable *age*.

Figure 3.7 (a) displays the cumulative parameter function for mothers 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 zero line and the slope decreases only during the first month of study time but a large portion of its 95% confidence interval is below the zero line. This is an indication that the hazard of death for newborns whose mother is under 20 years old is higher than that of newborns from mothers with 20 years old to 34 years old. A similar situation is observed for newborns whose mothers are 35 years old and above as indicates Figure 3.7 (b). The same results were found by applying the CPHM and were justified by Olausson et al. (1999) and Lampinen et al. (2009).

Figures 3.8 (a) and 3.8 (b) display the cumulative parameter functions with their 95% confidence limits for newborns whose mothers aborted once and newborns whose mothers aborted more than once, respectively. The plots fluctuate around the zero line. This indicates that the

differences among the levels of abortion are not significant.



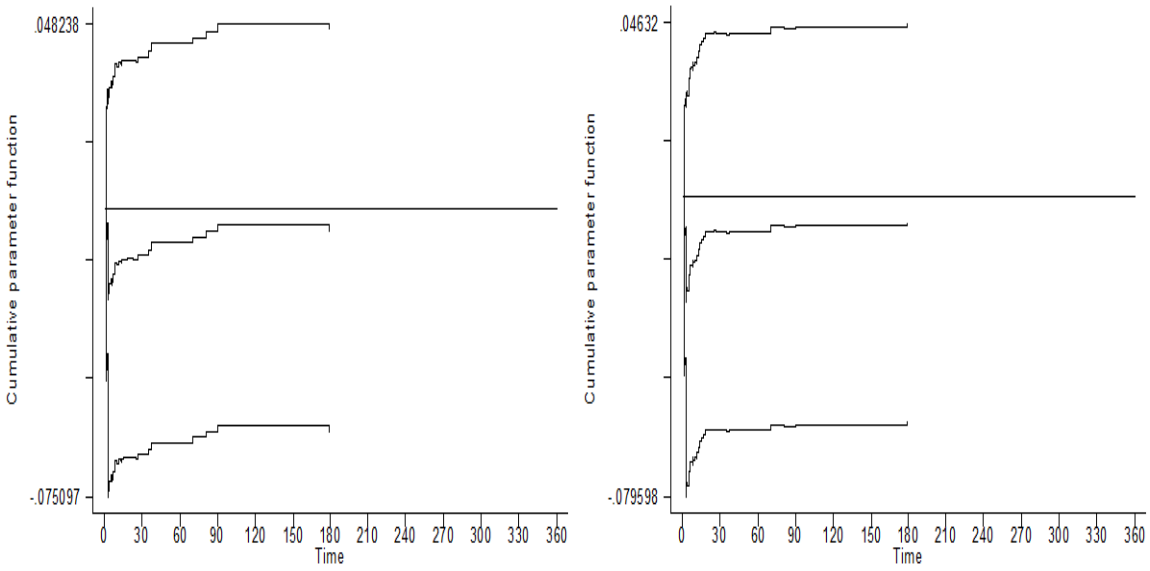
(a) Aborted once (b) Aborted more than once

Figure 3.8: Cumulative parameter function plots for variable *abortion*.

Figure 3.9 gives the plots of the cumulative parameter functions and their 95% confidence limits for levels of the variable *childbirth*. Both Figure 3.9 (a) and 3.9 (b) behave similarly: the plots are approximately horizontal and negative everywhere but the upper and lower limits of the confidence interval are on either side of the zero line. This indicates that the hazard of death of newborns by ventouse is slightly higher than that of newborns naturally and by surgery.

Figure 3.10 gives the plots of the cumulative parameter functions and their 95% confidence limits for levels of the variable *head*. Both Figure 3.10 (a) and 3.10 (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 the circumference of head less than 32cm is higher than that of newborns with the normal

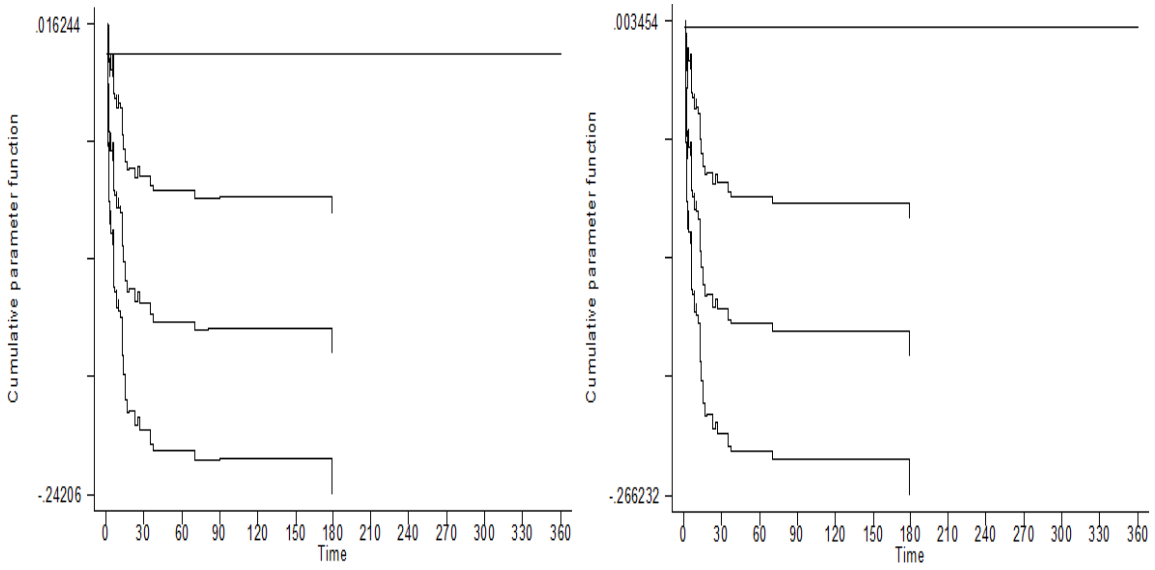
circumference of the head and that of newborns with an extra-normal circumference of the head.



(a) Born naturally (b) Born by surgery

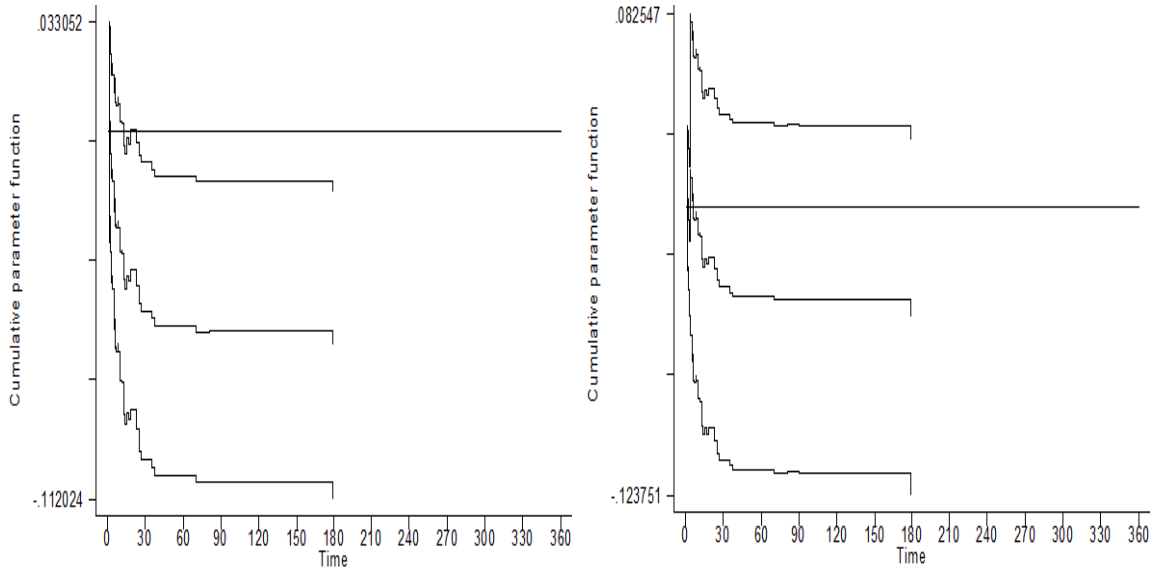
Figure 3.9: Cumulative parameter function plots for variable *childbirth*.

Figure 3.11 gives the plots of the cumulative parameter functions and their 95% confidence limits for levels of the variable *height*. Both Figure 3.11 (a) and 3.11 (b) behave similarly apart from the upper limit of the confidence interval of the plot 3.11 (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) 32cm to 36cm (b) Above 36cm

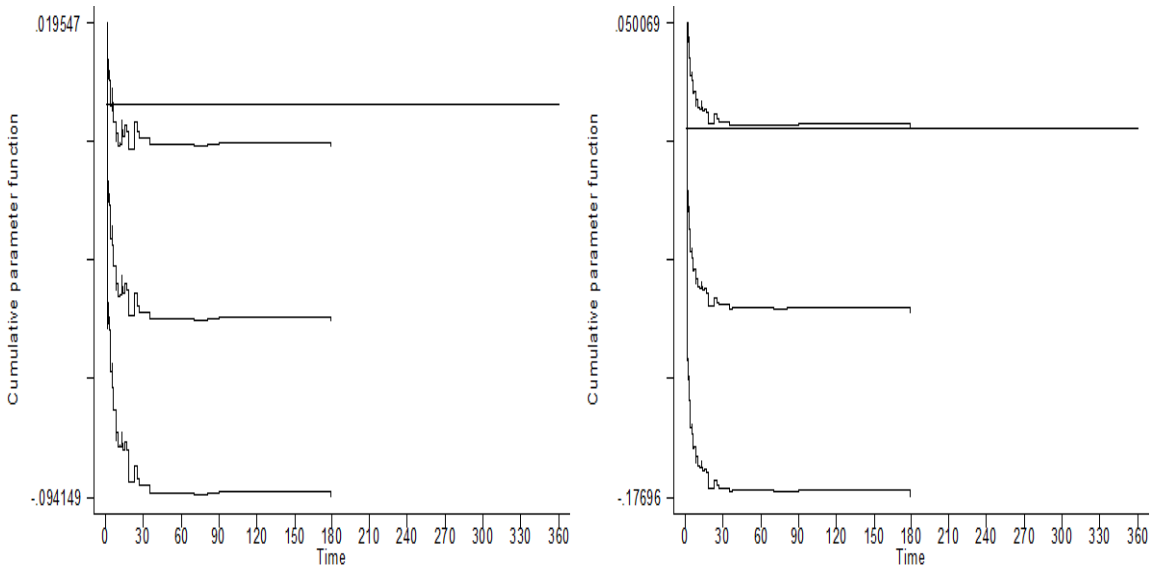
Figure 3.10: Cumulative parameter function plots for variable *head*.



(a) 46cm to 54cm (b) Above 54cm

Figure 3.11: Cumulative parameter function plots for variable *height*.

Figure 3.12 gives the plots of the cumulative parameter functions and their 95% confidence limits for levels of the variable *weight*. Both Figure 3.12 (a) and 3.12 (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) 2500g to 4500g (b) Above 4500g

Figure 3.12: Cumulative parameter function plots for variable *weight*.

Table 3.7 gives results on the test of significance of parameter functions. 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*.

Table 3.7: Tests for significance of covariates.

		Test 1		Test 2		Test 3		Test 4	
Covariate (reference)	Level	z	P	z	P	z	P	z	P
Age (Under 20 years old)	20 years old to 34 years old	-1.939	0.053	-1.943	0.052	-1.941	0.052	1.988	0.047
	35 years old and above	-1.992	0.046	-1.971	0.049	-1.992	0.046	-1.768	0.077
Residence (Rural)	Urban	-0.137	0.891	-0.233	0.816	-0.142	0.887	0.884	0.376
Antecedents (Not first new born)	first newborn	-2.190	0.028	-2.174	0.030	-2.189	0.029	-2.855	0.004
Abortion (not aborted)	aborted once	0.224	0.823	0.230	0.818	0.215	0.829	-5.086	$p < 0.001$
	Aborted more than once	0.771	0.440	0.510	0.610	0.762	0.466	-6.188	$p < 0.001$
Childbirth (born using ventouse)	Born naturally	-0.226	0.821	-0.256	0.798	-0.234	0.815	3.089	0.002
	Born by surgery	-0.246	0.805	-0.298	0.766	-0.254	0.799	3.079	0.002
Gender (Female)	Male	2.037	0.042	2.122	0.034	2.046	0.041	2.003	0.045
Number (Single)	Multiple	-3.488	$p < 0.001$	-3.365	0.001	-3.472	0.001	-6.177	$p < 0.001$
APGAR (under 4/10)	4/10 to 6/10	-1.299	0.194	-1.437	0.151	-1.325	0.185	-1.790	0.073
	7/10 and above	-3.452	0.001	-3.500	$p < 0.001$	-3.459	0.001	1.508	0.131
Weight (Under 2500 g)	2500 g to 4500g	-2.438	0.015	-2.418	0.016	-2.431	0.015	-2.643	0.008
	Above 4500 g	-1.981	0.048	-1.937	0.053	-1.970	0.049	-5.309	$p < 0.001$
Head (below 32 cm)	32 cm to 36 cm	-4.192	$p < 0.001$	-4.199	$p < 0.001$	-4.191	$p < 0.001$	-3.754	$p < 0.001$
	Above 36 cm	-4.686	$p < 0.001$	-4.730	$p < 0.001$	-4.688	$p < 0.001$	-4.855	$p < 0.001$
Height (Below 46 cm)	46 cm to 54 cm	-2.752	0.006	-2.638	0.008	-2.738	0.006	-3.750	$p < 0.001$
	Above 54 cm	-1.227	0.220	-1.098	0.272	-1.213	0.225	-3.964	$p < 0.001$

Test 1: weights equal to 1.0.

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

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

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

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 a significant difference between the 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 a 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).

Cox-Aalen Hazards Model (CAHM)

Multiplicative part of the CAHM (Table 3.8) 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 presents a huge difference in levels. The CAHM suggests that the hazard 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 3.13 summarises the additive part of the CAHM. The interpretation is not far from that of AAHM.

Table 3.8: 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]

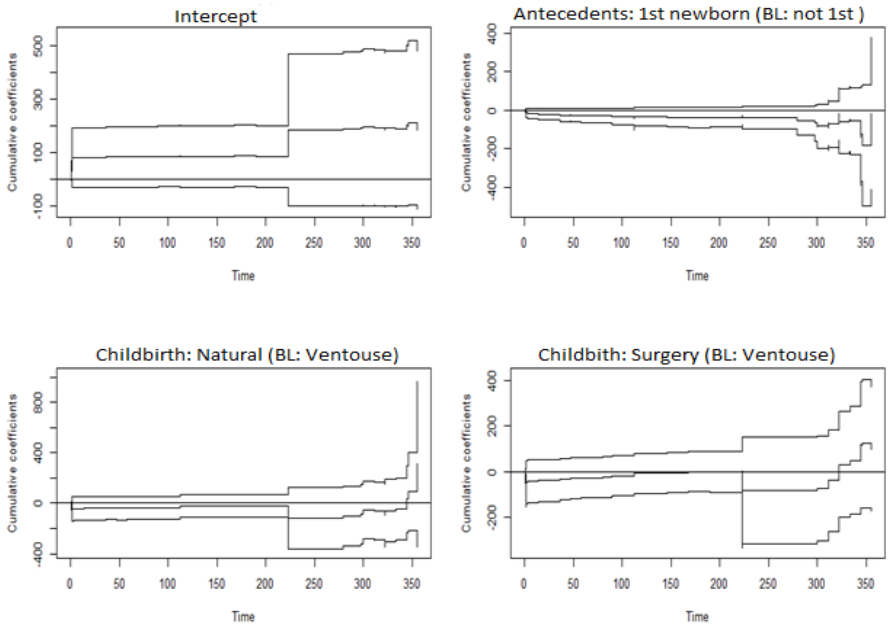


Figure 3.13: Additive part of the CAHM.

3.5 Conclusion

This chapter 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 (CAHM). These methods were used to analyse the dataset collected at the Kigali University Teaching Hospital for 2117 newborns during 366 days of the year 2016.

The results revealed that the hazard of death of an infant, is higher in male babies as compared to female babies; it is higher for babies whose mothers are under 20 years old mothers as compared to older mothers. 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 a relatively big head and a 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 the AAHM were significant for all variables except variable *residence* while for the CAHM, the significance was found on covariates *age*, *number*, *APGAR* and *weight*. The results on variable *height* were surprisingly not significant by the CPHM and by the CAHM, unlike expected results. Significance on variable *height* was rather observed in the AAHM in accordance with related tests, especially test 4.

This chapter reviewed different models with single event. However, additional events such as the incidence of chronic disease per subject along the study time would be of interest and this is beyond single event survival analysis. The extension to several events survival analysis is therefore targeted in the next chapter.

CHAPTER 4

MULTIPLE EVENTS MODELS

4.1 Introduction

The multiple events processes or processes that generate events repeatedly along the time are also known as the recurrent event processes (Cook and Lawless, 2007, p. 1). Such processes are adapted to the repeated event data found in medicine and public health, where the number of events exhibited is relatively small for a larger number of processes. Multiple events are met in the other domains such as social science, economics, manufacturing, insurance and reliability (Cook and Lawless, 2002). In multiple events studies, the number of events in distinct time intervals is termed as "**counts**", the **gaps** are the times between successive events, while the "**event intensity**" is the conditional probability of a new event, given the past event (Cook and Lawless, 2007, p. 1).

Cook and Lawless (2007) discussed different multiplicative models such as the **modulated Poisson model** which consist of modelling the intensity processes given the history $\mathscr{F}$, and the Cox models for ordered and unordered events. The interest in this study will be taken on the multiplicative models with the ordered events. Ordered events are based on the concept that the second event cannot occur before the first event, the third event cannot occur before the second event and so on. The models adapted to ordered events include the Andersen-Gill

Model (AGM), the Wei, Lin and Weissfeld Model (WLWM) and the Prentice, Williams and Peterson Model (PWPM) (Wei and Glidden, 1997).

The AGM known also as the counting process approach (Andersen and Gill, 1992), assumes that all event types are indistinguishable and all events within the same subject are assumed to be independent (Johnson et al., 2004). Therneau (1997) evokes a limitation of AGM of not allowing multiple events to occur at the same time. The WLWM is also known as the marginal risk sets model (Wei et al., 1989). The WLWM assumes that events are unordered where each event has its own stratum and each data point appears in all strata. This allows an analysis of multiple events occurring at the same time. The PWPM is also known as the conditional risk set model and was proposed by Prentice, Williams and Peterson (Prentice et al., 1981). In PWPM, the set up of the dataset is the same as that of the AGM but the analysis is stratified by failure order (Amorim and Cai, 2015). The PWPM can potentially analyse time to each event from the previous event, this is known as the gap-time model. Both AGM, WLWM and PWPM have been alternatively used on bladder cancer data and on the hospitalisation and death data presented by Castañeda and Gerritse (2010).

4.2 Mathematical formulation of Cox model with multiple events

Consider the time scale t, $t > 0$ and a sample of n individuals under study and let

$N_i(t)$ denotes the number of events for individual i, $i = 1, 2, ..., n$,

$T_{i1}, T_{i2}, ...$ denote the times of events for individual i,

$W_{ij} = T_{ij} - T_{ij-1}$ denote the gaps or times between successive events of the individual i,

$y_i(t)$ denote the fixed or time-varying covariates,

$N_i(t)$ denotes a counting process with intensity process

$$\lambda_i(t) = \lim_{\Delta t \to 0} \frac{P\left(N_i(t+\Delta t) - N_i(t) = 1 | \mathscr{F}_t\right)}{\Delta t}$$

where $\mathscr{F}_t$ is the history of events and covariates up to the time t (Sankaran and Anisha, 2011). The mean cumulative function (MCF) $\mu_i(t)$ and the corresponding rate of occurrence function $\rho_i(t)$ are defined in Cook and Lawless (2007) as:

$$\mu_i(t) = E\left[N_i(t)\right] \tag{4.1}$$

and

$$\rho_i(t) = \frac{d}{dt}\mu_i(t) \tag{4.2}$$

or

$$\rho_i(t)dt = d\mu_i(t). \tag{4.3}$$

Applying differentiation with respect to t on both sides of (4.1) and using (4.3) yields:

$$E\left[dN_i(t)\right] = \rho_i(t)dt. \tag{4.4}$$

Cook and Lawless (2007) discuss different multiplicative models such as the regression model for the rate function for both fixed and time dependent covariates expressed by

$$\rho_i(t) = \rho_0(t)e^{\boldsymbol{\beta}' y_i(t)} \tag{4.5}$$

and the regression model for the mean functions for the fixed covariates, expressed by

$$\mu_i(t) = \mu_0(t)e^{\boldsymbol{\beta}' y_i},$$

where $\mu_0(t) = \int_0^t \rho_0(v)dv$.

The second approach consists of modelling the intensity process $\lambda_i(t)$ given the history $\mathscr{F}$, that is

$$\lambda_{ik}(t|\mathscr{F}) = \lambda_{0k}(t)\, e^{\boldsymbol{\beta}' y_{ik}(t)}. \tag{4.6}$$

The expression λ_{0k} is the event specific baseline hazard for the k^{th} event over time. Model (4.6) incorporates the AGM, the WLWM and the PWPM according to the type of the dataset. Specifically Model (4.6) yields the PWPM gap model of the form

$$\lambda_{ik}(t|\mathscr{F}) = \lambda_{0k}[B(t)]\, e^{\boldsymbol{\beta}' y_{ik}(t)},$$

where $B(t) = t - T_{N(t^-)}$ is the time since the last event.

4.3 Likelihoods and maximum likelihood estimation

The likelihoods constructions and maximum likelihood estimates for the multiplicative multiple events models are well developed in Cook and Lawless (2007, p. 27-58), and specifically, Louzada (2007) discussed a parametric based estimation for the rate function model; Lawless and Nadeau (2012) addressed two ways of analyzing the rate function: the first one consists of specifying the distribution of the intensity process $\lambda_i(t)$ such as for example a Poisson process when $\lambda_i(t) = \rho_i(y)$, or a negative binomial process if $\lambda_i(t) = \frac{1+rN_i(t^-)}{1+r\mu_i(t^-)}\rho_i(t)$. In the second way, a distribution of the intensity process is not specified, this approach is known as "*robust*" and has potential to model means or variances (Sankaran and Anisha, 2011).

Assuming that two events cannot occur simultaneously in continuous time, let $]0, \tau_i[$, the interval of time in which the individual i is observed and n_i the number of events of individual i along $]0, \tau_i[$, then the likelihood function for the outcome n_i along $]0, \tau_i[$ is given by

$$L(\mathbf{\Phi}) = \prod_{i=1}^{n} L_i(\phi),$$

where

$$L_i(\phi) = \prod_{j=1}^{n_i} \rho_0 \left(T_{ij}, \boldsymbol{\alpha}\right) e^{\boldsymbol{\beta}' \mathbf{Y}_i} \, e^{-\int_0^\tau X_i(v)\rho_0(v,\boldsymbol{\alpha})e^{\boldsymbol{\beta}'\mathbf{Y}_i(v)dv}}. \tag{4.7}$$

In (4.7), $\mathbf{\Phi} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$; $\boldsymbol{\alpha}$ is called a baseline parameter, $\tau = \max(\tau_1, \tau_2, ...\tau_n)$ and

$$X_i(v) = \begin{cases} 1 & \text{if individual } i \text{ is at risk} \\ 0 & \text{otherwise.} \end{cases}$$

Using the relationship (4.4), the log-likelihood can be written as

$$\ln L(\Phi) = \sum_{i=1}^{n} \int_0^\tau X_i(v) \left[\ln \rho_i(v, \Phi) dN_i(v) - \rho_i(v, \Phi) dv\right].$$

The maximum likelihood estimates are obtained by solving a system

$$\begin{cases} \frac{\partial \ln L(\mathbf{\Phi})}{\partial \boldsymbol{\alpha}} = 0 \\ \frac{\partial \ln L(\mathbf{\Phi})}{\partial \boldsymbol{\beta}} = 0. \end{cases} \tag{4.8}$$

The numerical methods such as the Newton-Raphson method are used for solving system (4.8). The adequacy of parameters is checked by finding the elements $\boldsymbol{\mathcal{I}}_{\boldsymbol{\alpha\alpha}}$, $\boldsymbol{\mathcal{I}}_{\boldsymbol{\alpha\beta}}$, $\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\alpha}}$ and $\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\beta}}$ of the information matrix $\boldsymbol{\mathcal{I}}$ and assume that as $n \to \infty$, $\hat{\mathbf{\Phi}} - \mathbf{\Phi} \rightsquigarrow N\left(0, \boldsymbol{\mathcal{I}}^{-1}(\hat{\mathbf{\Phi}})\right)$ (Sankaran and Anisha, 2011).

4.4 Setup of dataset in AGM, PWPM and WLWM

Numerical examples on the layout of the dataset in the AGM, the PWPM and the WLWM are found in materials such as Hosmer et al. (2008, p. 289), Andersen et al. (1993), Fisher (1991), Lin and Ying (1994), O'Brien (1984), Lin and Wei (1992), Clayton and Cuzick (1985) and Kelly and Lim (2000). Assume that n is a maximum number of events per subject, and that τ_k, $k = 1, 2, \ldots n$, are times to events per subject along the study time with range $[0,\, T]$. Under the AGM, all events are assumed to be in one stratum along the study time. The study time T is subdivided into intervals defined by the times to events as $[0,\, \tau_1]\,;\,]\tau_1,\, \tau_2]\,;\ldots;\,]\tau_n,\, T]$, with event indicator for each time interval. The layout of the dataset for PWPM is the same as for the AGM where for each interval corresponds a specific stratum, making the number of time intervals per subject equal to the number of strata per subject. The alternative PWPM based on gap time takes 0 at the lower bound of each interval per subject, the upper bound is given by the gaps or $\tau_k - \tau_{k-1}$, $k = 1, 2, \ldots, n$, the first and the last intervals are respectively $[0,\, \tau_1]$ and $[0,\, T - \tau_n]$. Like in PWPM, the k^{th} time interval per subject in WLWM is in the k^{th} stratum, $k = 1, 2, \ldots, n$. In WLWM, the study time is subdivided into $n + 1$ intervals each with lower bound 0 and upper bound equal to the time to event, the first and the last intervals are respectively $[0,\, \tau_1]$ and $[0,\, T]$.

4.5 Application

Consider the data described in Table 2.1 regarding the survival times of 2117 newborns from the Kigali University Teaching Hospital (KUTH). The aim of this study was to apply a suitable

multiple events model for estimating the relative risk in each covariate.

Two events per subject are of interest: death and occurrence of at least one chronic disease or complication. The chronic disease or complications recorded at KUTH are severe oliguria, severe prematurity, very low birth weight, macrosomia, severe respiratory distress, gastroparesis, hemolytic, trisomy, asphyxia and laparoschisis.

The layout of the KUTH dataset follows the indication provided by the WLWM, Table 1 of Appendix A gives the first 50 entries, the full dataset can be found via the authors of this thesis.

Model (4.6) was implemented using STATA-15 and the dataset on infant mortality at KUTH with a portion given in Table 1 of Appendix A. The WLWM was used since death can occur without a previous chronic disease or complication and the two events could occur at the same time per subject.

Table 4.1: Unadjusted WLWM for the infant mortality at KUTH from 01-January-2016 to 31-December-2016 with Breslow method of ties handling.

Covariate (reference)	Level	Hazard Ratio	Std. Err	z	$P > z$	95% Conf. Int.
Age (Under 20 years old)	20 to 34 years old	0.277	0.997	-3.570	$p < 0.001$	[0.137; 0.560]
	35 years old and above	0.395	0.157	-2.330	0.020	[0.181; 0.863]
Residence (Rural)	Urban	0.847	0.139	-1.020	0.309	[0.614; 1.167]
Antecedents (Not 1st newborn)	1st new born	0.806	0.157	-1.100	0.270	[0.550; 1.182]
Abortion (Not aborted)	Aborted once	1.405	0.398	1.200	0.231	[0.806; 2.448]
	Aborted more than once	0.479	0.161	-2.190	0.028	[0.248; 0.925]
Childbirth (Ventouse)	Natural	0.873	0.491	-0.240	0.808	[0.290; 2.627]
	Surgery	1.115	0.613	0.200	0.843	[0.380; 3.274]
Gender (Female)	Male	1.740	0.296	3.260	0.001	[1.247; 2.429]
Number (Singleton)	Multiple	0.409	0.131	-2.790	0.005	[0.218; 0.766]
APGAR (Below 4/10)	4/10 to 6/10	0.377	0.112	-3.300	0.001	[0.211; 0.673]
	7/10 and above	0.130	0.036	-7.460	$p < 0.001$	[0.076; 0.222]
Weight (Under 2500 g)	2500 g to 4500 g	0.250	0.068	-5.070	$p < 0.001$	[0.146; 0.427]
	Above 4500 g	0.442	0.285	-1.270	0.206	[0.125; 1.565]
Head (Below 32 cm)	32 cm to 36 cm	0.456	0.128	-2.800	0.005	[0.263; 0.789]
	Above 36 cm	0.290	0.219	-1.640	0.102	[0.066; 1.278]
Height (Below 46 cm)	46 cm to 54 cm	0.894	0.276	-0.360	0.716	[0.488; 1.637]
	Above 54 cm	1.670	1.264	0.680	0.498	[0.379; 7.361]

Table 4.2: Unadjusted WLWM for the infant mortality at KUTH from 01-January-2016 to 31-December-2016 with Efron method of ties handling.

Covariate (reference)	Level	Hazard Ratio	Std. Err.	z	$P > z$	95% Conf. Int.
Age (Under 20 years old)	20 to 34 years old	0.230	0.083	-4.080	$p < 0.001$	[0.114; 0.466]
	35 years old and above	0.324	0.129	-2.840	0.005	[0.149; 0.706]
Residence (Rural)	Urban	0.831	0.137	-1.120	0.261	[0.602; 1.147]
Antecedents (Not 1st newborn)	1st newborn	0.756	0.149	-1.420	0.156	[0.513; 1.113]
Abortion (Not aborted)	Aborted once	1.393	0.396	1.170	0.244	[0.798; 2.430]
	Aborted more than once	0.452	0.154	-2.340	0.020	[0.232; 0.880]
Childbirth (Ventouse)	Natural	0.736	0.408	-0.550	0.580	[0.249; 2.179]
	Surgery	0.921	0.499	-0.150	0.880	[0.319; 2.661]
Gender (Female)	Male	1.823	0.312	3.520	$p < 0.001$	[1.304; 2.549]
Number (Singleton)	Multiple	0.324	0.106	-3.430	0.001	[0.170; 0.617]
APGAR (Below 4/10)	4/10 to 6/10	0.214	0.065	-5.090	$p < 0.001$	[0.118; 0.387]
	7/10 and above	0.070	0.020	-9.520	$p < 0.001$	[0.041; 0.121]
Weight (Under 2500 g)	2500 g to 4500 g	0.231	0.063	-5.340	$p < 0.001$	[0.135; 0.395]
	Above 4500 g	0.412	0.269	-1.360	0.174	[0.115; 1.479]
Head (Below 32 cm)	32 cm to 36 cm	0.422	0.119	-3.060	0.002	[0.243; 0.734]
	Above 36 cm	0.246	0.187	-1.840	0.065	[0.055; 1.093]
Height (Below 46 cm)	46 cm to 54 cm	0.917	0.285	-0.280	0.781	[0.499; 1.687]
	Above 54 cm	1.692	1.283	0.690	0.488	[0.383; 7.476]

Table 4.3: Unadjusted WLWM for the infant mortality at KUTH from 01-January-2016 to 31-December-2016 with Cox method of ties handling.

Covariate (reference)	Level	Hazard Ratio	Std. Err.	z	$P > z$	95% Conf. Int.
Age (Under 20 years old)	20 to 34 years old	0.193	0.085	-3.730	$p < 0.001$	[0.081; 0.458]
	35 years old and above	0.267	0.128	-2.760	0.006	[0.104; 0.682]
Residence (Rural)	Urban	0.766	0.150	-1.360	0.175	[0.521; 1.126]
Antecedents (Not 1st newborn)	1st newborn	0.763	0.185	-1.120	0.264	[0.475; 1.226]
Abortion (Not aborted)	Aborted once	1.404	0.453	1.050	0.293	[0.746; 2.643]
	Aborted more than once	0.378	0.152	-2.420	0.015	[0.172; 0.830]
Childbirth (Ventouse)	Natural	0.732	0.481	-0.470	0.635	[0.202; 2.653]
	Surgery	1.016	0.654	0.030	0.980	[0.288; 3.590]
Gender (Female)	Male	1.991	0.405	3.390	0.001	[1.336; 2.966]
Number (Singleton)	Multiple	0.218	0.111	-3.000	0.003	[0.080; 0.589]
APGAR (Below 4/10)	4/10 to 6/10	0.080	0.042	-4.810	$p < 0.001$	[0.029; 0.224]
	7/10 and above	0.021	0.011	-7.840	$p < 0.001$	[0.008; 0.056]
Weight (Under 2500 g)	2500 g to 4500 g	0.236	0.070	-4.850	$p < 0.001$	[0.131; 0.423]
	Above 4500 g	0.378	0.257	-1.430	0.153	[0.100; 1.436]
Head (Below 32 cm)	32 cm to 36 cm	0.391	0.119	-3.100	0.002	[0.216; 0.708]
	Above 36 cm	0.212	0.171	-1.920	0.055	[0.043; 1.033]
Height (Below 46 cm)	46 cm to 54 cm	0.828	0.283	-0.550	0.582	[0.423; 1.620]
	Above 54 cm	1.706	1.351	0.670	0.500	[0.361; 8.060]

Tables 4.1, 4.2 and 4.3 present the estimates of the hazard ratios of the unadjusted WLWM with ties handling by Breslow, Efron and Cox approaches, respectively for robustness check. The results in the latter two approaches were not far from that of the default method (Breslow). In all the three approaches, significant differences in levels were observed for the same covariates namely, the *age*, *abortion*, *gender*, *number*, *APGAR*, *weight* and *head* where p-values are less or equal to 0.05.

The adjusted WLWM with Breslow, Efron and Cox methods of ties handling are summarised in Table 4.4 and in Appendix A (Table 2 and Table 3) respectively. The results for different approaches are close.

Table 4.4: Adjusted WLWM for the infant mortality at KUTH from 01-January-2016 to 31-December-2016 with Breslow method of ties handling.

Covariate (reference)	Level	Hazard ratio	Std. Err.	z	P>z	95% Conf. Int.
Age (Under 20 years old)	20 to 34 years old	0.307	0.107	3.380	0.001	[0.155; 0.609]
	35 years old and above	0.472	0.179	-1.980	0.047	[0.225; 0.992]
Abortion (Not aborted)	Aborted once	1.482	0.406	1.430	0.152	[0.866; 2.537]
	Aborted more than once	0.541	0.175	-1.900	0.057	[0.287; 1.019]
Gender (Female)	Male	1.672	0.280	3.070	0.002	[1.204; 2.321]
Number (Singleton)	Multiple	0.401	0.128	-2.860	0.004	[0.214; 0.750]
APGAR (Below 4/10)	4/10 to 6/10	0.414	0.119	-3.080	0.002	[0.236; 0.726]
	7/10 and above	0.144	0.038	-7.350	$p < 0.001$	[0.086; 0.242]
Weight (Under 2500 g)	2500 g to 4500 g	0.238	0.060	-5.650	$p < 0.001$	[0.144; 0.391]
	Above 4500 g	0.447	0.284	-1.270	0.205	[0.129; 1.550]
Head (Below 32 cm)	32 cm to 36 cm	0.420	0.100	-3.660	0.000	[0.264; 0.669]
	Above 36 cm	0.284	0.210	-1.700	0.089	[0.067; 1.211]

The adjusted model by default (Breslow) suggests that the risk of death or attracting a chronic disease or complication of babies whose mothers are from 20 years and 34 years old is 0.307 times that of babies whose mothers are under 20 years old (95% CI:$0.155 - 0.609$, $p = 0.001$).

The risk of death or attracting a chronic disease or complication of babies whose mothers aborted more than once previously is 0.541 times that of babies whose mothers did not abort previously (95% CI:$0.287 - 1.019$, $p = 0.057$). The risk of death or attracting a chronic disease or complication of babies whose mothers are 35 years old and above is 0.472 times that of babies whose mothers are under 20 years old (95% CI:$0.225 - 0.992$, $p = 0.047$). The risk of death or attracting a chronic disease or complication for male babies is 1.672 times that of female babies (95% CI:$1.204 - 2.321$, $p = 0.002$). The risk of death or attracting a chronic disease or complication of multiple babies is 0.401 times that of singleton babies (95% CI:$0.214 - 0.750$, $p = 0.004$). The risk of death or attracting a chronic disease or complication for babies whose APGAR range from 4/10 to 6/10 is 0.414 times that of babies whose APGAR is below 4/10 (95% CI:$0.236 - 0.726$, $p = 0.002$). The risk of death or attracting a chronic disease or complication for babies whose APGAR range from 7/10 to 10/10 is 0.144 times that of babies whose APGAR is below 4/10 (95% CI:$0.086 - 0.242$, $p < 0.001$). The risk of death or attracting a chronic disease or complication for babies whose weight range from 2500 g to 4500 g is 0.238 times that of babies whose weight is below 2500 g (95% CI:$0.144 - 0.391$, $p < 0.001$). The risk of death or attracting a chronic disease or complication for babies whose circumference of the head range from 32 cm to 36 cm is 0.420 times that of babies whose circumference of the head is below 32 cm (95% CI:$0.264 - 0.669$, $p < 0.001$). The risk of death or attracting a chronic disease or complication for babies whose circumference of the head is above 36 cm is 0.284 times that of babies whose circumference of the head is below 32 cm (95% CI:$0.067 - 1.211$, $p = 0.067$).

4.6 Conclusion

This chapter reviewed different multiplicative multiple events regression models of the time to event survival data, namely, the mean function regression model, the rate function regression model and the intensity process regression model. The intensity process regression model incorporates the popular models such as Andersen- Gill Model (AGM), Wei, Lin and Weisfeld Model (WLWM) and Prentice, Williams and Peterson Model (PWPM) following on the layout of the dataset. It was found that data collected at Kigali University Teaching Hospital for 2117 newborns during 366 days of the year 2016 follows the conditions of the WLWM.

The results of the unadjusted WLWM by Breslow, Efron and Cox approaches of ties handling revealed significance on the age of mothers, information on previous abortion, the gender of the newborn, the number of newborns at a time, the APGAR, the weight of a newborn and the circumference of the head of a newborn. The results of the adjusted WLWM by Breslow, Efron and Cox approaches are close. The default approach (Breslow) indicated that the risk of death or attracting a chronic disease or clinical complication of infant was higher in male babies as compared to female babies; it was lower for babies whose mothers are from 20 to 34 years old and above 34 years old as compared to babies whose mothers are under 20 years old. Babies whose APGAR score falls in intervals 4/10 to 7/10 and 7/0 to 10/10 were found to have a better survival outcome than those born with APGAR score less than 4/10. Babies with normal weight and overweight, were found to have a lower risk as compared to underweight babies. Babies with a normal circumference of the head and those with a large circumference of the head were found to survive better than babies with a relatively small head.

This chapter extends single event models presented in Chapter 2 to multiple events models. The following chapter is devoted on using nonparametric re-sampling methods for measuring

consistency of the results of the CPHM developed in Chapter 2. This extension of the CPHM is limited on single event analysis while the extension of multiple events by applying re-sampling techniques follows from the next chapter.

CHAPTER 5

RE-SAMPLED COX PROPORTIONAL HAZARDS REGRESSION MODELS AND LENGTHS OF CONFIDENCE INTERVALS

5.1 Introduction

The re-sampling in Cox Proportional Hazards Model (CPHM) consists of conducting the CPHM on a given number of samples obtained after applying a relevant technique of re-sampling. The popular nonparametric techniques of re-sampling include a *bootstrap method* which is based on random sampling with replacement (Efron and Tibshirani, 1994), *jackknife method* which consists of making samples by leaving out one observation a time (Efron and Tibshirani, 1994), and *jackknife after bootstrap* (Efron, 1992). The interest in this study will be on Bootstrap Cox Proportional Hazards Model (BCPHM) and Jackknife Cox Proportional Hazards Model (JCPHM).

Hamada (1995) points out the aim of using the re-sampling technique in CPHM. Firstly the re-sampling allows the assessment of the stability of the CPHM. The instability may be caused by the correlation of the covariates. Secondly, the re-sampling may be used when the sample size is relatively small. Model adequacy may be satisfied by selecting variables on which the model is stable rather than testing the proportionality of variables.

BCPHM and JCPHM have been extensively applied to different studies. In Utzet and Sánchez (1992), bootstrap is applied for estimating the survival function and the hazard rate with respective standard errors. Bělašková et al. (2013) published a clinical study which uses BCPHM with consideration of right censoring and delayed entries. The study of Bělašková et al. adapts BCPHM due to the small sample size (N=61). Xu et al. (2014) conducted the BCPHM with consideration of a change-point along the study time with right censored survival data. The study proved consistency of the model by making a comparison with the model based on data simulation. The JCPHM was adopted by Xiao et al. (2012) together with a random weighting which consists of approximating the distribution of the maximum partial likelihood estimates in the CPHM (Wang et al., 2009; Zheng, 1987; Zheng and Tu, 1988). Several other manuscripts discuss the use of the re-sampled survival analysis such as James (1997), Quan and Tsai (1992), Sauerbrei and Schumacher (1992), Akritas (1986), Efron (1981), Hjort (1985) and Kim (1990).

In this chapter, due to the computing power of STATA-15 used in this thesis, the BCPHM with 1000 bootstrap replicates and the JCPHM were used and compared to the CPHM in modelling the risk of infant death at the Kigali University Teaching Hospital (KUTH) from 01-January-2016 to 31-December-2016 (Gatabazi et al., 2019a).

The present chapter also discusses the lengths of the confidence intervals in both CPHM, BCPHM and JCPHM. The length of a confidence interval is a function of the confidence level and the standard error and increases with the confidence level. In statistics, a narrow confidence interval reflects the accuracy in estimating a statistic of interest. Consequently, assuming that the confidence level is fixed, the minimum length of a confidence interval correspond to the minimum standard error (Gatabazi and Kabera, 2015). Many studies discuss different properties of a confidence interval and its length for a statistic of interest such as

for example Burr (1994) who discusses the confidence intervals for a Cox model in a re-sampled framework by considering three statistics namely, the hazard, the survival and the median survival functions; and Kivaranovich and Leeb (2019) who discuss the lengths of the confidence intervals in a post-model-selection with polyhedral constraints.

This chapter discusses the reason of high variability of the standard errors for specific covariates in re-sampled models as compared to the CPHM and generalises significance by means of stability of models.

The chapter comprises four sections including the introduction presented in Section 5.1. Section 5.2 presents the mathematical formulation of bootstrap and jackknife and their application in CPHM. Section 5.3 presents applications of re-sampled models to the infant mortality data and discusses the length of the confidence intervals in the Cox model.Section 5.4 gives a conclusion.

5.2 Nonparametric re-sampling techniques

5.2.1 Bootstrap method

Assume a sample

$$\mathbf{x} = x_1, x_2 \ldots, x_n,$$

where $x_{i,i\in[1,n]}$ are independent and identically distributed with distribution F_θ with θ, the statistical parameter of interest. Consider the distribution function F_{R_n} of a random variable $R_n(\mathbf{x}, F_\theta)$. A bootstrap method as described by Efron and Tibshirani (1994), consists of generating

$$\mathbf{x}^\star = \mathbf{x}^{\star 1}, \mathbf{x}^{\star 2} \ldots, \mathbf{x}^{\star B},$$

where $\mathbf{x}^{\star i}_{i\in[1,B]}$ are random samples of size n drawn with replacement from the sample $\mathbf{x}$. The variables of $\mathbf{x}^{\star i}_{i\in[1,B]}$ are independent and identically distributed with distribution $\widehat{F}_{\theta,n}$ given $\mathbf{x}$; $\widehat{F}_{\theta,n}$ is an estimator of F_θ from $\mathbf{x}$ and B is a number of bootstrap samples also (replications).

5.2.2 Bootstrap standard error

Assume B bootstrap samples $\mathbf{x}^{\star 1}, \mathbf{x}^{\star 2} \ldots, \mathbf{x}^{\star B}$. Efron and Tibshirani (1994) propose the estimated standard error of the bootstrap statistic of interest $\widehat{\theta}$ as

$$\widehat{se}_B = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left[\widehat{\theta}^\star(b) - \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta}^\star(b)\right]^2} \tag{5.1}$$

where $\widehat{\theta}^\star(b)$ is an estimate of the statistic of interest from the b^{th} bootstrap sample, $b = 1, 2, \ldots . B$.

5.2.3 Bootstrap Cox Proportional Hazards Model (BCPHM)

Assume a CPHM, $h(t|\mathbf{x}_i)$, over the p fixed covariates with values $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$ and the hazard function $h_0(t)$ when values of all covariates are zeros, that is

$$h(t|\mathbf{x}_i) = h_0(t)\exp(\boldsymbol{\beta}'\mathbf{x}_i) \tag{5.2}$$

(Collet, 2003), where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$ is a p-dimensional vector of model parameters.

Consider three approaches of approximating the partial likelihood in presence of tied events defined in Section 3.4.1 as Breslow, Efron, and Cox approaches.

The inference of model (5.2) based on bootstrap consists of applying model (5.2) to each of the B bootstrap samples $\mathbf{x}^{\star i}$, $\forall i \in [1, B]$ of covariates $\mathbf{x}_j$, $\forall j \in [1, p]$. Bootstrap model parameter estimation uses either Breslow, Efron or Cox approach. The bootstrap standard error is obtained by using Equation (5.1).

5.2.4 Jackknife method

Assume a sample

$$\mathbf{x} = \mathbf{x}_1, x_2 \ldots, \ \mathbf{x}_n,$$

where $\mathbf{x}_{j,j\in[1,n]}$ are the values of the covariate $\mathbf{x}$. Let θ be a statistic of interest. The jackknife samples consist of leaving out one observation at time, these are n samples $\mathbf{x}^{\star i} = (x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ $\forall i \in [1, n]$ (Efron and Tibshirani, 1994). The jackknife

standard error estimate as propose Efron and Tibshirani (1994), is

$$\widehat{se}_{jack} = \sqrt{\frac{n-1}{n}\sum_{i=1}^{n}\left[\widehat{\theta^{\star}}(i) - \frac{1}{n}\sum_{i=1}^{n}\widehat{\theta^{\star}}(i)\right]^2} \tag{5.3}$$

where $\theta^{\star}(i)$, $i \in [1, n]$ is a statistic of interest for the i^{th} jackknife sample.

5.2.5 Jackknife Cox Proportional Hazards Model (JCPHM)

Model (5.2) based on jackknife was made by applying it to each of the n jackknife samples $\mathbf{x}^{\star i}$, $\forall i \in [1, n]$ of covariates $\mathbf{x}_j$, $\forall j \in [1, p]$. Either Breslow, Efron or Cox approach can be used for estimating the jackknife model parameters, with standard error given by Equation (5.3).

5.3 Application

5.3.1 Re-sampled Cox models

Consider the data described in Table 2.1 regarding the survival times of 2117 newborns from the Kigali University Teaching Hospital (KUTH). The aim of this study is to apply the CPHM on bootstrap and jackknife samples of covariates. The relative risk in each covariate and related standard error was estimated and compared to those obtained by the usual CPHM. The significance in difference of levels of covariates was also explored in both CPHM and re-sampled CPHM.

STATA-15 displays the results in three tables: Table 5.1 presents the estimates of unadjusted

CPHM, BCPHM, JCPHM and corresponding adjusted models, by using Breslow estimation method. Both unadjusted and adjusted CPHM, BCPHM and JCPHM by Efron and Cox estimation are also presented in Tables 5.2 and 5.3. The results displayed by the JCPHM are relatively close to that of the CPHM (Table 5.1).

Standard errors in JCPHM and CPHM for all covariates are everywhere similar except for the upper levels of covariates *weight*, *head* and *height* where a standard error in JCPHM is more than 40 times that of CPHM. The critical difference in standard errors is also observed in BCPHM for the upper levels of covariates *weight*, *head* and *height*, for all levels of covariate *childbirth* and for the covariate *number* where a standard error is relatively higher in BCPHM. Also BCPHM does not take *age* and *number* as significant covariates unlike the fact of JCPHM and CPHM where these covariates are included in significant covariates. Following the arguments of Parzen and Lipsitz (1999), the χ^2 test statistics suggest a higher performance of the JCPHM as compared to the CPHM and BCPHM since the value of the χ^2 is relatively everywhere lower for the JCPHM.

The results by different approaches of ties handling are close as expected. The analysis is then made on the STATA-15 default method as proposed by Breslow (1974) . The similarity observed between the results of JCPHM and that of CPHM is strong as compared to that of BCPHM and CPHM. The similarity between CPHM and JCPHM suggests that the CPHM may be stable.

The overall analysis confirms the significance difference of levels of covariates *age*, *gender*, *number*, *APGAR*, *weight* and *head*. The results show a relatively higher risk of babies from under 20 years old mothers as compared to the older mothers, that is 4.651 times that of babies whose mothers' ages range from 20 to 34 years old; and 3.247 times that of babies whose mothers are 35 years old and above. The risk of male babies is 1.942 times that of

female babies. The risk of multiple babies is 0.264 times that of singleton babies. Babies with APGAR below 4/10 are at a relatively higher risk, that is 2.433 times that of babies with APGAR ranging from 4/10 to 6/10 and 16.949 times that of babies whose APGAR range from 7/10 to 10/10. The risk of babies whose weight is below 2500g is 5.525 times that of babies whose weight range from 2500g to 4500g and 2.688 times that of babies with weight above 4500g. The risk for babies born with the circumference of the head below 32cm is 4.808 times that of newborns whose circumference of head range from 32cm to 36cm; and 9.524 times that of newborns whose circumference of the head is above 36cm.

The results of BCPHM are also close to that of JCPHM and CPHM for all significant covariates but the model shows relatively high standard errors for non-significant levels of covariates. The critical discrepancy between standard errors after re-sampling for some covariates suggests instability of the CPHM at these specific covariates and this emphasizes their non-significance in the CPHM.

The stability of the adjusted CPHM is justified by the non-critical difference between the adjusted re-sampled models.

Table 5.1: Breslow estimation

CPHM							BCPHM					JCPHM				
Covariate (reference)	Level	HR	SE	z	P>z	95% CI	HR	SE	z	P>z	95% CI	HR	SE	z	P>z	95% CI
Age (Under 20 years old)	20 to 34 years old	0.172	0.086	-3.540	$p < 0.001$	[0.065; 0.456]	0.172	0.254	-1.190	0.234	[0.009; 3.124]	0.172	0.089	-3.400	0.001	[0.062; 0.475]
	35 years old and above	0.216	0.117	-2.840	0.005	[0.075; 0.623]	0.216	0.323	-1.020	0.306	[0.012; 4.058]	0.216	0.124	-2.660	0.008	[0.070; 0.667]
Residence (Rural)	Urban	1.014	0.240	0.060	0.954	[0.637; 1.614]	1.014	0.277	0.050	0.960	[0.594; 1.732]	1.014	0.285	0.050	0.961	[0.585; 1.758]
Antecedents (Not 1st newborn)	1st newborn	0.778	0.221	-0.880	0.377	[0.446; 1.358]	0.778	0.223	-0.880	0.381	[0.444; 1.364]	0.778	0.218	-0.900	0.370	[0.449; 1.347]
Abortion (Not aborted)	Aborted once	1.646	0.648	1.270	0.206	[0.761; 3.562]	1.646	0.695	1.180	0.238	[0.720; 3.763]	1.646	0.664	1.230	0.217	[0.746; 3.633]
	Aborted more than once	1.111	0.503	0.230	0.817	[0.457; 2.700]	1.111	2.084	0.060	0.955	[0.028; 43.927]	1.111	0.556	0.210	0.834	[0.416; 2.966]
Childbirth (Ventouse)	Natural	0.593	0.449	-0.690	0.490	[0.135; 2.612]	0.593	3.846	-0.080	0.936	$[0.000; 1.963 \times 10^{5}]$	0.593	0.469	-0.660	0.509	[0.126; 2.797]
	Surgery	0.777	0.580	-0.340	0.736	[0.180; 3.358]	0.777	5.021	-0.040	0.969	$[0.000; 2.443 \times 10^{5}]$	0.777	0.611	-0.320	0.749	[0.166; 3.630]
Gender (Female)	Male	1.964	0.472	2.810	0.005	[1.227; 3.146]	1.964	0.480	2.760	0.006	[1.217; 3.170]	1.964	0.504	2.630	0.009	[1.188; 3.248]
Number (Singleton)	Multiple	0.306	0.136	-2.660	0.008	[0.128; 0.732]	0.306	0.730	-0.500	0.620	[0.003; 32.826]	0.306	0.136	-2.670	0.008	[0.128; 0.729]
APGAR (Below 4/10)	4/10 to 6/10	0.335	0.133	-2.760	0.006	[0.154; 0.729]	0.335	0.160	-2.290	0.022	[0.131; 0.856]	0.335	0.157	-2.340	0.020	[0.134; 0.839]
	7/10 and above	0.049	0.019	-7.860	$p < 0.001$	[0.023; 0.103]	0.049	0.020	-7.300	$p < 0.001$	[0.022; 0.110]	0.049	0.020	-7.380	$p < 0.001$	[0.022; 0.109]
Weight (Under 2500 g)	2500 g to 4500 g	0.227	0.089	-3.790	$p < 0.001$	[0.105; 0.489]	0.227	0.102	-3.300	0.001	[0.094; 0.548]	0.227	0.105	-3.210	0.001	[0.091; 0.561]
	Above 4500 g	0.392	0.421	-0.870	0.383	[0.048; 3.213]	0.392	8.103	-0.050	0.964	$[0.000; 1.600 \times 10^{17}]$	0.392	17.310	-0.020	0.983	$[0.000; 1.740 \times 10^{37}]$
Head (Below 32 cm)	32 cm to 36 cm	0.288	0.111	-3.230	0.001	[0.136; 0.613]	0.288	0.121	-2.960	0.003	[0.127; 0.658]	0.288	0.116	-3.090	0.002	[0.131; 0.635]
	Above 36 cm	0.122	0.128	-2.010	0.045	[0.016; 0.951]	0.122	2.449	-0.100	0.917	$[0.000; 1.670 \times 10^{16}]$	0.122	5.426	-0.050	0.962	$[0.000; 1.220 \times 10^{37}]$
Height (Below 36 cm)	46 cm to 54 cm	0.567	0.235	-1.370	0.171	[0.251; 1.278]	0.567	0.240	-1.340	0.180	[0.247; 1.300]	0.567	0.247	-1.300	0.193	[0.241; 1.334]
	Above 54 cm	1.020	1.100	0.020	0.986	[0.123; 8.444]	1.020	21.073	0.000	0.999	$[0.000; 3.980 \times 10^{17}]$	1.020	44.687	0.000	1.000	$[0.000; 2.150 \times 10^{37}]$
Adjusted CPHM							Adjusted BCPHM					Adjusted JCPHM				
Age (Under 20 years old)	20 to 34 years old	0.215	0.105	-3.150	0.002	[0.083; 0.559]	-	-	-	-	-	0.215	0.104	-3.190	0.001	[0.084; 0.554]
	35 years old and above	0.308	0.159	-2.280	0.023	[0.112; 0.848]	-	-	-	-	-	0.308	0.160	-2.270	0.023	[0.111; 0.852]
Gender (Female)	Male	1.942	0.459	2.810	0.005	[1.222; 3.085]	1.562	0.350	1.990	0.046	[1.007; 2.424]	1.942	0.476	2.700	0.007	[1.200; 3.142]
Number (Singleton)	Multiple	0.264	0.115	-3.060	0.002	[0.112; 0.619]	-	-	-	-	-	0.264	0.117	-3.010	0.003	[0.111; 0.629]
APGAR (Below 4/10)	4/10 to 6/10	0.411	0.154	-2.380	0.017	[0.198; 0.856]	0.695	0.288	-0.880	0.379	[0.308; 1.565]	0.411	0.185	-1.970	0.049	[0.170; 0.995]
	7/10 and above	0.059	0.021	-7.850	$p < 0.001$	[0.029; 0.119]	0.100	0.039	-5.880	$p < 0.001$	[0.046; 0.215]	0.059	0.024	-6.810	$p < 0.001$	[0.026; 0.133]
Weight (Under 2500 g)	2500 g to 4500 g	0.181	0.064	-4.860	$p < 0.001$	[0.091; 0.361]	0.200	0.084	-3.840	$p < 0.001$	[0.088; 0.455]	0.181	0.071	-4.390	$p < 0.001$	[0.084; 0.389]
	Above 4500 g	0.372	0.384	-0.960	0.338	[0.049; 2.809]	0.438	8.985	-0.040	0.968	$[0.000; 1.280 \times 10^{17}]$	0.372	16.296	-0.020	0.982	$[0.000; 6.880 \times 10^{36}]$
Head (Below 32 cm)	32 cm to 36 cm	0.208	0.068	-4.830	$p < 0.001$	[0.110; 0.394]	0.216	0.088	-3.760	$p < 0.001$	[0.097; 0.480]	0.208	0.080	-4.060	$p < 0.001$	[0.098; 0.444]
	Above 36 cm	0.105	0.109	-2.180	0.029	[0.014; 0.797]	0.109	2.234	-0.110	0.914	$[0.000; 2.600 \times 10^{16}]$	0.105	4.680	-0.050	0.960	$[0.000; 9.160 \times 10^{36}]$
		$\chi^2 = 300.360$, $p < 0.001$					$\chi^2 = 296.290$, $p < 0.001$					$\chi^2 = 32.310$, $p < 0.001$				

Table 5.2: Efron estimation

CPHM							BCPHM					JCPHM				
Covariate (reference)	Level	HR	SE	z	P>z	95% CI	HR	SE	z	P>z	95% CI	HR	SE	z	P>z	95% CI
Age (Under 20 years old)	20 to 34 years old	0.160	0.079	-3.680	$p < 0.001$	[0.060; 0.424]	0.160	0.323	-0.910	0.364	[0.003; 8.374]	0.160	0.087	-3.370	0.001	[0.055; 0.464]
	35 years old and above	0.199	0.107	-2.990	0.003	[0.069; 0.573]	0.199	0.406	-0.790	0.429	[0.004; 10.896]	0.199	0.120	-2.680	0.007	[0.061; 0.648]
Residence (Rural)	Urban	1.029	0.246	0.120	0.907	[0.643; 1.645]	1.029	0.307	0.090	0.925	[0.573; 1.847]	1.029	0.314	0.090	0.927	[0.565; 1.871]
Antecedents (Not 1st newborn)	1st newborn	0.723	0.212	-1.110	0.268	[0.407; 1.283]	0.723	0.227	-1.030	0.301	[0.391; 1.337]	0.723	0.233	-1.010	0.314	[0.384; 1.359]
Abortion (Not aborted)	Aborted once	1.588	0.628	1.170	0.242	[0.732; 3.448]	1.588	0.696	1.060	0.291	[0.673; 3.749]	1.588	0.659	1.110	0.265	[0.704; 3.585]
	Aborted more than once	1.147	0.519	0.300	0.762	[0.473; 2.782]	1.147	4.651	0.030	0.973	$[0.000; 3.251 \times 10^{3}]$	1.147	0.587	0.270	0.789	[0.420; 3.127]
Childbirth (Ventouse)	Natural	0.532	0.400	-0.840	0.401	[0.122; 2.319]	0.532	3.646	-0.090	0.927	$[0.000; 3.605 \times 10^{5}]$	0.532	0.448	-0.750	0.454	[0.102; 2.772]
	Surgery	0.695	0.515	-0.490	0.624	[0.163; 2.969]	0.695	4.766	-0.050	0.958	$[0.000; 4.743 \times 10^{5}]$	0.695	0.579	-0.440	0.663	[0.136; 3.558]
Gender (Female)	Male	2.061	0.500	2.980	0.003	[1.282; 3.315]	2.061	0.556	2.680	0.007	[1.215; 3.496]	2.061	0.592	2.520	0.012	[1.173; 3.621]
Number (Singleton)	Multiple	0.243	0.113	-3.040	0.002	[0.098; 0.606]	0.243	0.135	-2.540	0.011	[0.082; 0.724]	0.243	0.141	-2.440	0.015	[0.078; 0.759]
APGAR (Below 4/10)	4/10 to 6/10	0.207	0.084	-3.880	$p < 0.001$	[0.094; 0.460]	0.207	0.116	-2.820	0.005	[0.070; 0.618]	0.207	0.120	-2.710	0.007	[0.066; 0.648]
	7/10 and above	0.030	0.012	-8.960	$p < 0.001$	[0.014; 0.065]	0.030	0.015	-7.070	$p < 0.001$	[0.011; 0.080]	0.030	0.016	-6.750	$p < 0.001$	[0.011; 0.083]
Weight (Under 2500 g)	2500 g to 4500 g	0.222	0.088	-3.800	$p < 0.001$	[0.102; 0.483]	0.222	0.105	-3.180	0.001	[0.088; 0.562]	0.222	0.107	-3.110	0.002	[0.086; 0.574]
	Above 4500 g	0.389	0.426	-0.860	0.389	[0.045; 3.338]	0.389	8.081	-0.050	0.964	$[0.000; 1.950 \times 10^{17}]$	0.389	17.369	-0.020	0.983	$[0.000; 4.530 \times 10^{37}]$
Head (Below 32 cm)	32 cm to 36 cm	0.284	0.110	-3.250	0.001	[0.133; 0.607]	0.284	0.115	-3.100	0.002	[0.129; 0.629]	0.284	0.119	-3.000	0.003	[0.125; 0.647]
	Above 36 cm	0.110	0.117	-2.070	0.038	[0.014; 0.886]	0.110	2.350	-0.100	0.918	$[0.000; 1.590 \times 10^{17}]$	0.110	3.679	-0.070	0.947	$[0.000; 3.080 \times 10^{27}]$
Height (Below 36 cm)	46 cm to 54 cm	0.569	0.238	-1.350	0.177	[0.251; 1.291]	0.569	0.252	-1.270	0.202	[0.239; 1.354]	0.569	0.273	-1.180	0.240	[0.222; 1.457]
	Above 54 cm	1.010	1.094	0.010	0.993	[0.121; 8.431]	1.010	21.269	0.000	1.000	$[0.000; 1.8.460 \times 10^{17}]$	1.010	44.776	0.000	1.000	$[0.000; 5.730 \times 10^{37}]$
Adjusted CPHM							Adjusted BCPHM					Adjusted JCPHM				
Age (Under 20 years old)	20 to 34 years old	0.201	0.098	-3.280	0.001	[0.077; 0.524]	-	-	-	-	-	0.201	0.102	-3.170	0.002	[0.075; 0.543]
	35 years old and above	0.293	0.152	-2.360	0.018	[0.106; 0.811]	-	-	-	-	-	0.293	0.160	-2.250	0.025	[0.101; 0.856]
Gender (Female)	Male	2.071	0.495	3.050	0.002	[1.297; 3.308]	1.562	0.400	1.740	0.081	[0.946; 2.579]	2.071	0.587	2.570	0.010	[1.188; 3.611]
Number (Singleton)	Multiple	0.205	0.092	-3.520	$p < 0.001$	[0.085; 0.495]	-	-	-	-	-	0.205	0.118	-2.740	0.006	[0.066; 0.637]
APGAR (Below 4/10)	4/10 to 6/10	0.273	0.103	-3.430	0.001	[0.130; 0.573]	0.545	0.273	-1.210	0.226	[0.204; 1.457]	0.273	0.169	-2.100	0.036	[0.081; 0.919]
	7/10 and above	0.038	0.014	-8.980	$p < 0.001$	[0.019; 0.078]	0.077	0.036	-5.440	$p < 0.001$	[0.030; 0.193]	0.038	0.023	-5.530	$p < 0.001$	[0.012; 0.122]
Weight (Under 2500 g)	2500 g to 4500 g	0.179	0.063	-4.890	$p < 0.001$	[0.090; 0.356]	0.201	0.083	-3.880	0.000	[0.089; 0.452]	0.179	0.071	-4.360	$p < 0.001$	[0.082; 0.388]
	Above 4500 g	0.379	0.396	-0.930	0.353	[0.049; 2.938]	0.477	9.872	-0.040	0.971	$[0.000; 2.040 \times 10^{17}]$	0.379	16.849	-0.020	0.983	$[0.000; 2.970 \times 10^{37}]$
Head (Below 32 cm)	32 cm to 36 cm	0.205	0.067	-4.860	$p < 0.001$	[0.108; 0.388]	0.215	0.090	-3.680	$p < 0.001$	[0.095; 0.487]	0.205	0.081	-4.030	$p < 0.001$	[0.095; 0.443]
	Above 36 cm	0.095	0.100	-2.250	0.025	[0.012; 0.740]	0.105	2.180	-0.110	0.914	$[0.000; 5.960 \times 10^{16}]$	0.095	4.226	-0.050	0.958	$[0.000; 5.340 \times 10^{36}]$
		$\chi^2 = 316.160$, $p < 0.001$					$\chi^2 = 297.200$, $p < 0.001$					$\chi^2 = 29.760$, $p < 0.001$				

Table 5.3: Cox estimation

CPHM							BCPHM					JCPHM				
Covariate (reference)	Level	HR	SE	z	P>z	95% CI	HR	SE	z	P>z	95% CI	HR	SE	z	P>z	95% CI
Age (Under 20 years old)	20 to 34 years old	0.140	0.075	-3.690	$p < 0.001$	[0.050; 0.398]	0.140	0.257	-1.070	0.283	[0.004; 5.064]	0.140	0.084	-3.260	0.001	[0.043; 0.457]
	35 years old and above	0.171	0.098	-3.090	0.002	[0.056; 0.523]	0.171	0.313	-0.960	0.335	[0.005; 6.216]	0.171	0.111	-2.710	0.007	[0.048; 0.613]
Residence (Rural)	Urban	1.003	0.258	0.010	0.990	[0.606; 1.660]	1.003	0.347	0.010	0.993	[0.510; 1.974]	1.003	0.342	0.010	0.993	[0.514; 1.956]
Antecedents (Not 1st newborn)	1st newborn	0.726	0.231	-1.010	0.313	[0.389; 1.353]	0.726	0.280	-0.830	0.406	[0.341; 1.545]	0.726	0.268	-0.870	0.386	[0.351; 1.498]
Abortion (Not aborted)	Aborted once	1.671	0.686	1.250	0.211	[0.748; 3.735]	1.671	0.763	1.120	0.261	[0.683; 4.091]	1.671	0.722	1.190	0.234	[0.717; 3.897]
	Aborted more than once	1.388	0.697	0.650	0.514	[0.519; 3.712]	1.388	0.756	0.600	0.548	[0.477; 4.038]	1.388	0.849	0.540	0.593	[0.418; 4.609]
Childbirth (Ventouse)	Natural	0.533	0.422	-0.790	0.427	[0.113; 2.517]	0.533	3.473	-0.100	0.923	$[0.000; 1.883 \times 10^{5}]$	0.533	0.449	-0.750	0.456	[0.102; 2.786]
	Surgery	0.759	0.590	-0.360	0.722	[0.166; 3.479]	0.759	4.946	-0.040	0.966	$[0.000; 2.683 \times 10^{5}]$	0.759	0.628	-0.330	0.739	[0.150; 3.850]
Gender (Female)	Male	2.195	0.570	3.030	0.002	[1.319; 3.652]	2.195	0.672	2.570	0.010	[1.204; 3.999]	2.195	0.695	2.480	0.013	[1.179; 4.086]
Number (Singleton)	Multiple	0.203	0.110	-2.950	0.003	[0.071; 0.585]	0.203	0.693	-0.470	0.640	[0.000; 162.000]	0.203	0.196	-1.650	0.099	[0.031; 1.353]
APGAR (Below 4/10)	4/10 to 6/10	0.167	0.085	-3.500	$p < 0.001$	[0.061; 0.455]	0.167	0.602	-0.500	0.620	[0.000; 197.300]	0.167	0.180	-1.660	0.098	[0.020; 1.392]
	7/10 and above	0.022	0.010	-8.140	$p < 0.001$	[0.009; 0.055]	0.022	0.078	-1.070	0.284	[0.000; 24.091]	0.022	0.021	-3.880	$p < 0.001$	[0.003; 0.151]
Weight (Under 2500 g)	2500 g to 4500 g	0.221	0.088	-3.790	$p < 0.001$	[0.101; 0.482]	0.221	0.105	-3.180	0.001	[0.087; 0.560]	0.221	0.105	-3.170	0.002	[0.087; 0.562]
	Above 4500 g	0.324	0.362	-1.010	0.313	[0.036; 2.892]	0.324	6.266	-0.060	0.954	$[0.000; 9.150 \times 10^{15}]$	0.324	10.526	-0.030	0.972	$[0.000; 1.450 \times 10^{27}]$
Head (Below 32 cm)	32 cm to 36 cm	0.285	0.110	-3.240	0.001	[0.133; 0.609]	0.285	0.119	-3.020	0.003	[0.126; 0.644]	0.285	0.117	-3.050	0.002	[0.127; 0.639]
	Above 36 cm	0.106	0.114	-2.090	0.036	[0.013; 0.866]	0.106	2.091	-0.110	0.909	$[0.000; 5.660 \times 10^{15}]$	0.106	3.780	-0.060	0.950	$[0.000; 1.910 \times 10^{29}]$
Height (Below 36 cm)	46 cm to 54 cm	0.539	0.226	-1.480	0.140	[0.237; 1.225]	0.539	0.236	-1.410	0.158	[0.229; 1.270]	0.539	0.252	-1.320	0.186	[0.216; 1.346]
	Above 54 cm	1.037	1.120	0.030	0.973	[0.125; 8.613]	1.037	20.074	0.000	0.998	$[0.000; 3.080 \times 10^{16}]$	1.037	41.221	0.000	0.999	$[0.000; 7.290 \times 10^{33}]$
Adjusted CPHM							Adjusted BCPHM					Adjusted JCPHM				
Age (Under 20 years old)	20 to 34 years old	0.173	0.092	-3.310	0.001	[0.061; 0.488]	-	-	-	-	-	0.181	0.096	-3.230	0.001	[0.064; 0.511]
	35 years old and above	0.250	0.139	-2.490	0.013	[0.084; 0.745]	-	-	-	-	-	0.248	0.139	-2.490	0.013	[0.083; 0.744]
Gender (Female)	Male	2.150	0.550	2.990	0.003	[1.302; 3.549]	2.031	0.473	3.050	0.002	[1.287; 3.205]	1.778	0.506	2.020	0.043	[1.018; 3.106]
Number (Singleton)	Multiple	0.176	0.091	-3.350	0.001	[0.064; 0.486]	-	-	-	-	-	-	-	-	-	-
APGAR (Below 4/10)	4/10 to 6/10	0.249	0.114	-3.030	0.002	[0.101; 0.612]	-	-	-	-	-	0.516	0.330	-1.030	0.301	[0.147; 1.809]
	7/10 and above	0.030	0.013	-8.220	$p < 0.001$	[0.013; 0.069]	-	-	-	-	-	0.060	0.035	-4.820	$p < 0.001$	[0.019; 0.188]
Weight (Under 2500 g)	2500 g to 4500 g	0.176	0.062	-4.910	$p < 0.001$	[0.088; 0.352]	0.149	0.053	-5.380	$p < 0.001$	[0.075; 0.299]	0.209	0.082	-3.990	$p < 0.001$	[0.097; 0.451]
	Above 4500 g	0.325	0.347	-1.050	0.293	[0.040; 2.636]	0.367	6.399	-0.060	0.954	$[0.000; 2.450 \times 10^{14}]$	0.425	16.781	-0.020	0.983	$[0.000; 1.840 \times 10^{33}]$
Head (Below 32 cm)	32 cm to 36 cm	0.196	0.064	-5.020	$p < 0.001$	[0.103; 0.370]	0.120	0.038	-6.700	$p < 0.001$	[0.065; 0.224]	0.198	0.077	-4.180	$p < 0.001$	[0.093; 0.423]
	Above 36 cm	0.090	0.095	-2.290	0.022	[0.011; 0.706]	0.073	1.284	-0.150	0.882	$[0.000; 7.170 \times 10^{13}]$	0.098	4.324	-0.050	0.958	$[0.000; 3.360 \times 10^{36}]$
		$\chi^2 = 316.430$, $p < 0.001$					$\chi^2 = 210.070$, $p < 0.001$					$\chi^2 = 31.380$, $p < 0.001$				

5.3.2 Lengths of the confidence intervals in CPHM, BCPHM and JCPHM

Table 5.4 presents the Breslow estimates of the adjusted CPHM, BCPHM and JCPHM with the confidence intervals of the relative hazards in the last column of the table.

The lengths of the confidence intervals of the relative hazards are relatively narrow and approximately the same for all the three models for covariates *gender*, *APGAR* and the intermediate levels of covariates *weight* and *head*. The two re-sampled models reveal that the upper levels of covariates *weight* and *head* are not significant as the confidence interval is infinitely wide due to the observed rapid increase of the standard error of the relative hazards.

The study shows that re-sampling contradicts the significance found at the upper lever of covariate *head* in the CPHM, but also re-sampled models strongly confirms non-significance of the upper level of covariate *weight* in CPHM.

Table 5.4: Breslow estimation

CPHM							
Covariate (reference)	Level	HR	SE	Z	PV	95% CI	L
Age (Under 20 years old)	20 to 34 years old	0.215	0.105	-3.150	0.002	[0.083; 0.559]	0.476
	35 years old and above	0.308	0.159	-2.280	0.023	[0.112; 0.848]	0.736
Gender (Female)	Male	1.942	0.459	2.810	0.005	[1.222; 3.085]	1.862
Number (Singleton)	Multiple	0.264	0.115	-3.060	0.002	[0.112; 0.619]	0.507
APGAR (Below 4/10)	4/10 to 6/10	0.411	0.154	-2.380	0.017	[0.198; 0.856]	0.658
	7/10 and above	0.059	0.021	-7.850	$p < 0.01$	[0.029; 0.119]	0.090
Weight (Under 2500g)	2500 g to 4500 g	0.181	0.064	-4.860	$p < 0.001$	[0.091; 0.361]	0.270
	Above 4500 g	0.372	0.384	-0.960	0.338	[0.049; 2.809]	2.760
Head (Below 32 cm)	32 cm to 36 cm	0.208	0.068	-4.830	$p < 0.001$	[0.110; 0.394]	0.284
	Above 36 cm	0.105	0.109	-2.180	0.029	[0.014; 0.797]	0.783
BCPHM							
Age (Under 20 years old)	20 to 34 years old						
	35 years old and above						
Gender (Female)	Male	1.562	0.350	1.990	0.046	[1.007; 2.424]	1.417
Number (Singleton)	Multiple						
APGAR (Below 4/10)	4/10 to 6/10	0.695	0.288	-0.880	0.379	[0.308; 1.565]	1.257
	7/10 and above	0.100	0.039	-5.880	$p < 0.001$	[0.046; 0.215]	0.169
Weight (Under 2500g)	2500 g to 4500 g	0.200	0.084	-3.840	$p < 0.001$	[0.088; 0.455]	0.367
	Above 4500 g	0.438	8.985	-0.040	0.968	[0.000; 1.280*10^17]	1.280*10^17
Head (Below 32 cm)	32 cm to 36 cm	0.216	0.088	-3.760	$p < 0.001$	[0.097; 0.480]	0.383
	Above 36 cm	0.109	2.234	-0.110	0.914	[0.000; 2.600*10^16]	2.600*10^16
JCPHM							
Age (Under 20 years old)	20 to 34 years old	0.215	0.104	-3.190	0.001	[0.084; 0.554]	0.470
	35 years old and above	0.308	0.160	-2.270	0.023	[0.111; 0.852]	0.741
Gender (Female)	Male	1.942	0.476	2.700	0.007	[1.200; 3.142]	1.942
Number (Singleton)	Multiple	0.264	0.117	-3.010	0.003	[0.111; 0.629]	0.518
APGAR (Below 4/10)	4/10 to 6/10	0.411	0.185	-1.970	0.049	[0.170; 0.995]	0.825
	7/10 and above	0.059	0.024	-6.810	$p < 0.001$	[0.026; 0.133]	0.107
Weight (Under 2500g)	2500 g to 4500 g	0.181	0.071	-4.390	$p < 0.001$	[0.084; 0.389	0.304
	Above 4500 g	0.372	16.296	-0.020	0.982	[0.000; 6.880*10^36]	6.880*10^36
Head (Below 32 cm)	32 cm to 36 cm	0.208	0.080	-4.060	$p < 0.001$	[0.098; 0.444]	0.346
	Above 36 cm	0.105	4.680	-0.050	0.960	[0.000; 9.160*10^36]	9.160*10^36

5.4 Conclusion

This chapter reviewed different methods of re-sampling in the Cox Proportional hazards Model (CPHM) namely, the Bootstrap Cox Proportional Hazards Model (BCPHM) and the Jackknife Cox Proportional Hazards Model (JCPHM) and discussed the lengths of the confidence intervals in CPHM, BCPHM and JCPHM. The results after re-sampling are compared to that of the CPHM for three different ties handling methods namely, Breslow, Efron and Cox approximation. The χ^2 test statistics show everywhere a higher performance of the JCPHM as compared to the CPHM and BCPHM.

The results displayed by the JCPHM and CPHM are very close and suggested significance on covariates *age*, *abortion*, *gender*, *number*, *APGAR*, *weight* and *head*. Male babies are at a relatively higher risk as compared to female babies. The risk is higher for babies whose mothers are under 20 years old as compared to older mothers. Babies born with APGAR less than 4/10 were found to have a higher risk as compared to newborns with APGAR greater than 4/10. Underweight babies were found to have a higher risk as compared to babies with normal weight and overweight. Babies with a normal circumference of the head were found to survive better than those with a relatively big head and a relatively small head. Underheight babies were found to have a higher risk as compared to babies born with normal height and over-height newborns. The results of the BCPHM are not far from that of JCPHM and CPHM but the non-significant covariates displayed relatively higher standard errors, leading to the relatively wide lengths of the corresponding confidence intervals. The overall results of the re-sampled models showed a relatively higher standard error per non-significant covariate. Due to a relatively higher risk to death of infant from under 20 years old mothers, pregnancy of mothers belonging in such range of age should be avoided. Also as abnormality leads to a

relatively higher risk to the infant mortality, clinically recommended nutrition during pregnancy would decrease abnormality of a newborn; decreasing then the infant mortality.

The present chapter uses the nonparametric re-sampled model for measuring the consistency of the results of the CPHM presented in Chapter 2. The results of the re-sampled models with multiple events are presented in the next chapter.

CHAPTER 6

RE-SAMPLED MARGINAL RISK SET MODEL

6.1 Introduction

The multiple events model for the infant mortality at the Kigali University Teaching hospital analysed in Chapter 4 leaves a question on the stability of the adopted model. The analysis used the primary dataset of the year 2016 with two events per subject, namely, death and the occurrence of at least one of the common conditions that may also cause long term death to infants. It was found that the Marginal Risk Set Model (MRSM) also known as the Wei, Lin and Weissfeld Model (WLWM) is appropriate for fitting the data. The WLWM is among the multiplicative methods for analysing ordered events found in Cook and Lawless (2007). Other multiplicative models include the Andersen-Gill Model (AGM) and the Prentice, Williams and Peterson Model (PWPM) (Wei and Glidden, 1997).

The AGM is known also as the counting process approach (Andersen and Gill, 1992). The AGM assumes that all event types are not different and all events within the same subject are assumed to be independent (Johnson et al., 2004). A limitation of AGM as evoked by Therneau (1997) is found on not allowing multiple events to occur at the same time.

The PWPM also known as the Conditional Risk Set Model was proposed by Prentice, Williams and Peterson (Prentice et al., 1981). In PWPM, the idea is the same as that of the AGM

apart from stratifying the data by failure order (Amorim and Cai, 2015). The PWPM is also potential to analyse time to each event from the previous event known as a gap-time model. However, as for the AGM, the PWPM is unable to model several events occurring at the same time per individual.

The MRSM assumes that events are unordered where each event has its own stratum and each data point appears in all strata Wei et al. (1989). This allows an analysis of simultaneous events per subject.

The present study uses two popular nonparametric methods of re-sampling namely, *bootstrap* based on random sampling with replacement (Efron and Tibshirani, 1994), and *jackknife method* which consists of making samples by leaving out one observation a time (Efron and Tibshirani, 1994). The re-sampling in modelling survival analysis allows the assessment of the stability of the survival regression models. One cause of the instability of the model may be the small sample size (Hamada, 1995). The size of the sample in the KUTH data is 2117 and the record is effective in the year 2016. The long term results could be assumed due to the stability observed after re-sampling. Several manuscripts on re-sampling in survival analysis are limited on the re-sampled Cox proportional hazards model and on estimating standard errors of the survival and hazard functions such as in Utzet and Sánchez (1992), Bělašková et al. (2013), Xu et al. (2014) and Xiao et al. (2012) where bootstrap is involved; Wang et al. (2009), Zheng (1987), Zheng and Tu (1988) in which the jackknife is implicated or James (1997), Quan and Tsai (1992), Sauerbrei and Schumacher (1992), Akritas (1986), Efron (1981), Hjort (1985) and Kim (1990) where hazard and survival functions with their respective standard errors are of interest.

This study utilises the bootstrap-based MRSM with 1000 replicates and the jackknife-based MRSM. The results were then compared to that of the MRSM found in Gatabazi et al. (2018)

and in Gatabazi et al. (2019b).

6.2 Marginal Risk Set Model (MRSM)

Let $h(t|\mathbf{x}_i)$ be the hazard function of the survival time T given the p fixed covariates $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$. Assume that $h_0(t)$ is the hazard function when values of all covariates are zeros, then

$$h(t|\mathbf{x}_i) = h_0(t)\exp(\boldsymbol{\beta}'\mathbf{x}_i) \tag{6.1}$$

(Collet, 2003), where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$ is a p-dimensional vector of model parameters. Define an indicator function

$$\delta_{ij}(t) = \begin{cases} 1 & \text{if individual } i \text{ is at risk of the } j^{th} \text{ event} \\ 0 & \text{otherwise.} \end{cases}$$

The marginal risk set model or the Wei Lin and Weisfeld Model (WLWM) assumes that the k^{th} time interval per subject is in the k^{th} stratum, $k = 1, 2, \ldots, n$. In WLWM, the study time is subdivided into $n+1$ intervals each with lower bound 0 and upper bound equal to the time to an event, the first and the last intervals are respectively $[0, \tau_1]$ and $[0, T]$. The hazard function for the j^{th} event for individual i is given by

$$h(t|\mathbf{x}_i) = \delta_{ij}(t)h_{0j}(t)\exp(\boldsymbol{\beta}_j'\mathbf{x}_i). \tag{6.2}$$

6.3 Maximum likelihood and parameter estimation

Assuming that two events cannot occur simultaneously in continuous time, let $]0, \tau_i[$, the interval of time in which the individual i is observed and n_i the number of events of individual i along $]0, \tau_i[$, then the probability density function for the outcome n_i along $]0, \tau_i[$ is given by

$$L(\mathbf{\Phi}) = \prod_{i=1}^{n} L_i(\phi)$$

where

$$L_i(\phi) = \prod_{j=1}^{n_i} h(t|\mathbf{x}_i)\, e^{-\int_0^{\tau_i} \delta_{ij}(v) h(v|\mathbf{x}_i) dv}. \tag{6.3}$$

In (6.3), individual i has n_i events with $n_i \geq 0$ at times $t_{i1} \leq t_{i2} \leq \cdots \leq t_{in_i}$.

The appropriate partial likelihood functions for tied time to event data is well described in Collet (2003) and in Gatabazi and Kabera (2015) and include Breslow's, Efron's and Cox's techniques. The maximum likelihood estimates are obtained by solving a system

$$\begin{cases} \frac{\partial \ln L(\mathbf{\Phi})}{\partial \boldsymbol{\alpha}} = 0 \\ \frac{\partial \ln L(\mathbf{\Phi})}{\partial \boldsymbol{\beta}} = 0 \end{cases} \tag{6.4}$$

where $\boldsymbol{\alpha}$ is known as the baseline parameter vector while $\boldsymbol{\beta}$ is a vector of model parameters. The numerical methods such as the Newton-Raphson method are used for solving system (6.4). The adequacy of parameters is checked by finding the elements $\boldsymbol{\mathcal{I}}_{\boldsymbol{\alpha\alpha}}$, $\boldsymbol{\mathcal{I}}_{\boldsymbol{\alpha\beta}}$, $\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\alpha}}$ and $\boldsymbol{\mathcal{I}}_{\boldsymbol{\beta\beta}}$ of the information matrix $\boldsymbol{\mathcal{I}}$ and assume that as $n \to \infty$, $\hat{\mathbf{\Phi}} - \mathbf{\Phi} \rightsquigarrow N\left(0, \boldsymbol{\mathcal{I}}^{-1}(\hat{\mathbf{\Phi}})\right)$ (Sankaran and Anisha, 2011).

6.4 Re-sampled MRSM

6.4.1 Bootstrap Marginal Risk Set Model (BMRSM)

Consider the p fixed covariates $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in})$ in Equation (6.2) where $x_{ij,i\in[1,p]}$ are independent and identically distributed possibly with distribution F_θ where θ is the statistical parameter of interest. Consider the distribution function F_{R_n} of a random variable $R_n(\mathbf{x}, F_\theta)$. A bootstrap method as described in Efron and Tibshirani (1994), consists of generating

$$\mathbf{x}_i^\star = \mathbf{x}_i^{\star 1}, \mathbf{x}_i^{\star 2} \ldots, \mathbf{x}_i^{\star B},$$

where $\mathbf{x}_i^{\star k}$, $k \in [1, B]$ are random samples of size n drawn with replacement from the sample $\mathbf{x}_i$. The variables of $\mathbf{x}_i^{\star k}$ are independent and identically distributed with distribution $\widehat{F}_{\theta,n}$ given $\mathbf{x}$; $\widehat{F}_{\theta,n}$ is an estimator of F_θ from $\mathbf{x}_i$; B is a number of bootstrap samples also known as replications.

The estimated standard error of the bootstrap statistic of interest $\widehat{\theta}$ is given in Efron and Tibshirani (1994) as

$$\widehat{se}_B = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left[\widehat{\theta}^\star(b) - \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta}^\star(b)\right]^2} \tag{6.5}$$

where $\widehat{\theta}^\star(b)$ is an estimate of the statistic of interest from the b^{th} bootstrap sample, $b = 1, 2, \ldots . B$. The inference of model (6.2) based on bootstrap consists of applying model (6.2) to each of the B bootstrap samples $\mathbf{x}_i^{\star k}$, $\forall k \in [1, B]$ of covariates $\mathbf{x}_i$, $\forall i \in [1, p]$. Bootstrap model parameter estimation in the presence of tied events uses either Breslow, Efron or Cox approach. The bootstrap standard error is obtained by using Equation (6.5).

6.4.2 Jackknife Marginal Risk Set Model (JMRSM)

Consider the p fixed covariates $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in})$ in Equation (6.2). Let θ be a statistic of interest. The jackknife samples consist of leaving out one observation at time, these are n samples $\mathbf{x}_i^\star = (x_{i1}, x_{i2}, \ldots, x_{i\ k-1}, x_{i\ k+1}, \ldots, x_{in})\ \forall j \in [1, n]$ (Efron and Tibshirani, 1994). The jackknife standard error estimate as proposed by Efron and Tibshirani (1994), is

$$\widehat{se}_{jack} = \sqrt{\frac{n-1}{n}\sum_{i=1}^{n}\left[\widehat{\theta^\star}(i) - \frac{1}{n}\sum_{i=1}^{n}\widehat{\theta^\star}(i)\right]^2} \tag{6.6}$$

where $\theta^\star(i)$, $i \in [1, n]$ is a statistic of interest for the i^{th} jackknife sample.

The JMRSM consists of applying model (6.2) to each of the n jackknife samples $\mathbf{x}_i^{\star k}$ of covariates $\mathbf{x}_i$, $i \in [1, p]$. Either Breslow, Efron or Cox approach is used for estimating the jackknife model parameters, with standard error given by Equation (6.6).

6.5 Application

Consider the data described in Table 2.1 regarding the survival times of 2117 newborns from the Kigali University Teaching Hospital (KUTH). The aim of this study is to apply the MRSM on bootstrap and jackknife samples of covariates. The relative risk in each covariate and related standard error were estimated and compared to that obtained by the usual MRSM obtained in Chapter 4. The significance in difference of levels of covariates are also explored in both MRSM and re-sampled MRSM. Using Breslow estimation, Table 6.1 presents unadjusted MRSM, BMRSM, JMRSM and corresponding adjusted models. Unadjusted and adjusted

MRSM, BMRSM and JMRSM by Efron and Cox approaches are also presented in Tables 6.2 and 6.3, respectively.

The results of the unadjusted JMRSM were relatively close to that of the unadjusted MRSM (Table 6.1). The standard errors in JMRSM and MRSM are close for all covariates. The standard errors in BMRSM and MRSM are also close for covariates except all levels of covariate *childbirth* where a standard error in BMRSM is about 4 times that of MRSM and the upper levels of covariates *weight*, *head* and *height* where a standard error in BMRSM is about 20 times that of MRSM. Significance difference in levels of covariates is found at the same covariates for both MRSM, BMRSM and JMRSM except at the upper level of the covariate *abortion* where significance is suggested by the MRSM.

The bootstrap and jackknife re-sampling techniques were applied to the MRSM and then constructed the BMRSM and JMRSM through three different approaches of ties handling. The overall results of MRSM, BMRSM and JMRSM by different approaches of ties handling are similar as expected. The STATA default method (Breslow) is then of interest in the analysis. The JMRSM is adopted for checking stability since the results are closer to that of MRSM compared to the results of BMRSM. The similarity between MRSM and JMRSM suggests that the MRSM may be stable.

Table 6.1: Breslow estimation

MRSM							BMRSM					JMRSM				
Covariate (reference)	Level	HR	SE	z	P>z	95% CI	HR	SE	z	P>z	95% CI	HR	SE	z	P>z	95% CI
Age (Under 20 years old)	20 to 34 years old	0.277	0.100	-3.570	0.000	[0.137; 0.560]	0.277	0.088	-4.060	0.000	[0.149; 0.515]	0.277	0.081	-4.360	0.000	[0.155; 0.493]
	35 years old and above	0.395	0.157	-2.330	0.020	[0.181; 0.863]	0.395	0.132	-2.780	0.005	[0.205; 0.761]	0.395	0.127	-2.890	0.004	[0.210; 0.741]
Residence (Rural)	Urban	0.847	0.139	-1.020	0.309	[0.614; 1.167]	0.847	0.148	-0.950	0.341	[0.601; 1.193]	0.847	0.158	-0.890	0.372	[0.587; 1.220]
Antecedents (Not 1st newborn)	1st newborn	0.806	0.157	-1.100	0.270	[0.550; 1.182]	0.806	0.138	-1.260	0.207	[0.577; 1.126]	0.806	0.134	-1.300	0.193	[0.582; 1.116]
Abortion (Not aborted)	Aborted once	1.405	0.398	1.200	0.231	[0.806; 2.448]	1.405	0.459	1.040	0.298	[0.741; 2.664]	1.405	0.471	1.010	0.311	[0.728; 2.710]
	Aborted more than once	0.479	0.161	-2.190	0.028	[0.248; 0.925]	0.479	0.280	-1.260	0.208	[0.152; 1.507]	0.479	0.360	-0.980	0.328	[0.110; 2.094]
Childbirth (Ventouse)	Natural	0.873	0.491	-0.240	0.808	[0.290; 2.627]	0.873	1.973	-0.060	0.952	[0.010; 73.427]	0.873	0.329	-0.360	0.718	[0.416; 1.829]
	Surgery	1.115	0.613	0.200	0.843	[0.380; 3.274]	1.115	2.517	0.050	0.962	[0.013; 93.040]	1.115	0.372	0.330	0.744	[0.580; 2.143]
Gender (Female)	Male	1.740	0.296	3.260	0.001	[1.247; 2.429]	1.740	0.324	2.980	0.003	[1.209; 2.505]	1.740	0.337	2.860	0.004	[1.191; 2.544]
Number (Singleton)	Multiple	0.409	0.131	-2.790	0.005	[0.218; 0.766]	0.409	0.107	-3.420	0.001	[0.245; 0.682]	0.409	0.100	-3.640	0.000	[0.252; 0.661]
APGAR (Below 4/10)	4/10 to 6/10	0.377	0.112	-3.300	0.001	[0.211; 0.673]	0.377	0.127	-2.900	0.004	[0.195; 0.729]	0.377	0.139	-2.640	0.008	[0.182; 0.778]
	7/10 and above	0.130	0.036	-7.460	0.000	[0.076; 0.222]	0.130	0.033	-8.130	0.000	[0.079; 0.212]	0.130	0.031	-8.470	0.000	[0.081; 0.208]
Weight (Under 2500 g)	2500 g to 4500 g	0.250	0.068	-5.070	0.000	[0.146; 0.427]	0.250	0.064	-5.430	0.000	[0.151; 0.412]	0.250	0.063	-5.540	0.000	[0.153; 0.408]
	Above 4500 g	0.442	0.285	-1.270	0.206	[0.125; 1.565]	0.442	4.002	-0.090	0.928	[0.000; 2.290×10^{7}]	0.442	0.508	-0.710	0.478	[0.046; 4.222
Head (Below 32 cm)	32 cm to 36 cm	0.456	0.128	-2.800	0.005	[0.263; 0.789]	0.456	0.115	-3.100	0.002	[0.277; 0.749]	0.456	0.117	-3.070	0.002	[0.275; 0.753]
	Above 36 cm	0.290	0.219	-1.640	0.102	[0.066; 1.278]	0.290	4.156	-0.090	0.931	[0.000; 4.470×10^{11}]	0.290	0.284	-1.270	0.206	[0.043; 1.971]
Height (Below 36 cm)	46 cm to 54 cm	0.894	0.276	-0.360	0.716	[0.488; 1.637]	0.894	0.241	-0.420	0.677	[0.527; 1.516]	0.894	0.253	-0.400	0.692	[0.513; 1.557]
	Above 54 cm	1.670	1.264	0.680	0.498	[0.379; 7.361]	1.670	22.884	0.040	0.970	[0.000; 7.73×10^{11}]	1.670	1.612	0.530	0.596	[0.251; 11.093]
Adjusted MRSM							Adjusted BMRSM					Adjusted JMRSM				
Covariate (reference)	Level	HR	SE	z	P>z	95% CI	HR	SE	z	P>z	95% CI	HR	SE	z	P>z	95% CI
Age (Under 20 years old)	20 to 34 years old	0.307	0.107	-3.380	0.001	[0.155; 0.609]	0.309	0.089	-4.080	0.000	[0.176; 0.543]	0.309	0.083	-4.370	0.000	[0.182; 0.523]
	35 years old and above	0.472	0.179	-1.980	0.047	[0.225; 0.992]	0.489	0.145	-2.420	0.016	[0.274; 0.874]	0.489	0.137	-2.550	0.011	[0.282; 0.848]
Abortion (Not aborted)	Aborted once	1.482	0.406	1.430	0.152	[0.866; 2.537]	-	-	-	-	-	-	-	-	-	-
	Aborted more than once	0.541	0.175	-1.900	0.057	[0.287; 1.019]	-	-	-	-	-	-	-	-	-	-
Gender (Female)	Male	1.672	0.280	3.070	0.002	[1.204; 2.321]	1.607	0.304	2.510	0.012	[1.109; 2.328]	1.607	0.316	2.410	0.016	[1.093; 2.363]
Number (Singleton)	Multiple	0.401	0.128	-2.860	0.004	[0.214; 0.750]	0.417	0.106	-3.450	0.001	[0.254; 0.686]	0.417	0.103	-3.550	0.000	[0.258; 0.677]
APGAR (Below 4/10)	4/10 to 6/10	0.414	0.119	-3.080	0.002	[0.236; 0.726]	0.412	0.137	-2.660	0.008	[0.215; 0.791]	0.412	0.142	-2.580	0.010	[0.210; 0.809]
	7/10 and above	0.144	0.038	-7.350	0.000	[0.086; 0.242]	0.150	0.034	-8.370	0.000	[0.096; 0.234]	0.150	0.033	-8.580	0.000	[0.098; 0.232]
Weight (Under 2500 g)	2500 g to 4500 g	0.238	0.060	-5.650	0.000	[0.144; 0.391	0.240	0.057	-6.030	0.000	[0.151; 0.381]	0.240	0.057	-6.040	0.000	[0.151; 0.381]
	Above 4500 g	0.447	0.284	-1.270	0.205	[0.129; 1.550	0.478	4.519	-0.080	0.938	[0.000; 5.32×10^{7}]	0.478	0.419	-0.840	0.400	[0.086; 2.669]
Head (Below 32 cm)	32 cm to 36 cm	0.420	0.100	-3.660	0.000	[0.264; 0.669]	0.439	0.103	-3.500	0.000	[0.277; 0.696]	0.439	0.107	-3.390	0.001	[0.273; 0.707]
	Above 36 cm	0.284	0.210	-1.700	0.089	[0.067; 1.211]	0.303	4.200	-0.090	0.931	[0.000; 1.970×10^{11}]	0.303	0.298	-1.210	0.225	[0.044; 2.084]

Table 6.2: Efron estimation

MRSM							BMRSM					JMRSM				
Covariate (reference)	Level	HR	SE	z	P>z	95% CI	HR	SE	z	P>z	95% CI	HR	SE	z	P>z	95% CI
Age (Under 20 years old)	20 to 34 years old	0.230	0.083	-4.080	0.000	[0.114; 0.466]	0.230	0.086	-3.940	0.000	[0.111; 0.478]	0.230	0.083	-4.090	0.000	[0.114; 0.466]
	35 years old and above	0.324	0.129	-2.840	0.005	[0.149; 0.706]	0.324	0.128	-2.850	0.004	[0.149; 0.703]	0.324	0.125	-2.920	0.004	[0.152; 0.691]
Residence (Rural)	Urban	0.831	0.137	-1.120	0.261	[0.602; 1.147]	0.831	0.160	-0.960	0.337	[0.570; 1.212]	0.831	0.174	-0.890	0.376	[0.552; 1.252]
Antecedents (Not 1st newborn)	1st newborn	0.756	0.149	-1.420	0.156	[0.513; 1.113]	0.756	0.149	-1.420	0.155	[0.514; 1.112]	0.756	0.143	-1.480	0.140	[0.521; 1.096]
Abortion (Not aborted)	Aborted once	1.393	0.396	1.170	0.244	[0.798; 2.430]	1.393	0.470	0.980	0.326	[0.719; 2.699]	1.393	0.522	0.880	0.377	[0.668; 2.904]
	Aborted more than once	0.452	0.154	-2.340	0.020	[0.232; 0.880]	0.452	0.322	-1.110	0.265	[0.112; 1.826]	0.452	0.391	-0.920	0.359	[0.083; 2.465]
Childbirth (Ventouse)	Natural	0.736	0.408	-0.550	0.580	[0.249; 2.179]	0.736	1.482	-0.150	0.879	[0.014; 38.109]	0.736	0.336	-0.670	0.502	[0.301; 1.801]
	Surgery	0.921	0.499	-0.150	0.880	[0.319; 2.661]	0.921	1.858	-0.040	0.968	[0.018; 47.963]	0.921	0.388	-0.190	0.846	[0.403; 2.104]
Gender (Female)	Male	1.823	0.312	3.520	0.000	[1.304; 2.549]	1.823	0.361	3.040	0.002	[1.238; 2.687]	1.823	0.400	2.740	0.006	[1.186; 2.804]
Number (Singleton)	Multiple	0.324	0.106	-3.430	0.001	[0.170; 0.617]	0.324	0.100	-3.670	0.000	[0.177; 0.591]	0.324	0.096	-3.810	0.000	[0.181; 0.578]
APGAR (Below 4/10)	4/10 to 6/10	0.214	0.065	-5.090	0.000	[0.118; 0.387]	0.214	0.080	-4.100	0.000	[0.102; 0.447]	0.214	0.093	-3.550	0.000	[0.091; 0.501]
	7/10 and above	0.070	0.020	-9.520	0.000	[0.041; 0.121]	0.070	0.019	-9.660	0.000	[0.041; 0.120	0.070	0.019	-9.830	0.000	[0.041; 0.119]
Weight (Under 2500 g)	2500 g to 4500 g	0.231	0.063	-5.340	0.000	[0.135; 0.395]	0.231	0.064	-5.310	0.000	[0.134; 0.396]	0.231	0.062	-5.440	0.000	[0.136; 0.391]
	Above 4500 g	0.412	0.269	-1.360	0.174	[0.115; 1.479]	0.412	3.892	-0.090	0.925	[0.000; 4.57×10^{7}]	0.412	0.485	-0.750	0.451	[0.041; 4.149]
Head (Below 32 cm)	32 cm to 36 cm	0.422	0.119	-3.060	0.002	[0.243; 0.734]	0.422	0.115	-3.160	0.002	[0.247; 0.720]	0.422	0.118	-3.090	0.002	[0.244; 0.729]
	Above 36 cm	0.246	0.187	-1.840	0.065	[0.055; 1.093]	0.246	3.784	-0.090	0.927	[0.000; 3.030×10^{12}]	0.246	0.251	-1.370	0.169	[0.033; 1.819]
Height (Below 36 cm)	46 cm to 54 cm	0.917	0.285	-0.280	0.781	[0.499; 1.687]	0.917	0.290	-0.270	0.784	[0.494; 1.704]	0.917	0.294	-0.270	0.788	[0.489; 1.721]
	Above 54 cm	1.692	1.283	0.690	0.488	[0.383; 7.476]	1.692	24.567	0.040	0.971	[0.000; 3.890×10^{12}]	1.692	1.700	0.520	0.601	[0.236; 12.140]
Adjusted MRSM							Adjusted BMRSM					Adjusted JMRSM				
Covariate (reference)	Level	HR	SE	z	P>z	95% CI	HR	SE	z	P>z	95% CI	HR	SE	z	P>z	95% CI
Age (Under 20 years old)	20 to 34 years old	0.262	0.092	-3.810	0.000	[0.132; 0.522]	0.265	0.088	-3.980	0.000	[0.138; 0.509]	0.265	0.088	-4.000	0.000	[0.138; 0.508]
	35 years old and above	0.407	0.155	-2.360	0.018	[0.193; 0.859]	0.421	0.151	-2.410	0.016	[0.208; 0.850]	0.421	0.146	-2.490	0.013	[0.213; 0.833]
Abortion (Not aborted)	Aborted once	1.487	0.408	1.440	0.149	[0.868; 2.546]	-	-	-	-	-	-	-	-	-	-
	Aborted more than once	0.520	0.170	-2.000	0.046	[0.274; 0.987]	-	-	-	-	-	-	-	-	-	-
Gender (Female)	Male	1.764	0.297	3.370	0.001	[1.268; 2.453]	1.684	0.336	2.610	0.009	[1.138; 2.490]	1.684	0.367	2.390	0.017	[1.098; 2.582]
Number (Singleton)	Multiple	0.308	0.101	-3.580	0.000	[0.162; 0.586]	0.322	0.097	-3.750	0.000	[0.178; 0.583]	0.322	0.101	-3.630	0.000	[0.175; 0.594]
APGAR (Below 4/10)	4/10 to 6/10	0.249	0.073	-4.730	0.000	[0.140; 0.442]	0.246	0.093	-3.720	0.000	[0.117; 0.515]	0.246	0.100	-3.450	0.001	[0.110; 0.546]
	7/10 and above	0.081	0.022	-9.400	0.000	[0.048; 0.137]	0.085	0.021	-9.940	0.000	[0.052; 0.138]	0.085	0.021	-9.830	0.000	[0.052; 0.138]
Weight (Under 2500 g)	2500 g to 4500 g	0.222	0.057	-5.910	0.000	[0.135; 0.366]	0.225	0.057	-5.910	0.000	[0.137; 0.369]	0.225	0.056	-5.990	0.000	[0.138; 0.367]
	Above 4500 g	0.430	0.276	-1.310	0.189	[0.122; 1.512]	0.487	5.083	-0.070	0.945	[0.000; 3.730×10^{8}]	0.487	0.453	-0.770	0.440	[0.078; 3.023]
Head (Below 32 cm)	32 cm to 36 cm	0.388	0.093	-3.940	0.000	[0.243; 0.622]	0.403	0.105	-3.490	0.000	[0.242; 0.671]	0.403	0.108	-3.380	0.001	[0.238; 0.683]
	Above 36 cm	0.235	0.175	-1.940	0.052	[0.054; 1.014]	0.252	3.678	-0.090	0.925	[0.000; 6.680×10^{11}]	0.252	0.259	-1.340	0.180	[0.034; 1.889]

Table 6.3: Cox estimation

MRSM							BMRSM					JMRSM				
Covariate (reference)	Level	HR	SE	z	P>z	95% CI	HR	SE	z	P>z	95% CI	HR	SE	z	P>z	95% CI
Age (Under 20 years old)	20 to 34 years old	0.193	0.085	-3.730	0.000	[0.081; 0.458]	0.193	0.094	-3.370	0.001	[0.074; 0.502]	0.193	0.088	-3.600	0.000	[0.079; 0.472]
	35 years old and above	0.267	0.128	-2.760	0.006	[0.104; 0.682]	0.267	0.131	-2.700	0.007	[0.102; 0.697]	0.267	0.124	-2.850	0.004	[0.107; 0.662]
Residence (Rural)	Urban	0.766	0.150	-1.360	0.175	[0.521; 1.126]	0.766	0.221	-0.920	0.356	[0.435; 1.349]	0.766	0.221	-0.920	0.356	[0.435; 1.350]
Antecedents (Not 1st newborn)	1st newborn	0.763	0.185	-1.120	0.264	[0.475; 1.226]	0.763	0.219	-0.940	0.345	[0.435; 1.338]	0.763	0.194	-1.060	0.289	[0.463; 1.258]
Abortion (Not aborted)	Aborted once	1.404	0.453	1.050	0.293	[0.746; 2.643]	1.404	0.627	0.760	0.448	[0.585; 3.369]	1.404	0.593	0.800	0.422	[0.613; 3.215]
	Aborted more than once	0.378	0.152	-2.420	0.015	[0.172; 0.830]	0.378	0.336	-1.100	0.274	[0.066; 2.155]	0.378	0.446	-0.830	0.409	[0.038; 3.814]
Childbirth (Ventouse)	Natural	0.732	0.481	-0.470	0.635	[0.202; 2.653]	0.732	0.369	-0.620	0.537	[0.273; 1.968]	0.732	0.365	-0.630	0.532	[0.276; 1.945]
	Surgery	1.016	0.654	0.030	0.980	[0.288; 3.590]	1.016	0.480	0.030	0.973	[0.403; 2.565]	1.016	0.455	0.040	0.971	[0.423; 2.443]
Gender (Female)	Male	1.991	0.405	3.390	0.001	[1.336;2.966]	1.991	0.534	2.570	0.010	[1.177; 3.368]	1.991	0.601	2.280	0.023	[1.101; 3.599]
Number (Singleton)	Multiple	0.218	0.111	-3.000	0.003	[0.080; 0.589]	0.218	0.155	-2.140	0.033	[0.054; 0.882]	0.218	0.131	-2.530	0.011	[0.067; 0.709]
APGAR (Below 4/10)	4/10 to 6/10	0.080	0.042	-4.810	0.000	[0.029; 0.224]	0.080	0.056	-3.580	0.000	[0.020; 0.319]	0.080	0.052	-3.870	0.000	[0.022; 0.287]
	7/10 and above	0.021	0.011	-7.840	0.000	[0.008; 0.056]	0.021	0.014	-5.970	0.000	[0.006; 0.076]	0.021	0.011	-7.230	0.000	[0.008; 0.061]
Weight (Under 2500 g)	2500 g to 4500 g	0.236	0.070	-4.850	0.000	[0.131; 0.423]	0.236	0.077	-4.420	0.000	[0.124; 0.448]	0.236	0.068	-5.000	0.000	[0.134; 0.415]
	Above 4500 g	0.378	0.257	-1.430	0.153	[0.100; 1.436]	0.378	4.696	-0.080	0.938	[0.000; 1.410×10^{10}]	0.378	0.473	-0.780	0.437	[0.033; 4.386]
Head (Below 32 cm)	32 cm to 36 cm	0.391	0.119	-3.100	0.002	[0.216; 0.708]	0.391	0.101	-3.640	0.000	[0.236; 0.649]	0.391	0.115	-3.180	0.001	[0.219; 0.698]
	Above 36 cm	0.212	0.171	-1.920	0.055	[0.043; 1.033]	0.212	3.376	-0.100	0.922	[0.000; 7.780×10^{12}]	0.212	0.238	-1.380	0.167	[0.023; 1.913]
Height (Below 36 cm)	46 cm to 54 cm	0.828	0.283	-0.550	0.582	[0.423; 1.620]	0.828	0.254	-0.610	0.539	[0.454; 1.512]	0.828	0.284	-0.550	0.582	[0.423; 1.622]
	Above 54 cm	1.706	1.351	0.670	0.500	[0.361; 8.060]	1.706	28.569	0.030	0.975	[0.000; 3.090×10^{14}]	1.706	1.747	0.520	0.602	[0.229; 12.707]
Adjusted MRSM							Adjusted BMRSM					Adjusted JMRSM				
Covariate (reference)	Level	HR	SE	z	P>z	95% CI	HR	SE	z	P>z	95% CI	HR	SE	z	P>z	95% CI
Age (Under 20 years old)	20 to 34 years old	0.218	0.094	-3.520	0.000	[0.094; 0.509]	0.219	0.078	-4.280	0.000	[0.109; 0.439]	0.219	0.087	-3.830	0.000	[0.101; 0.476]
	35 years old and above	0.341	0.157	-2.340	0.019	[0.138; 0.841]	0.352	0.133	-2.760	0.006	[0.167; 0.738]	0.352	0.141	-2.610	0.009	[0.160; 0.771]
Abortion (Not aborted)	Aborted once	1.479	0.459	1.260	0.208	[0.804; 2.719]	-	-	-	-	-	-	-	-	-	-
	Aborted more than once	0.424	0.161	-2.260	0.024	[0.201; 0.892]	-	-	-	-	-	-	-	-	-	-
Gender (Female)	Male	1.886	0.374	3.200	0.001	[1.278; 2.783]	1.833	0.544	2.040	0.041	[1.025; 3.278]	1.833	0.528	2.100	0.036	[1.042; 3.225]
Number (Singleton)	Multiple	0.214	0.108	-3.050	0.002	[0.079; 0.576]	0.227	0.136	-2.480	0.013	[0.070; 0.732]	0.227	0.135	-2.490	0.013	[0.070; 0.730]
APGAR (Below 4/10)	4/10 to 6/10	0.098	0.050	-4.550	0.000	[0.036; 0.267]	0.091	0.053	-4.100	0.000	[0.029; 0.286]	0.091	0.062	-3.530	0.000	[0.024; 0.345]
	7/10 and above	0.026	0.012	-7.680	0.000	[0.010; 0.066]	0.026	0.013	-7.660	0.000	[0.010; 0.067]	0.026	0.013	-7.420	0.000	[0.010; 0.069]
Weight (Under 2500 g)	2500 g to 4500 g	0.213	0.057	-5.730	0.000	[0.125; 0.361]	0.215	0.060	-5.540	0.000	[0.125; 0.371]	0.215	0.057	-5.810	0.000	[0.128; 0.362]
	Above 4500 g	0.364	0.245	-1.500	0.134	[0.097; 1.364]	0.398	4.183	-0.090	0.930	[0.000; 3.590×10^{8}]	0.398	0.385	-0.950	0.340	[0.060; 2.650]
Head (Below 32 cm)	32 cm to 36 cm	0.349	0.090	-4.080	0.000	[0.211; 0.579]	0.374	0.102	-3.590	0.000	[0.219; 0.640]	0.374	0.105	-3.510	0.000	[0.216; 0.648]
	Above 36 cm	0.199	0.160	-2.020	0.044	[0.042; 0.957]	0.222	3.684	-0.090	0.928	[0.000; 7.970×10^{13}]	0.222	0.253	-1.320	0.186	[0.024; 2.067]

The overall analysis confirms the significant difference in all levels of covariates *age*, *gender*, *number* and *APGAR* and intermediate levels of covariates *weight* and *head*. The re-sampled adjusted models by Breslow technique of handling tied events suggest that the risk of death attracting a chronic disease of babies whose mothers' age range from 20 to 34 years old is 0.265 times that of babies whose mothers are under 20 years old. The risk of babies whose mothers are 35 years old and above is 0.421 times that of babies whose mothers are under 20 years old. The risk for male babies is 1.684 times that of female babies. The risk of multiple babies is 0.322 times that of singleton babies. The risk of babies whose APGAR range from 4/10 to 6/10 is 0.246 times that of babies whose APGAR is below 4/10. The risk of babies whose APGAR range from 7/10 to 10/10 is 0.085 times that of babies whose APGAR is below 4/10. The risk of babies whose weight range from 2500g to 4500g is 0.225 times that of babies whose weight is below 2500g. The risk of babies whose weight is above 4500g is 0.372 times that of babies whose weight is below 2500g. The risk of babies whose circumference of the head range from 32cm to 36cm is 0.403 times that of babies whose circumference of the head is below 32cm.

The results of BMRSM are close to that of JMRSM and MRSM for all significant covariates but the re-sampled models show relatively higher standard errors for some non-significant covariates. The discrepancy between standard errors after re-sampling for covariates such as *childbirth*,*weight*, *head* and *height* suggests instability of the MRSM at these specific covariates and this emphasizes their non-significance in the MRSM.

6.6 Conclusion

Different methods of re-sampling in MRSM were described in this chapter namely, the bootstrap and jackknife. The results of Bootstrap Marginal Risk Set Model (BMRSM) and Jackknife Marginal Risk Set Model (JMRSM) are compared to that of the MRSM for three different methods of ties handling namely, Breslow, Efron and Cox approximations.

The JMRSM and MRSM have displayed relatively close results with significance on covariates *age*, *abortion*, *gender*, number, *APGAR*, *weight* and *head*. The models revealed that the risk of death or attracting chronic disease of an infant is higher in male babies as compared to female babies. The risk is higher for babies whose mothers are under 20 years old mothers as compared to older mothers. 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 4/10 and those whose APGAR range between 4/10 and 6/10. The risk is lower for underweight babies as compared to babies with normal weight and overweight. The survival outcome for babies with normal circumference of the head was found to be better than those with a relatively small head.

Both BMRSM, JMRSM and MRSM displayed close results for significant covariates. However, the BMRSM displayed relatively higher standard error for some non-significant covariates and this emphasizes their insignificance in MRSM. Babies from under 20 years old mothers were found at relatively higher risk and hence pregnancy of mothers belonging in such range of age should be avoided. Also abnormality in infant's weight and head lead to relatively higher risk to infant mortality, clinically recommended ways of keeping pregnancy against any cause of infant abnormality should be reinforced.

CHAPTER 7

CONCLUSIONS

The research work provided in this dissertation described the main points of the non-parametric classical survival analysis and their application to the infant mortality data.

The basic functions of survival analysis reviewed were the survival function, the hazard function and the cumulative hazard function by using Kaplan-Meier estimation. Non-parametric survival regression models were also reviewed namely, the Cox Proportional Hazards Model (CPHM), the Aalen Additive Hazards Model (AAHM) and the Cox-Aalen Hazards Model (CAHM). Multiple events methods reviewed include the Andersen-Gill Model (AGM), the Marginal Risk Set Model (MRSM) also known as the Wei, Lin and Weissfeld Model (WLWM) and the Prentice, Williams and Peterson Model (PWPM); the MRSM was selected due to the structure of the dataset of interest. Re-sampled CPHM and re-sampled MRSM were also conducted and compared respectively to CPHM and MRSM.

The theoretical results described were applied to data on infant mortality collected at the Kigali University Teaching Hospital (KUTH) during the period from the first January 2016 to the 31^{st} December 2016. The total number of newborns with complete information were 2117 of whom 69 were stillborn and 82 died during the study time.

The covariates reported in the data were the age of mothers, the gender of a newborn, residential area of parents, indicator on whether a newborn is the first or not, indicator on whether a mother aborted previously, number of newborns at a time, APGAR score of a newborn, weight

of a newborn, circumference of the head of a newborn and height of a newborn. The death of a newborn was considered as an event in single event analysis while multiple events analysis included the death of a newborn and the occurrence of at least one of the conditions that may also cause long term death to infants such as severe oliguria, severe prematurity, very low birth weight, macrosomia, severe respiratory distress, gastroparesis, hemolytic, trisomy, asphyxia and laparoschisis.

Exploratory analysis of the data was done using Kaplan-Meier estimation. Proportional Hazards Assumption (PHA) was checked by assessing Kaplan-Meier estimates of survival functions per groups of covariates. It was found that PHA is violated for covariates *antecedents* and *childbirth*. The data analysis was done using the CPHM, AAHM and CAHM as single event regression models and MRSM as multiple events regression model.

The CPHM for all covariates showed significance on covariates *age*, *gender*, *number*, *APGAR*, *weight* and *head* while the unadjusted re-sampled CPHM models revealed significance on the *age*, *abortion*, *gender*, *number*, *APGAR*, *weight* and *head*. The four tests of AAHM showed significance for the present covariates except *residence*. AAHM indicated time dependent covariates and revealed that fixed covariates may be *age*, *APGAR* and *childbirth*. The CAHM combined multiplicative and additive parts with fixed covariates taken at the multiplicative part, while time dependent covariates are included in the additive part. The CPHM assumes the presence of fixed covariates and would be preferred if the time dependent covariates are dropped out. The CAHM takes the covariates *antecedents* and *childbirth* that violate the PHA at its additive part and all other covariates at its multiplicative part. It was found that CAHM would be appropriate if the model is based on covariates *age*, *number*, *APGAR* and *weight* where significance is read. The AAHM distinguishes time dependent and fixed covariates and shown advantage at the point of significance of covariates.

The results obtained from the multiple events analysis revealed that the MRSM fit the data well. The covariates *age*, *abortion*, *gender*, *number*, *APGAR*, *weight* and *head* were found to have significant effect. The re-sampled MRSM confirmed significant difference of the levels of covariates *age*, *gender*, *number*, *APGAR*,*weight* and *head*.

The overall study suggests that infant from under 20 years mothers was at relatively higher risk of death. This complies with studies such as that of Basinga et al. (2012), Olausson et al. (1999) or Lampinen et al. (2009). Pregnancy for under 20 years old mothers should then be avoided. Also a newborn's head and weight abnormality led to a relatively higher risk to infant mortality in line of the recommendations of the clinical medicine as compiled for example by Janssen et al. (2007). Clinically recommended nutrition during pregnancy would decrease abnormality of the newborn, leading to the infant mortality decrease.

The work described in this dissertation has a number of limitations. For example, it would be good to investigate subject effect or groups of subjects effects. This could be done by introducing a frailty variable in the CPHM or in the AAHM to account for the heterogeneity of the subjects or groups of subjects (Hosmer et al., 2008, p. 296, Aalen et al., 2008, p. 231).

The analysis was limited to only 11 variables. Unavailable variables concerning mothers 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 *sports activity level*. These variables were not recorded in the registry at KUTH. This study for Rwanda would be extended to further countries in Sub-Saharan Africa, Africa or worldwide.

References

Aalen, O. O. (1975). *Statistical inference for a family of counting process*. Ph.D. Thesis, University of California, Berkeley.

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*, 2^{nd} edition. The International Bank for Reconstruction and Development/ The World Bank.
URL http://www.ncbi.nlm.nih.gov/books/NBK2292/

Akritas, M. G. (1986). Bootstrapping the Kaplan-Meier Estimator. *Journal of the American Statistical Association* **81** *(396), 1032-1038*.

Amorim, L. D., and Cai, J. (2015). Modeling recurrent events: A tutorial for analysis in epidemiology. *International journal of epidemiology,* **44** *(1), 324-333*.

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical models based on counting processes*. Spring-Verlag, New York.

Andersen, P. K., and Gill, R. D. (1992). Cox's regression model for counting processes: A large sample study. *Ann Statist,* **10**, *1100-1120*.

Autar, K. (2009). Newton-Raphson method of solving nonlinear equations.
URL http://numericalmethods.eng.usf.edu

Basinga, P., Moore, A. M., Singh, S., Remez, L., Birungi, F., and Nyirazinyoye, L. (2012). *Unintended Pregnancy and Induced Abortion in Rwanda: Causes and Consequences*. Guttmacher Institute, New York.

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*.

Burr, D. (1994). A comparison of certain bootstrap confidence intervals in the Cox model. *Journal of the American Statistical Association* **89**, *1290-1302*.

Bělašková, S., Fišerová, E., and Krupičková, S. (2013). Study of bootstrap estimates in Cox regression model with delayed entry. *Mathematica* **52** *(2), 21-30*.

Cai, J., and Schaubel, D. (2004). Analysis of recurrent event data. *Advances in Survival Analysis,* **23**, *603-623*.

Castañeda, J., and Gerritse, B. (2010). Appraisal of several methods to model time to multiple

events per subject: Modeling time to hospitalisations and death. *Revista Colombiana de Estaística,* **33** *(1), 43-61*.

Clayton, D. G., and Cuzick, J. (1985). Multivariate generalisations of the proportional hazards model (with discussion). *Journal of the Royal Statistical Society, Series B,* **148**, *82-117*.

Cleves, M. (2000). Analysis of multiple failure-time survival data. *Stata Technical Bulletin,* **9**.

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

Cook, R. J., and Lawless, J. F. (2002). Analysis of repeated events. *Statistical Methods in Medical Research,* **11**, *141-166*.

Cook, R. J., and Lawless, J. F. (2007). *The statistical analysis of recurrent events*. Springer, New York.

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

Doob, J. L. (1953). *Stochastic processes*. Wiley, New York.

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

Efron, B. (1981). Censored data and the bootstrap. *Journal of the American Statistical Association* **76**, *312-319*.

Efron, B. (1992). Jackknife-After-Bootstrap standard errors and influence functions. *Wiley for the Royal Statistical Society* **54** *(1), 83-127*.

Efron, B., and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Chapman & Hall/CRC.

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*.

Fisher, L. D. (1991). A review of methods for handling multiple endpoints in clinical trials. *Proceedings of the American Statistical Association, 43-46*.

Flemming, T. R., and Harrington, D. P. (1991). *Counting process and survival analysis*. John Wiley & Sons, Inc., Haboken.

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

Gatabazi, P., and Kabera, G. (2015). Survival analysis and its stochastic process approach with application to diabetes data.
URL https://www.researchgate.net/profile/Paul_Gatabazi

Gatabazi, P., Melesse, S. F., and Ramroop, S. (2018). Multiple events model for the infant mortality at Kigali University Teaching Hospital. *The Open Public Health Journal* **11**, *464-473*.

Gatabazi, P., Melesse, S. F., and Ramroop, S. (2019a). Infant mortality at the Kigali University Teaching Hospital: Application of Aalen additive hazards model and comparison with other classical survival models. *African Population Studies* **33** *(2), 4834-4851*.

Gatabazi, P., Melesse, S. F., and Ramroop, S. (2019b). Re-sampled Cox proportional hazard model for the infant mortality at the Kigali university teaching hospital. *The Open Public Health Journal* **12**, *136-144*.

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

Greenwood, M. J. (1926). The natural duration of cancer. *Reports of Public Helph and related subjects* **33** *HMSO, London*.

Hamada, C. (1995). Bootstrap Cox regression using SAS procedures. *SAS Institute Japan Ltd.* **211**, *1258-1262*.

Hjort, N. (1985). Bootstrapping Cox's regression model. *Technical Report, Stanford University, Dept. of Statistics* **241**.

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*, 2^{nd} edition. John Wiley & Sons Inc., Haboken.

James, L. F. (1997). A study of a class of weighted bootstraps for censored data. *Ann Statist* **25**, *1595-1621*.

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*.

Johnson, C. J., Boyce, M. S., Schwartz, C. C., and Haroldson, M. A. (2004). Modeling survival: Application of the Andersen-Gill model to Yellowstone Grizzly Bears. *The Journal of Wildlife Management,* **68**, *966-978*.

Kaplan, E., and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assn,* **53** *(282), 457-481*.

Kelly, P. J., and Lim, L. L. (2000). Survival analysis for recurrent event data: An allpication to childhood infectious diseases. *Statistics in Medicine,* **19**, *13-33*.

Kim, J. (1990). *Conditional bootstrap methods for censored data*. PhD dissertation, Florida State University, Dept, of Statistics.

Kivaranovich, D., and Leeb, H. (2019). On the length of post-model-selection confidence intervals conditional on polyhedral constraints. *ArXiv: 1803.01665* **2**.

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

Lampinen, R., Julkunen, K. V., and Kankkunen, P. (2009). A review of pregnancy in women over 35 years of age. *The Open Nursing Journal* **3**, *33-38*.

Lawless, J. F., and Nadeau, C. (2012). Some simple robust methods for the analysis of recurrent events. *Technometrics,* **37**, *158-168*.

Lin, D. Y., and Ying, Z. (1994). Semiparametric analysis o the additive risk model. *Biometrika,* **81** *(1), 61-71*.

Lin, J. S., and Wei, L. J. (1992). Linear regression analysis for multivariate failure time observations. *Journal of the American Statistical Association,* **87**, *1091-1097*.

Lindor, K. D., Dickson, E. R., Baldus, W. P., Jorgensen, R. A., Ludwig, J., Murtaugh, P. A., Harrison, J. M., Wiesner, R. H., Anderson, M. L., and Lange, S. M. (1994). Ursodeoxycholic acid in the treatment of primary biliary cirrhosis. *Gastroenterology,* **106** *(5), 1284-1290*.

Louzada, F. (2007). Intensity models for parametric analysis of recurrent events data. *Brazilian Journal of Probability and Statistics,* **22**, *No 1, 23-33*.

Macdonald, A. S. (1996). Competing risks, nonparametric and regression models. *B A J,* **2**, *429-448*.

Monden, C. W. S., and Smits, J. (2017). Mortality among twins and singletons in Sub-Saharan Africa between 1995 and 2014: a pooled analysis of data from 90 demographic and health surveys in 30 countries. *Proc Natl Acad Sci U S A* **5** *(7), E673-E679*.

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*, 2^{nd} edition. Chapman Hall/ CRC, Boca Raton.

O'Brien, P. C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics,* **40**, *1079-1087*.

Olausson, P. O., Cnattinius, S., and Haglund, B. (1999). Teenage pregnancies and risk of late fetal death and infant mortality. *British Journal of Obstetricsand Gynaecology* **106**, *116-121*.

Parzen, M., and Lipsitz, S. R. (1999). A global goodness-of-fit statistic for Cox regression models. *Biometrics* **55**, *580-584*.

Peduzzi, P., Concato, J., Kemper, E., Holford, T., and Feinstein, A. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology* **49**, *1373-1379*.

Pongou, R. (2013). Why is infant mortality higher in boys than in girls? A new hypothesis based on preconception environment and evidence from a large sample of twins. *Demography* **50**, *421-444*.

Pongou, R., Shapiro, D., and Tenikue, M. (2019). Mortality convergence of twins and singletons in Sub-Saharan Africa. *Demographic research* **41** *(36), 1047-1058*.

Prentice, R. L., Williams, B. J., and Peterson, A. V. (1981). On the regression analysis of multivariate incomplete failure time data. *Biometrika,* **68**, *373-379*.

Quan, H., and Tsai, W. Y. (1992). Jackknife for the proportional hazards model. *J Statist Comput Simul* **43**, *163-176*.

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

Sagara, I., Giolgi, R., Doumbo, O., Piarroux, R., and Gaudart, J. (2014). Modeling recurrent events: Comparison of statistical models with continuous and discontinuous risk intervals on recurrent malaria episodes data. *Malaria Journal* **13**, *293*.
URL https://hal.archives-ouvertes.fr/hal-01208210/document

Sankaran, P. G., and Anisha, P. (2011). Shared frailty model for recurrent event data with multiple causes. *Journal of Applied Statistics,* **38**,*No 12, 2859-2868*.

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/PMC3250938/

Sauerbrei, W., and Schumacher, M. (1992). A bootstrap re-sampling procedure for model building: Application to the Cox regression model. *Statistics in Medicine* **11** *(16), 2093-2109*.

Schaubel, D., Johansen, H., Dutta, M., Desmeules, M., Becker, A., and Mao, Y. (1996). Neonatal characteristics as risk factors for preschool asthma. *J. Asthma,* **33**, *255-264*.

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*.

Sun, L., Park, D., and Sun, J. (2006). The additive hazards model for recurrent gap times. *Statistica Sinica,* **16**, *919-932*.

Therneau, T. M. (1997). Extending the Cox's model. *Proceedings of the First Seattle Symposium in Biostatistics. New York: Springer*.

Utzet, F., and Sánchez, A. (1992). Some applications of the bootstrap to survival analysis. *Anuario de Psicologia* **55**, *155-167*.

Wang, Z., Wu, Y., and Zhao, L. C. (2009). Approximation by randomly weighting method in censored regression model. *Sci China Ser A* **52**, *567-576*.

Wasserman, C. (2013). Infant mortality. *Health of Washington State, Washington State Department of Health-2013*.

URL http://www.doh.wa.gov/portals/1/documents/5500/mch-im2013.pdf

Wei, L. J., and Glidden, D. V. (1997). An overview of statistical methods for multiple failure time data in clinical trials. *Statistics in medecine,* **16**, *833-839*.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of mulyivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association,* **84** *(408), 1065-1073*.

Wise, P. (2003). The anatomy of a disparity in infant mortality. *Annu Rev Public Health* **24**, *341-362*.

Xiao, L., Yao-Hua, W., and Dong-Sheng, T. (2012). Jackknifed random weighting for Cox proportional hazards model. *Sci China Math* **55***(4), 775-786*.

Xu, G., Sen, B., and Ying, Z. (2014). Bootstrapping a change-point Cox model for survival data. *Electron J Stat.* **8** *(1), 1345-1379*.

Zarulli, V., Jones, J. A. B., Oksuzyanc, A., Lindahl-Jacobsen, R., Christensen, K., and Vaupel, J. W. (2018). Women live longer than men even during severe famines and epidemics. *Proc Natl Acad Sci U S A* **115** *(4), E832-E840*.

Zheng, Z. G. (1987). Random weighting method. *Acta Math Appl Sin* **10**, *247-253*.

Zheng, Z. G., and Tu, D. (1988). Random weighting method in regression models. *Sci Sinica Series A* **31**, *1442-1459*.

APPENDICES

Appendix A

Tables and results

Table 1: First 50 entries of the dataset on infant mortality at KUTH from 01-January-2016 to 31-December-2016

no	id	age	residence	antecedents	abortion	childbirth	gender	number	apgar	weight	head	height	tstart	tstop	event	n_events
1	1	1	0	0	0	2	0	0	2	1	1	1	0	39	0	1
2	2	1	0	0	1	1	1	0	2	1	1	1	0	212	0	1
3	3	1	1	1	0	1	1	0	2	1	1	1	0	196	0	1
4	4	1	1	0	0	2	0	0	2	1	1	1	0	128	0	1
5	5	1	0	0	0	2	1	0	2	1	1	1	0	335	0	1
6	6	1	1	0	0	1	0	0	2	1	1	1	0	262	0	1
7	7	2	0	0	0	1	0	0	2	1	1	1	0	214	0	1
8	8	1	0	0	0	1	1	0	2	1	1	1	0	228	0	1
9	9	2	1	0	0	1	0	0	2	1	1	1	0	355	0	1
10	10	1	1	1	0	2	1	0	2	0	1	1	0	25	0	1
11	11	1	1	0	0	2	0	0	2	0	0	0	0	256	0	1
12	12	2	0	1	0	2	0	1	2	0	1	1	0	179	0	1
13	13	2	0	1	0	2	0	1	2	0	1	1	0	179	0	1
14	14	2	0	1	0	2	1	1	2	0	0	0	0	179	0	1
15	15	2	0	0	0	2	1	0	2	1	1	1	0	348	0	1
16	16	1	1	0	1	2	0	0	2	1	1	1	0	305	0	1
17	17	1	1	0	0	1	0	0	2	1	1	1	0	45	0	1
18	18	1	0	0	0	2	1	0	2	1	1	1	0	129	0	1
19	19	1	1	0	0	2	1	0	0	2	1	1	0	0	1	1
20	19	1	1	0	0	2	1	0	0	2	1	1	0	0	1	2
21	20	1	1	1	0	2	0	0	2	1	1	1	0	137	0	1
22	21	0	1	1	0	1	0	0	2	1	1	1	0	293	0	1
23	22	2	1	0	2	1	0	0	2	1	1	1	0	70	0	1
24	25	1	0	1	0	1	1	0	1	1	1	1	0	1	1	1
25	25	1	0	1	0	1	1	0	1	1	1	1	0	1	1	2
26	23	1	1	0	0	1	0	0	2	1	1	1	0	218	0	1
27	24	1	1	0	0	1	1	0	2	1	1	1	0	260	0	1
28	26	2	0	0	0	2	0	0	2	1	1	1	0	24	0	1
29	27	1	1	0	0	2	0	0	2	1	1	1	0	16	0	1
30	28	1	0	1	0	2	1	0	2	0	0	0	0	318	0	1
31	29	2	0	1	0	2	0	0	2	0	0	0	0	1	1	1
32	29	2	0	1	0	2	0	0	2	0	0	0	0	6	1	2
33	30	1	1	0	0	1	0	0	2	1	1	1	0	249	0	1
34	31	1	1	1	0	2	0	0	2	1	1	1	0	311	0	1
35	32	1	1	0	0	1	1	0	2	1	1	1	0	357	0	1
36	33	1	1	0	0	1	1	0	2	0	1	1	0	232	0	1
37	34	1	1	1	0	1	1	0	2	1	1	1	0	356	0	1
38	35	1	1	0	0	1	0	0	2	1	1	1	0	140	0	1
39	36	1	1	1	0	2	0	0	2	1	1	2	0	272	0	1
40	37	2	0	1	0	2	0	0	2	1	1	1	0	203	0	1
41	38	1	1	0	0	2	1	0	2	1	2	0	0	235	0	1
42	39	1	1	0	0	2	0	0	2	1	1	1	0	305	0	1
43	40	1	1	1	0	1	0	0	2	1	1	1	0	263	0	1
44	41	2	0	0	2	2	0	0	2	0	0	0	0	192	0	1
45	42	1	1	0	0	1	1	0	2	1	1	1	0	248	0	1
46	43	1	1	0	0	1	1	0	2	0	0	0	0	1	1	1
47	43	1	1	0	0	1	1	0	2	0	0	0	0	4	1	2
48	44	1	1	0	0	1	1	0	2	1	1	1	0	254	0	1
49	45	2	1	0	0	0	0	0	2	1	1	1	0	333	0	1
50	46	1	1	0	0	2	1	0	2	1	1	1	0	39	0	1

Table 2: Adjusted WLWM for the infant mortality at KUTH from 01-January-2016 to 31-December-2016 with Efron method of ties handling.

Covariate (reference)	Level	Hazard Ratio	Std. Err.	z	P>z	95% Conf. Int.
Age (Under 20 years old)	20 to 34 years old	0.262	0.092	-3.810	$p < 0.001$	[0.132; 0.522]
	35 years old and above	0.407	0.155	-2.360	0.018	[0.193; 0.859]
Abortion (Not aborted)	Aborted once	1.487	0.408	1.440	0.149	[0.868; 2.546]
	Aborted more than once	0.520	0.170	-2.000	0.046	[0.274; 0.987]
Gender (Female)	Male	1.764	0.297	3.370	0.001	[1.268; 2.453]
Number (Singleton)	Multiple	0.308	0.101	-3.580	$p < 0.001$	[0.162; 0.586]
APGAR (Below 4/10)	4/10 to 6/10	0.249	0.073	-4.730	$p < 0.001$	[0.140; 0.442]
	7/10 and above	0.081	0.022	-9.400	$p < 0.001$	[0.048; 0.137]
Weight (Under 2500 g)	2500 g to 4500 g	0.222	0.057	-5.910	$p < 0.001$	[0.135; 0.366]
	Above 4500 g	0.430	0.276	-1.310	0.189	[0.122; 1.512]
Head (Below 32 cm)	32 cm to 36 cm	0.388	0.093	-3.940	$p < 0.001$	[0.243; 0.622]
	Above 36 cm	0.235	0.175	-1.940	0.052	[0.054; 1.014]

Table 3: Adjusted WLWM for the infant mortality at KUTH from 01-January-2016 to 31-December-2016 with Cox method of ties handling.

Covariate (reference)	Level	Hazard Ratio	Std. Err.	z	P>z	95% Conf. Int.
Age (Under 20 years old)	20 to 34 years old	0.218	0.094	-3.520	$p < 0.001$	[0.094; 0.509]
	35 years old and above	0.341	0.157	-2.340	0.019	[0.138; 0.841]
Abortion (Not aborted)	Aborted once	1.479	0.459	1.260	0.208	[0.804; 2.719]
	Aborted more than once	0.424	0.161	-2.260	0.024	[0.201; 0.892]
Gender (Female)	Male	1.886	0.374	3.200	0.001	[1.278; 2.783]
Number (Singleton)	Multiple	0.214	0.108	-3.050	0.002	[0.079; 0.576]
APGAR (Below 4/10)	4/10 to 6/10	0.098	0.050	-4.550	$p < 0.001$	[0.036; 0.267]
	7/10 and above	0.026	0.012	-7.680	$p < 0.001$	[0.010; 0.066]
Weight (Under 2500 g)	2500 g to 4500 g	0.213	0.057	-5.730	$p < 0.001$	[0.125; 0.361]
	Above 4500 g	0.364	0.245	-1.500	0.134	[0.097; 1.364]
Head (Below 32 cm)	32 cm to 36 cm	0.349	0.090	-4.080	$p < 0.001$	[0.211; 0.579]
	Above 36 cm	0.199	0.160	-2.020	0.044	[0.042; 0.957]

Appendix B

Published papers

Paper 1

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 $\boldsymbol{x}_i = (x_{i1}, x_{i2}, ..., x_{ip})'$ and $h_0(t)$ a hazard function when values of all covariates are zeros. The CPHM is given in [7] as

$$h(t|\boldsymbol{x}_i) = h_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{x}_i) \quad (1)$$

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

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

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 i^{th} of n individuals with vector of covariates $\boldsymbol{x}_i(t) = (x_{i1}, x_{i2}, ..., x_{ip})'$. That is given by

$$h[t|\boldsymbol{x}_i(t)] = \beta_0(t) + \beta_1(t)x_{i1}(t) + \beta_2(t)x_{i2}(t) + ... + \beta_p(t)x_{ip}(t) \quad (3)$$

where $\boldsymbol{\beta}(t) = (\beta_0(t), \beta_1(t), ..., \beta_p(t))'$ 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

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

$k = 0, 1, 2, \ldots, p$. 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

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

where $x_{i0} = 1$.

Model (5) has the form of a multiple linear regression model for the i^{th} 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 $k = 0, 1, 2, \ldots, p$ and $i = 0, 1, 2, \ldots, n$. Model (5) can be written in matrix form as

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

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 $Y_i(t)$, $Y_i(t)x_{i1}(t), \ldots, Y_i(t)x_{ip}(t)$,

$d\mathbf{B}(t) = (dB_0(t), dB_1(t), \ldots, dB_p(t))'$ 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

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

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

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

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

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

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

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

and

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

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-α) % confidence interval for the k^{th} 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 k^{th} 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: $\beta_k(t) = 0, \forall t \in [0, t_0]$
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

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

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

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

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

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

Weights 4: $\mathbf{K}_j = \mathrm{diag}[\hat{S}_{KM}(t_{j-1})/\mathrm{se}(\hat{\beta}_{kk}(tj))]$ where $\hat{\beta}_{kk}(tj)$ is the k^{th} 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

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

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

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

Where $\hat{u}_k$ is the k^{th} element of $\hat{\mathbf{u}}$ given in (14) and $\mathrm{se}(\hat{u}_k)$ is the square root of the k^{th} 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−α) % 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; $\boldsymbol{\beta}(t)$ is a $(p\times 1)$ of time-varying regression coefficient and $\boldsymbol{\alpha}$ is a $(q\times 1)$ vector of relative risk regression coefficients. Then the hazard function is given by

$$h(t|x) = Y(t)[X'(t)\boldsymbol{\beta}(t)]\exp(Z'(t)\boldsymbol{\alpha}) \quad (17)$$

Estimation is based on cumulative parameter functions $B(t) = \int_0^t \beta(v)dv$ and model parameters $\boldsymbol{\alpha}$. 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]

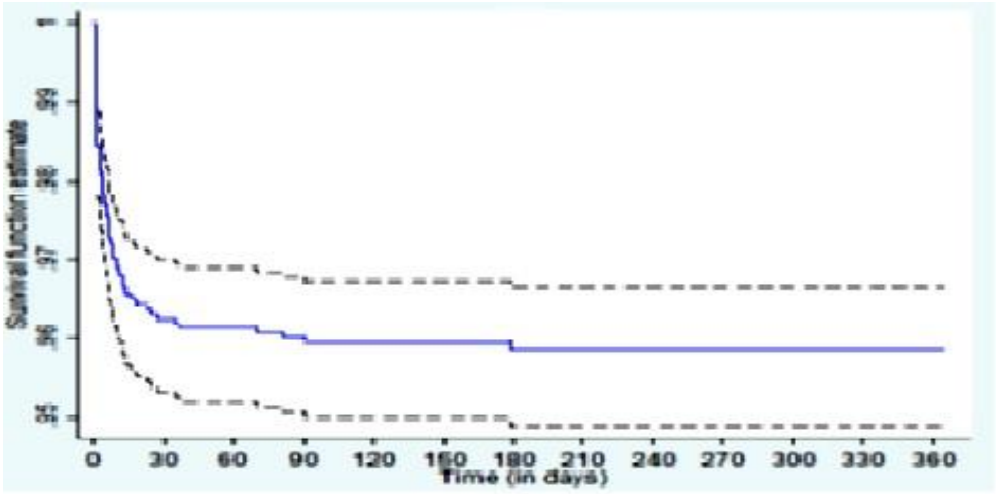


Figure 1: Survival function estimates and 95% confidence limits.

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.

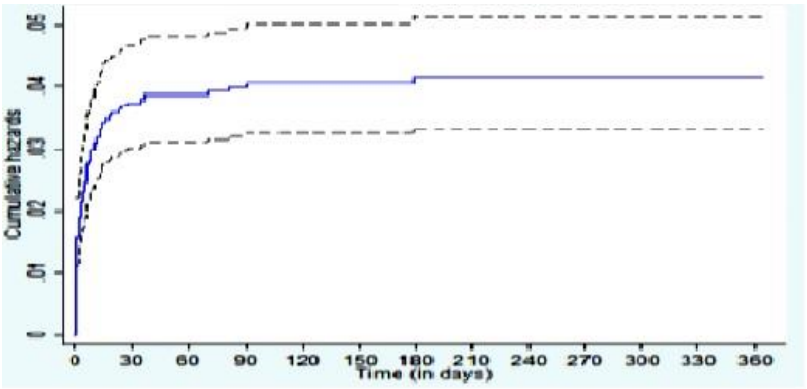


Figure 2: Cumulative hazard function estimate and 95% confidence interval

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

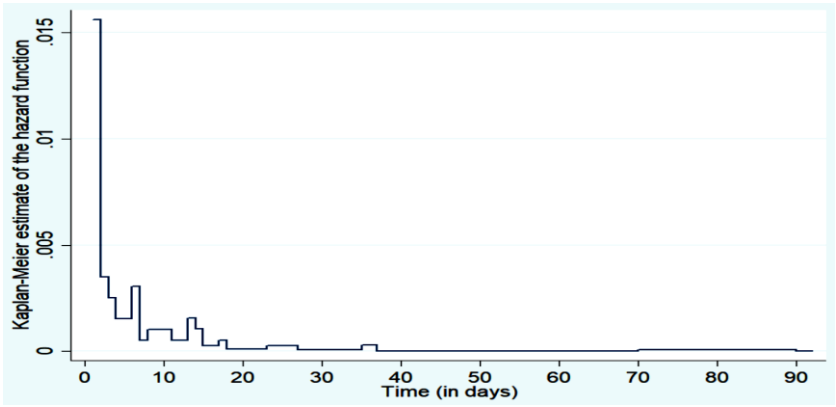


Figure 3: Hazard function estimate and 95% confidence limits.

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.

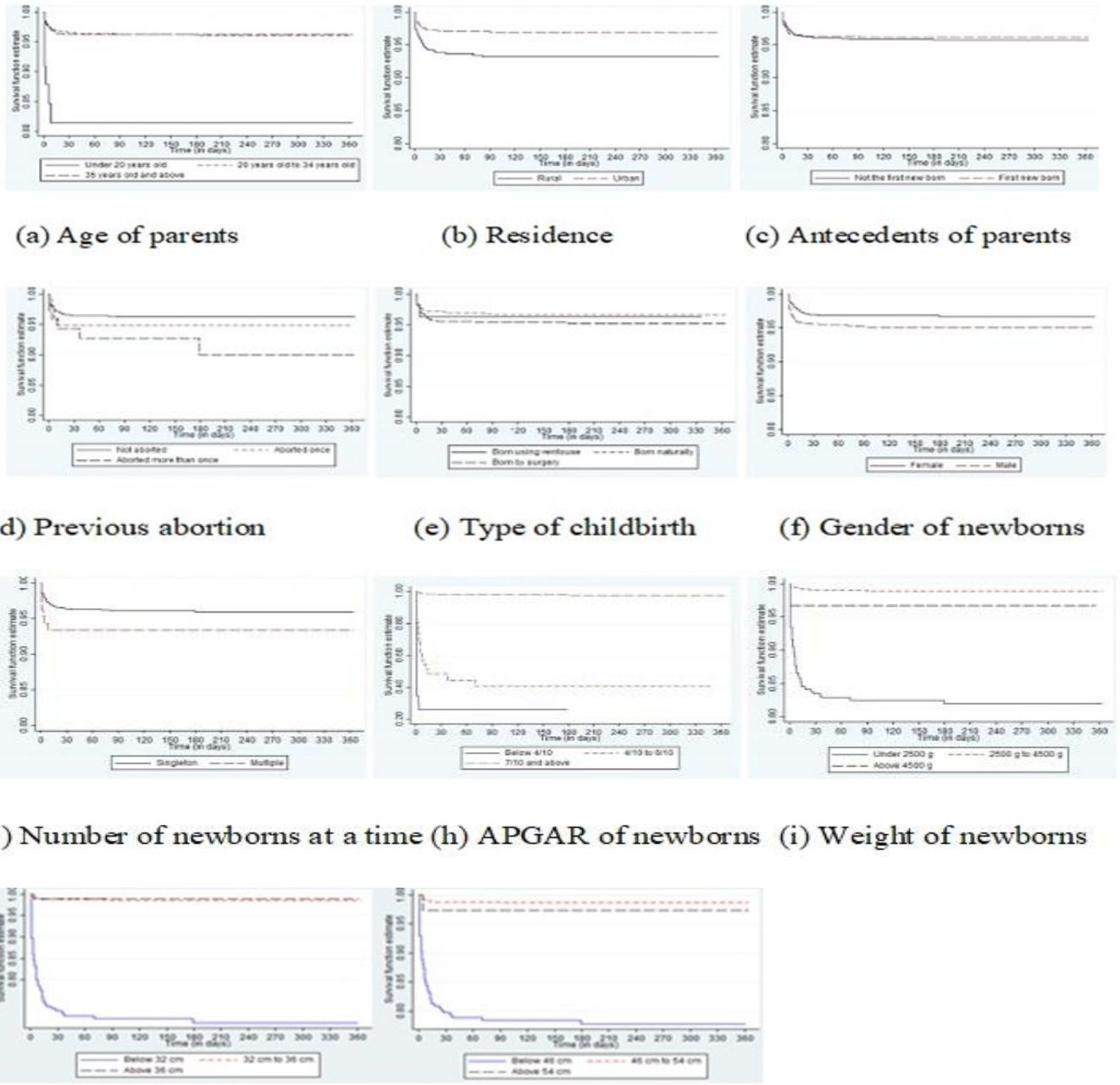


Figure 4: Plots of the Kaplan-Meier estimates of the survival function for variables (a) *age*, (b) *residence*, (c) *antecedents*, (d) *abortion*, (e) *childbirth*, (f) *gender*, (g) *number*, (h) *APGAR*, (i) *weight* (j) *head* and (k) *height* for dataset on newborns at KUTH, year 2016.

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]

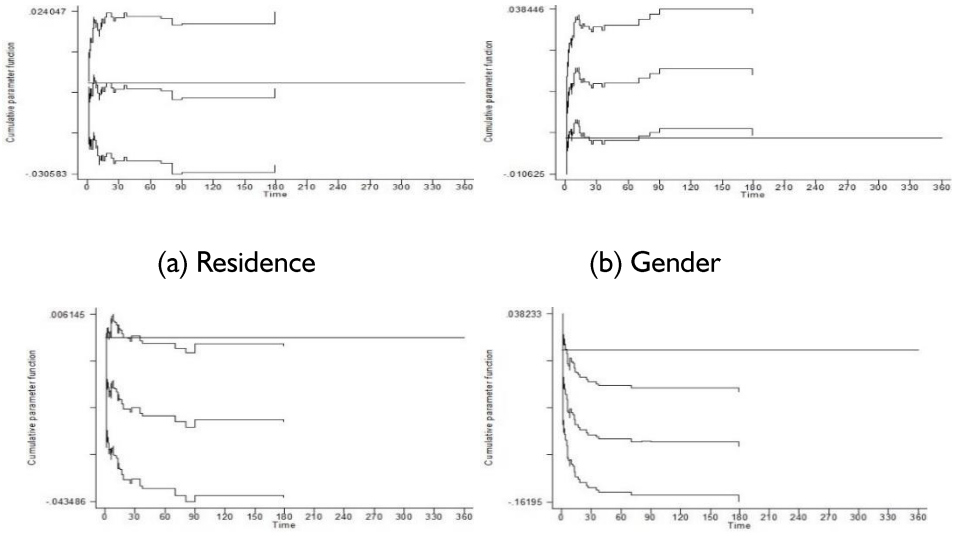
The results in Table 7 indicate significant differences in levels of covariates age, gender, number, APGAR, weight and head where p-values are less or equal to 0.05. The model suggests that the hazard of death of babies whose parents are from 20 years and 34 years old is 0.216 (95% CI:0.084−0.554, p = 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.

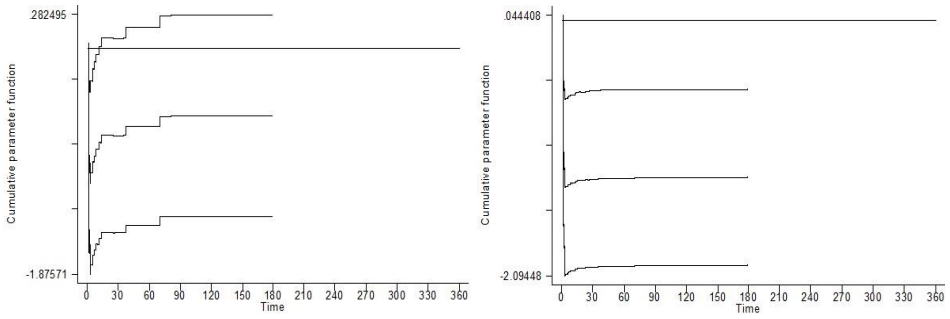
Figure 5 (b) represents the estimated cumulative parameter function with its 95% confidence limits for the male compared to female babies. The plot is approximately horizontal and positive everywhere with the upper and lower limits of the confidence interval situated approximately above the zero line.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

.

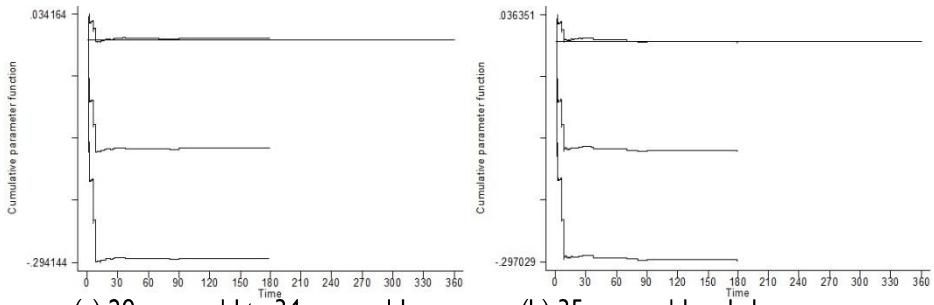


(a) APGAR=4/10 to APGAR=6/10 (b) APGAR=7/10 and above

Figure 6: Cumulative parameter function plots for variable APGAR.

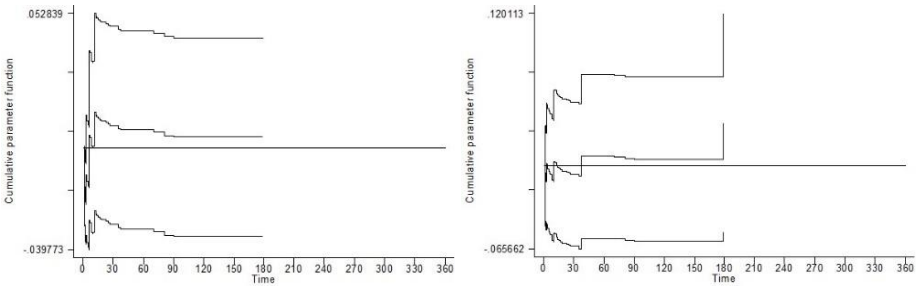
Figure 6 gives the plots of the cumulative parameter functions and their 95% confidence limits for levels of the variable APGAR Figure 6 (a) represents the estimated cumulative parameter function with its 95% confidence limits for the newborn's APGAR from 4/10 to 6/10 compared to the APGAR < 4/10. The 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.

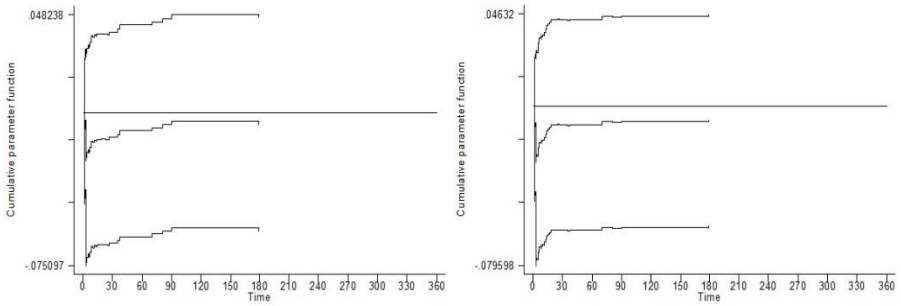


(a) Aborted once (b) Aborted more than once

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.

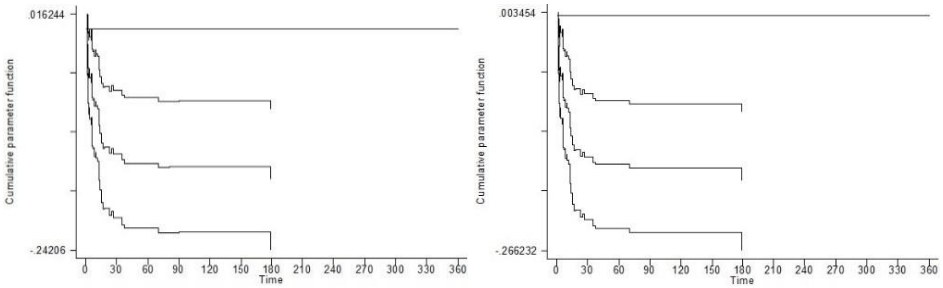
Figure 9 gives the plots of the cumulative parameter functions and their 95% confidence limits for levels of the variable childbirth. Both Figure 9 (a) and 9 (b) behave similarly: the plots are approximately horizontal and negative everywhere, but the upper and lower limits of the confidence interval 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.

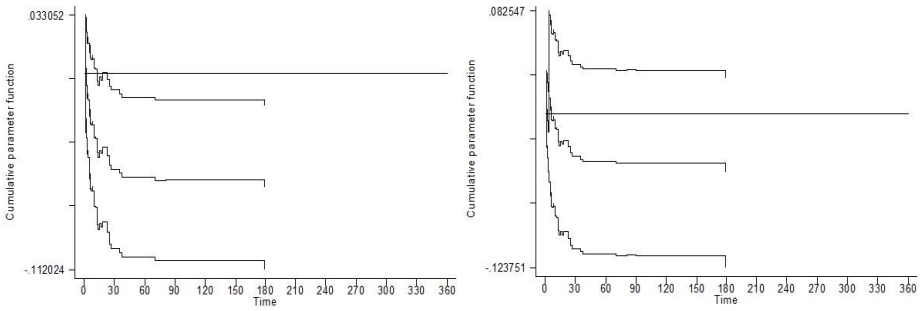
Figure 10 gives the plots of the cumulative parameter functions and their 95% confidence limits for levels of the variable head. Both Figure 10 (a) and 10 (b) behave similarly: the plots are decreasing below the zero line with the upper and lower limits of the confidence interval below the zero line.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

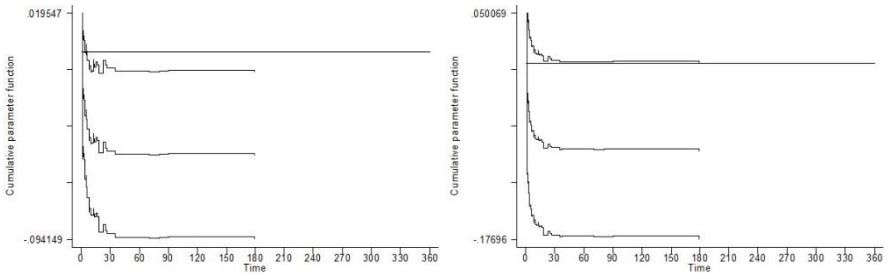
Figure 11 gives the plots of the cumulative parameter functions and their 95% confidence limits for levels of the variable height. Both Figure 11 (a) and 11 (b) behave similarly apart from the upper limit of the confidence interval of at the plot 11 (b) situated above the zero line: the plots are decreasing 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 $KM/\text{se}\left(\hat{\beta}(t)\right)$.

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]

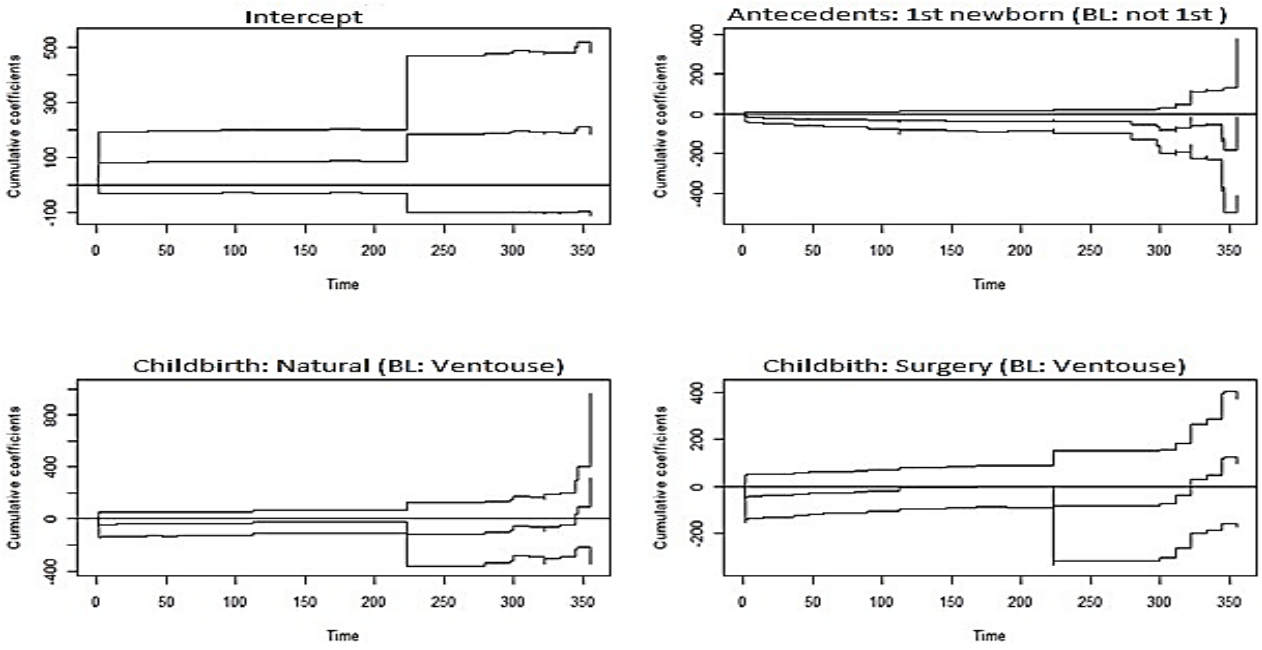


Figure 13: Additive part of the CAHM.

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

This work was supported through the DELTAS Africa Initiative. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [grant 107754/Z/15/Z-DELTAS Africa Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB) programme] and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

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/PMC3250938/

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.

Paper 2

Send Orders for Reprints to reprints@benthamscience.ae

BENTHAM OPEN

CrossMark

The Open Public Health Journal

Content list available at: www.benthamopen.com/TOPHJ/

DOI: 10.2174/1874944501811010464, 2018, *11*, 464-473

RESEARCH ARTICLE

Multiple Events Model for the Infant Mortality at Kigali University Teaching Hospital

P. Gatabazi*, S. F. Melesse and S. Ramroop

School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Private Bag X01 Scottsville - 3209 South Africa

Received: June 22, 2018 Revised: October 4, 2018 Accepted: October 15, 2018

Abstract:

Introduction:

The present study applies multiple events survival analysis to infant mortality at the Kigali University Teaching Hospital (KUTH) in Rwanda.

Materials and Methods:

The primary dataset consists of newborns from KUTH recorded in the year 2016 and in the current paper, a complete case analysis was used. Two events per subject were modeled namely death and the occurrence of at least one of the following conditions that may also cause long-term death to infants such as severe oliguria, severe prematurity, very low birth weight, macrosomia, severe respiratory distress, gastroparesis, hemolytic, trisomy, asphyxia and laparoschisis. Covariates of interest include demographic covariates namely 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 babies include *APGAR*, *gender*, *number* of births at a time, *weight*, circumference of the *head*, and *height*.

Results/Conclusion:

The results revealed that Wei, Lin and Weissfeld Model (WLWM) fit the data well. The covariates *age*, *abortion*, *gender*, *number*, *APGAR*, *weight* and *head* were found to have a significant effect.

Keywords: Survival analysis, Multiple events, Rate function, Mean function, Intensity process, Infant mortality.

1. INTRODUCTION

The multiple events processes or processes that generate events repeatedly along the time are also known as the recurrent event processes [1]. Such processes are adapted to the repeated event data found in medicine and public health, where the number of events exhibited is relatively small for a larger number of processes. Multiple events are met in other domain such as social science, economics, manufacturing, insurance and reliability [2]. In multiple events studies, the number of events in distinct time intervals is termed as "**counts**", the **gaps** are the times between successive event, while the "**event intensity**" is the conditional probability of new event, given the past event, per unit of time [1].

Cook and Lawless [1] discuss different multiplicative approach models such as the **modulated Poisson model** which consists of modelling the intensity processes given the history, and the Cox Models for ordered and unordered events. The interest in this study will be taken on the multiplicative model with the ordered events. Ordered events are based on the concept that the second event cannot occur before the first event, the third event cannot occur before the second event and so on. The models adapted to ordered events include the Andersen-Gill Model (AGM), the Wei, Lin

* Address correspondence to this author at the School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Private Bag X01 Scottsville - 3209 South Africa; Tel: +27710513309; E-mail: gatabazi001@gmail.com

and Weissfeld Model (WLWM) and the Prentice, Williams and Peterson Model (PWPM) [3].

The AGM known also as the counting process approach [4], assumes that all event types are indistinguishable and all events within the same subject are assumed to be independent [5]. Therneau [6] evokes a limitation of AGM of not allowing multiple events to occur at the same time. The WLWM is also known as the marginal risk sets model [7]. The WLWM assumes that events are unordered where each event has its own stratum and each data point appears in all strata. This allows an analysis of multiple events occurring at the same time. The PWPM also known as the conditional risk set model was proposed by Prentice, Williams and Peterson [8]. In PWPM, the set up of the dataset is the same as that of the AGM but the analysis is stratified by failure order [9]. The PWPM can potentially analyse time to each event from the previous event, this is known as the gap-time model. AGM, WLWM and PWPM have been alternatively used on bladder cancer data and on the hospitalisation and death data presented by Castañeda and Gerritse [10].

The WLWM will be used in this study for modelling the risk of infant at Kigali University Teaching Hospital from 01-January-2016 to 31-December-2016 with two events namely death or occurrence of a chronic disease or complication that is due to the type of the dataset where the events of interest can occur on the same day taken as a unit in this study.

Including the introduction, the study comprises four sections: Section 2 is the methodology of the study where the mathematical formulation of AGM, PWPM and WLWM is described. Section 3 gives the main results and interpretation and Section 4 gives conclusion.

2. METHODOLOGY

2.1. Mathematical Formulation of Cox Model with Multiple Events

Consider the time scale t, $t > 0$ and a sample of n individuals under study and let

$N_i(t)$ denotes the number of events for individual i, $i = 1, 2,..., n$,

$T_{i1}T_{i2}$ denote the times of events for individual i,

$W_{ij} = T_{ij}\text{-}T_{i,j\text{-}1}$ denote the gaps or times between successive events of the individual i,

$y_i(t)$'s denote the fixed or time-varying covarites.

$N_i(t)$ is a counting process with intensity process

$$\lambda_i(t) = \lim_{\Delta t \to 0} \frac{P(N_i(t+\Delta t) - N_i(t) = 1 | \mathcal{F}_t}{\Delta t}$$

with $\mathcal{F}_t$ the history of events and covariates up to the time t [11]. The mean cumulative function (MCF) $\mu_i(t)$ and the corresponding rate of occurrence function $\rho_i(t)$ are defined in [1] as:

$$\mu_i(t) = E[N_i(t)] \quad \textbf{(1)}$$

and

$$\rho_i(t) = \frac{d}{dt}\mu_i(t). \quad \textbf{(2)}$$

or

$$\rho_i(t)dt = d\mu_i(t). \quad \textbf{(3)}$$

Applying differentiation on both sides of (1) and using (3) yields:

$$E[dN_i(t] = \rho_i(t)dt \quad \textbf{(4)}$$

Cook and Lawless [1] discuss different multiplicative approaches models such as the regression model for the rate function for both fixed and time-dependent covariates expressed by:

$$\rho_i(t) = \rho_0(t)e^{\beta' y_i(t)}. \quad (5)$$

and the regression model for the mean functions for the fixed covariates, expressed by

$$\mu_i(t) = \mu_0(t)e^{\beta' y_i}$$

where $\mu_0(t) = \int_0^t \rho_0(v)dv.$

The second approach consisted of modelling the intensity process $\lambda_i(t)$ given the history F, that is

$$\lambda_{ik}(t|\mathcal{F}) = \lambda_{0k}(t)e^{\beta' y_{ik}(t)}. \quad (6)$$

The expression λ_k is the event specific baseline hazard for the k^{th} event over time. Model (6) incorporate the AGM, WLWM and the PWPM according to the type of the dataset. Specifically Model (6) yield PWPM gap model of the form

$$\lambda_{ik}(t|\mathcal{F}) = \lambda_{0k}[B(t)]e^{\beta' y_{ik}(t)}.$$

where $B(t) = t\text{-}T_{N(t)}$ is the time since the last event.

2.2. Likelihoods and Maximum Likelihood Estimation

The likelihoods constructions and maximum likelihood estimates for the multiplicative multiple events models are well developed in [1], and specifically [12], discussed a parametric based estimation for the rate function model; Lawless and Nadeau [13] addressed two ways of analyzing the rate function: The first one consists of specifying the distribution of the intensity process $\lambda_i(t)$ such as for example a Poisson process when $\lambda_i(t) = \rho_i(y)$, or a negative binomial process if $\lambda_i(t) = \frac{1+rN_i(t^-)}{1+r\mu_i(t^-)}\rho_i(t)$. In the second way, a distribution of the intensity process is not specified, this approach known as "*robust*" is potential to model means or variances [11].

Assuming that two events cannot occur simultaneously in continuous time, let $]0, \tau_i[$, the interval of time in which the individual i is observed and n_i the number of events of individual i along $]0, \tau_i[$, then the probability density function for the outcome n_i along $]0, \tau_i[$ is given by:

$$L(\mathbf{\Phi}) = \prod_{i=1}^{n} L_i(\phi)$$

where

$$L_i(\mathbf{\Phi}) = \prod_{j=1}^{n_i} \rho_0(T_{ij}, \alpha)e^{\beta' Y_i}\, e^{-\int_0^\tau X_i(v)\rho_0(v,\alpha)e^{\beta' Y_i(v)}dv}. \quad (7)$$

In (7), $\phi = (\alpha, \beta)$; α is called a baseline parameter, $\tau = \max(\tau_1, \tau_2, ...\tau_n)$ and:

$$X_i(v) = \begin{cases} 1 & \text{if individual } i \text{ is at risk} \\ 0 & \text{otherwise.} \end{cases}$$

Using the relationship (4), the log-likelihood can be written:

$$\ln L(\Phi) = \sum_{i=1}^{n} \int_0^{\tau} X_i(v)[\ln \rho_i(v,\Phi) dN_i(v) - \rho_i(v,\Phi) dv]$$

The maximum likelihood estimates are obtained by solving a system:

$$\begin{cases} \frac{\partial \ln L(\boldsymbol{\Phi})}{\partial \boldsymbol{\alpha}} = 0 \\ \frac{\partial \ln L(\boldsymbol{\Phi})}{\partial \boldsymbol{\beta}} = 0 \end{cases} \quad \textbf{(8)}$$

The numerical methods such as the Newton-Raphson method are used for solving the system (8). The adequacy of parameters is checked by finding the elements $I_{\alpha\alpha}$, $I_{\alpha\beta}$, $I_{\beta\alpha}$ and $I_{\beta\beta}$ of the information matrix I and assume that as $n \to \infty$, $\hat{\boldsymbol{\Phi}} - \boldsymbol{\Phi} \rightsquigarrow N\left(0, \mathcal{I}^{-1}(\hat{\boldsymbol{\Phi}})\right)$ [11].

2.3. Setup of Dataset in AGM, PWPM and WLWM

Numerical examples on the layout of dataset in the AGM, the PWPM and the WLWM are found in materials such as [14 - 21]. Assume that n is a maximum number of events per subject, and that τ_k, $k = 1, 2,...n$, are times to events per subject along the study time with range [0, T]. Under the AGM, All events are assumed to be in one stratum along the study time. The study time T is subdivided into intervals defined by the times to events such as 0 - τ_1; τ_1-τ_2; τ_n-T, with event indicator for each time interval. The layout of dataset for PWPM is the same as for the AGM where for each interval corresponds a specific stratum, making the number of time intervals per subject equal to the number of strata per subject. The alternative PWPM based on gape time take 0 at lower bound of each interval per subject, the upper bound is given by the gaps or τ_k-τ_{k-1}, $k = 1, 2,...,n$; the first and the last intervals are respectively 0 - τ_1 and T-τ_n. Like in PWPM, the k^{th} time interval per subject in WLWM is in the k^{th} stratum, $k = 1, 2,...,n$. In WLWM, the study time is subdivided into $n + 1$ intervals each with lower bound 0 and upper bound equal to the time to event, the first and the last intervals are respectively 0 - τ_1 and 0 - T.

2.4. Dataset

The primary dataset of newborns at KUTH is recorded from 1st January to 31st December 2016 and a complete case analysis is considered. Two events per subject are of interest: death and occurrence of at least one chronic disease or complication. The chronic disease or complications recorded at KUTH are severe oliguria, severe prematurity, very low birth weight, macrosomia, severe respiratory distress, gastroparesis, hemolytic, trisomy, asphyxia and laparoschisis. Beside the *event* status and the *time* to an event, eleven covariates are of interest: demographic covariates that include the *age* and the place of *residence* for parents; clinical covariates for parents which 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*, Table **1** describes the variables of interest.

Table 1. 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	Variable Description Codes/Values/Unit	Variable Description Codes/Values/Unit
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, 1aborted 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 to, 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*

(Table 1) contd.....

Variable Description Codes/Values/Unit	Variable Description Codes/Values/Unit	Variable Description Codes/Values/Unit
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 time or not	0 = censored, 1 = dead
	Indicator on the rank of records per subject	1 = first record, 2 = second record

The layout follows the indication provided by the WLWM, Table **2** gives the first 50 entries, the full dataset can be found *via* the authors of this article.

Table 2. First 50 entries of the dataset on infant mortality at KUTH from 01-January-2016 to 31-December-2016.

No	Id	Age	Residence	Antecedents	Abortion	Childbirth	Gender	Number	Apgar	Weight	Head	Height	Tstart	Tstop	Event	N_events
	1	1	0	0	0	2	0	0	2	1	1	1	0	39	0	1
	2	1	0	0	1	1	1	0	2	1	1	1	0	212	0	1
	3	1	1	1	0	1	1	0	2	1	1	1	0	196	0	1
	4	1	1	0	0	2	0	0	2	1	1	1	0	128	0	1
	5	1	0	0	0	2	1	0	2	1	1	1	0	335	0	1
	6	1	1	0	0	1	0	0	2	1	1	1	0	262	0	1
	7	2	0	0	0	1	0	0	2	1	1	1	0	214	0	1
	8	1	0	0	0	1	1	0	2	1	1	1	0	228	0	1
	9	2	1	0	0	1	0	0	2	1	1	1	0	355	0	1
	10	1	1	1	0	2	1	0	2	0	1	1	0	25	0	1
	11	1	1	0	0	2	0	0	2	0	0	0	0	256	0	1
	12	2	0	1	0	2	0	1	2	0	1	1	0	179	0	1
	13	2	0	1	0	2	0	1	2	0	1	1	0	179	0	1
	14	2	0	1	0	2	1	1	2	0	0	0	0	179	0	1
	15	2	0	0	0	2	1	0	2	1	1	1	0	348	0	1
	16	1	1	0	1	2	0	0	2	1	1	1	0	305	0	1
	17	1	1	0	0	1	0	0	2	1	1	1	0	45	0	1
	18	1	0	0	0	2	1	0	2	1	1	1	0	129	0	1
	19	1	1	0	0	2	1	0	0	2	1	1	0	0	1	1
	19	1	1	0	0	2	1	0	0	2	1	1	0	0	1	2
	20	1	1	1	0	2	0	0	2	1	1	1	0	137	0	1
	21	0	1	1	0	1	0	0	2	1	1	1	0	293	0	1
	22	2	1	0	2	1	0	0	2	1	1	1	0	70	0	1
	25	1	0	1	0	1	1	0	1	1	1	1	0	1	1	1
	25	1	0	1	0	1	1	0	1	1	1	1	0	1	1	2
	23	1	1	0	0	1	0	0	2	1	1	1	0	218	0	1
	24	1	1	0	0	1	1	0	2	1	1	1	0	260	0	1
	26	2	0	0	0	2	0	0	2	1	1	1	0	24	0	1
	27	1	1	0	0	2	0	0	2	1	1	1	0	16	0	1
	28	1	0	1	0	2	1	0	2	0	0	0	0	318	0	1
	29	2	0	1	0	2	0	0	2	0	0	0	0	1	1	1
	29	2	0	1	0	2	0	0	2	0	0	0	0	6	1	2
	30	1	1	0	0	1	0	0	2	1	1	1	0	249	0	1
	31	1	1	1	0	2	0	0	2	1	1	1	0	311	0	1
	32	1	1	0	0	1	1	0	2	1	1	1	0	357	0	1
	33	1	1	0	0	1	1	0	2	0	1	1	0	232	0	1
	34	1	1	1	0	1	1	0	2	1	1	1	0	356	0	1
	35	1	1	0	0	1	0	0	2	1	1	1	0	140	0	1
	36	1	1	1	0	2	0	0	2	1	1	2	0	272	0	1

(Table 2) contd.....

No	Id	Age	Residence	Antecedents	Abortion	Childbirth	Gender	Number	Apgar	Weight	Head	Height	Tstart	Tstop	Event	N_events
	37	2	0	1	0	2	0	0	2	1	1	1	0	203	0	1
	38	1	1	0	0	2	1	0	2	1	2	0	0	235	0	1
	39	1	1	0	0	2	0	0	2	1	1	1	0	305	0	1
	40	1	1	1	0	1	0	0	2	1	1	1	0	263	0	1
	41	2	0	0	2	2	0	0	2	0	0	0	0	192	0	1
	42	1	1	0	0	1	1	0	2	1	1	1	0	248	0	1
	43	1	1	0	0	1	1	0	2	0	0	0	0	1	1	1
	43	1	1	0	0	1	1	0	2	0	0	0	0	4	1	2
	44	1	1	0	0	1	1	0	2	1	1	1	0	254	0	1
	45	2	1	0	0	0	0	0	2	1	1	1	0	333	0	1
	46	1	1	0	0	2	1	0	2	1	1	1	0	39	0	1

3. RESULTS AND INTERPRETATION

Model is implemented by using STATA package, version 14 and the dataset on infant mortality at KUTH with a portion given in Table **2**. The WLWM is used since death can occur without a previous chronic disease or complication and the two events could occur at the same time per subject.

Tables **3**, **4** and **5** present the estimates of the hazard ratios of the unadjusted WLWM with ties handling by Breslow, Efron and Cox approaches, respectively. The results in the later two approaches are not far from that of the default method (Breslow). Significant differences in levels are observed for the same covariates in all approaches for the *age*, *abortion*, *gender*, *number*, *APGAR*, *weight* and *head* where p-values are less or equal to.

Table 3. Unadjusted WLWM for the infant mortality at KUTH from 01-January-2016 to 31-December-2016 with Breslow method of ties handling.

Covariate (reference)	Level	Hazard Ratio	Std. Err	z	$P > z$	95% Conf. Int.
Age (Under 20 years old)	20 to 34 years old	0.277	0.997	-3.570	$p < 0.001$	[0.137; 0.560]
	35 years old and above	0.395	0.157	-2.330	0.020	[0.181; 0.863]
Residence (Rural)	Urban	0.847	0.139	-1.020	0.309	[0.614; 1.167]
Antecedents (Not 1st newborn)	1st new born	0.806	0.157	-1.100	0.270	[0.550; 1.182]
Abortion (Not aborted)	Aborted once	1.405	0.398	1.200	0.231	[0.806; 2.448]
	Aborted more than once	0.479	0.161	-2.190	0.028	[0.248; 0.925]
Childbirth (Ventouse)	Natural	0.873	0.491	-0.240	0.808	[0.290; 2.627]
	Surgery	1.115	0.613	0.200	0.843	[0.380; 3.274]
Gender (Female)	Male	1.740	0.296	3.260	0.001	[1.247; 2.429]
Number (Singleton)	Multiple	0.409	0.131	-2.790	0.005	[0.218; 0.766]
APGAR (Below 4/10)	4/10 to 6/10	0.377	0.112	-3.300	0.001	[0.211; 0.673]
	7/10 and above	0.130	0.036	-7.460	$p < 0.001$	[0.076; 0.222]
Weight (Under 2500 g)	2500 g to 4500 g	0.250	0.068	-5.070	$p < 0.001$	[0.146; 0.427]
	Above 4500 g	0.442	0.285	-1.270	0.206	[0.125; 1.565]
Head (Below 32 cm)	32 cm to 36 cm	0.456	0.128	-2.800	0.005	[0.263; 0.789]
	Above 36 cm	0.290	0.219	-1.640	0.102	[0.066; 1.278]
Height (Below 46 cm)	46 cm to 54 cm	0.894	0.276	-0.360	0.716	[0.488; 1.637]
	Above 54 cm	1.670	1.264	0.680	0.498	[0.379; 7.361]

Table 4. Unadjusted WLWM for the infant mortality at KUTH from 01-January-2016 to 31-December-2016 with Efron method of ties handling.

Covariate (reference)	Level	Hazard Ratio	Std. Err.	z	$P > z$	95% Conf. Int.
Age (Under 20 years old)	20 to 34 years old	0.230	0.083	-4.080	$p < 0.001$	[0.114; 0.466]
	35 years old and above	0.324	0.129	-2.840	0.005	[0.149; 0.706]
Residence (Rural)	Urban	0.831	0.137	-1.120	0.261	[0.602; 1.147]
Antecedents (Not 1st newborn)	1st newborn	0.756	0.149	-1.420	0.156	[0.513; 1.113]
Abortion (Not aborted)	Aborted once	1.393	0.396	1.170	0.244	[0.798; 2.430]
	Aborted more than once	0.452	0.154	-2.340	0.020	[0.232; 0.880]

(Table 4) contd.....

Covariate (reference)	Level	Hazard Ratio	Std. Err.	z	*P* > *z*	95% Conf. Int.
Childbirth (Ventouse)	Natural	0.736	0.408	-0.550	0.580	[0.249; 2.179]
	Surgery	0.921	0.499	-0.150	0.880	[0.319; 2.661]
Gender (Female)	Male	1.823	0.312	3.520	$p < 0.001$	[1.304; 2.549]
Number (Singleton)	Multiple	0.324	0.106	-3.430	0.001	[0.170; 0.617]
APGAR (Below 4/10)	4/10 to 6/10	0.214	0.065	-5.090	$p < 0.001$	[0.118; 0.387]
	7/10 and above	0.070	0.020	-9.520	$p < 0.001$	[0.041; 0.121]
Weight (Under 2500 g)	2500 g to 4500 g	0.231	0.063	-5.340	$p < 0.001$	[0.135; 0.395]
	Above 4500 g	0.412	0.269	-1.360	0.174	[0.115; 1.479]
Head (Below 32 cm)	32 cm to 36 cm	0.422	0.119	-3.060	0.002	[0.243; 0.734]
	Above 36 cm	0.246	0.187	-1.840	0.065	[0.055; 1.093]
Height (Below 46 cm)	46 cm to 54 cm	0.917	0.285	-0.280	0.781	[0.499; 1.687]
	Above 54 cm	1.692	1.283	0.690	0.488	[0.383; 7.476]

Table 5. Unadjusted WLWM for the infant mortality at KUTH from 01-January-2016 to 31-December-2016 with Cox method of ties handling.

Covariate (reference)	Level	Hazard Ratio	Std. Err.	z	*P* > *z*	95% Conf. Int.
Age (Under 20 years old)	20 to 34 years old	0.193	0.085	-3.730	$p < 0.001$	[0.081; 0.458]
	35 years old and above	0.267	0.128	-2.760	0.006	[0.104; 0.682]
Residence (Rural)	Urban	0.766	0.150	-1.360	0.175	[0.521; 1.126]
Antecedents (Not 1st newborn)	1st newborn	0.763	0.185	-1.120	0.264	[0.475; 1.226]
Abortion (Not aborted)	Aborted once	1.404	0.453	1.050	0.293	[0.746; 2.643]
	Aborted more than once	0.378	0.152	-2.420	0.015	[0.172; 0.830]
Childbirth (Ventouse)	Natural	0.732	0.481	-0.470	0.635	[0.202; 2.653]
	Surgery	1.016	0.654	0.030	0.980	[0.288; 3.590]
Gender (Female)	Male	1.991	0.405	3.390	0.001	[1.336; 2.966]
Number (Singleton)	Multiple	0.218	0.111	-3.000	0.003	[0.080; 0.589]
APGAR (Below 4/10)	4/10 to 6/10	0.080	0.042	-4.810	$p < 0.001$	[0.029; 0.224]
	7/10 and above	0.021	0.011	-7.840	$p < 0.001$	[0.008; 0.056]
Weight (Under 2500 g)	2500 g to 4500 g	0.236	0.070	-4.850	$p < 0.001$	[0.131; 0.423]
	Above 4500 g	0.378	0.257	-1.430	0.153	[0.100; 1.436]
Head (Below 32 cm)	32 cm to 36 cm	0.391	0.119	-3.100	0.002	[0.216; 0.708]
	Above 36 cm	0.212	0.171	-1.920	0.055	[0.043; 1.033]
Height (Below 46 cm)	46 cm to 54 cm	0.828	0.283	-0.550	0.582	[0.423; 1.620]
	Above 54 cm	1.706	1.351	0.670	0.500	[0.361; 8.060]

The adjusted WLWM with Breslow74, Efron77 and Cox72 methods of ties handling is summarised in Tables **6**, **7** and **8** and the results are not critically different.

Table 6. Adjusted WLWM for the infant mortality at KUTH from 01-January-2016 to 31-December-2016 with Breslow method of ties handling.

Covariate (reference)	Level	Hazard ratio	Std. Err.	z	*P* > *z*	95% Conf. Int.
Age (Under 20 years old)	20 to 34 years old	0.307	0.107	3.380	0.001	[0.155; 0.609]
	35 years old and above	0.472	0.179	-1.980	0.047	[0.225; 0.992]
Abortion (Not aborted)	Aborted once	1.482	0.406	1.430	0.152	[0.866; 2.537]
	Aborted more than once	0.541	0.175	-1.900	0.057	[0.287; 1.019]
Gender (Female)	Male	1.672	0.280	3.070	0.002	[1.204; 2.321]
Number (Singleton)	Multiple	0.401	0.128	-2.860	0.004	[0.214; 0.750]
APGAR (Below 4/10)	4/10 to 6/10	0.414	0.119	-3.080	0.002	[0.236; 0.726]
	7/10 and above	0.144	0.038	-7.350	$p < 0.001$	[0.086; 0.242]
Weight (Under 2500 g)	2500 g to 4500 g	0.238	0.060	-5.650	$p < 0.001$	[0.144; 0.391]
	Above 4500 g	0.447	0.284	-1.270	0.205	[0.129; 1.550]
Head (Below 32 cm)	32 cm to 36 cm	0.420	0.100	-3.660	0.000	[0.264; 0.669]
	Above 36 cm	0.284	0.210	-1.700	0.089	[0.067; 1.211]

Table 7. Adjusted WLWM for the infant mortality at KUTH from 01-January-2016 to 31-December-2016 with Efron method of ties handling.

Covariate (reference)	Level	Hazard Ratio	Std. Err.	z	$P>z$	95% Conf. Int.
Age (Under 20 years old)	20 to 34 years old	0.262	0.092	-3.810	$p<0.001$	[0.132; 0.522]
	35 years old and above	0.407	0.155	-2.360	0.018	[0.193; 0.859]
Abortion (Not aborted)	Aborted once	1.487	0.408	1.440	0.149	[0.868; 2.546]
	Aborted more than once	0.520	0.170	-2.000	0.046	[0.274; 0.987]
Gender (Female)	Male	1.764	0.297	3.370	0.001	[1.268; 2.453]
Number (Singleton)	Multiple	0.308	0.101	-3.580	$p<0.001$	[0.162; 0.586]
APGAR (Below 4/10)	4/10 to 6/10	0.249	0.073	-4.730	$p<0.001$	[0.140; 0.442]
	7/10 and above	0.081	0.022	-9.400	$p<0.001$	[0.048; 0.137]
Weight (Under 2500 g)	2500 g to 4500 g	0.222	0.057	-5.910	$p<0.001$	[0.135; 0.366]
	Above 4500 g	0.430	0.276	-1.310	0.189	[0.122; 1.512]
Head (Below 32 cm)	32 cm to 36 cm	0.388	0.093	-3.940	$p<0.001$	[0.243; 0.622]
	Above 36 cm	0.235	0.175	-1.940	0.052	[0.054; 1.014]

Table 8. Adjusted WLWM for the infant mortality at KUTH from 01-January-2016 to 31-December-2016 with Cox method of ties handling.

Covariate (reference)	Level	Hazard Ratio	Std. Err.	z	$P>z$	95% Conf. Int.
Age (Under 20 years old)	20 to 34 years old	0.218	0.094	-3.520	$p<0.001$	[0.094; 0.509]
	35 years old and above	0.341	0.157	-2.340	0.019	[0.138; 0.841]
Abortion (Not aborted)	Aborted once	1.479	0.459	1.260	0.208	[0.804; 2.719]
	Aborted more than once	0.424	0.161	-2.260	0.024	[0.201; 0.892]
Gender (Female)	Male	1.886	0.374	3.200	0.001	[1.278; 2.783]
Number (Singleton)	Multiple	0.214	0.108	-3.050	0.002	[0.079; 0.576]
APGAR (Below 4/10)	4/10 to 6/10	0.098	0.050	-4.550	$p<0.001$	[0.036; 0.267]
	7/10 and above	0.026	0.012	-7.680	$p<0.001$	[0.010; 0.066]
Weight (Under 2500 g)	2500 g to 4500 g	0.213	0.057	-5.730	$p<0.001$	[0.125; 0.361]
	Above 4500 g	0.364	0.245	-1.500	0.134	[0.097; 1.364]
Head (Below 32 cm)	32 cm to 36 cm	0.349	0.090	-4.080	$p<0.001$	[0.211; 0.579]
	Above 36 cm	0.199	0.160	-2.020	0.044	[0.042; 0.957]

The adjusted model by default (Breslow) suggests that the risk of death or attracting a chronic disease or complication of babies whose parents are from 20 years and 34 years old is 0.307 times that of babies whose parents are under 20 years old (95% CI:0.155-0.609, $p = 0.001$). The risk of death or attracting a chronic disease or complication of babies whose parents aborted more than once previously is 0.541 times that of babies whose parents did not aborted previously (95% CI:0.287-1.019, $p = 0.057$). The risk of death or attracting a chronic disease or complication of babies whose parents are 35 years old and above is 0.472 times that of babies whose parents are under 20 years old (95% CI:0.225-0.992, $p = 0.047$). The risk of death or attracting a chronic disease or complication for male babies is 1.672 times that of female babies (95% CI:1.204-2.321, $p = 0.002$). The risk of death or attracting a chronic disease or complication of multiple babies is 0.401 times that of singleton babies (95% CI:0.214-0.750, $p = 0.004$) The risk of death or attracting a chronic disease or complication for babies whose APGAR range from 4/10 to 6/10 is 0.414 times that of babies whose APGAR is below 4/10 (95% CI:0.236-0.726, $p = 0.002$). The risk of death or attracting a chronic disease or complication for babies whose APGAR range from 7/10 to 10/10 is 0.144 times that of babies whose APGAR is below 4/10 (95% CI:0.086-0.242, $p < 0.001$). The risk of death or attracting a chronic disease or complication for babies whose weight range from 2500 g to 4500 g is 0.238 times that of babies whose weight is below 2500 g (95% CI:0.144-0.391, $p < 0.001$). The risk of death or attracting a chronic disease or complication for babies whose circumference of head range from 32 cm to 36 cm is 0.420 times that of babies whose circumference of head is is below 32 cm (95% CI:0.264-0.669, $p < 0.001$). The risk of death or attracting a chronic disease or complication for babies whose circumference of head is above 36 cm is 0.284 times that of babies whose circumference of head is is below 32 cm (95% CI:0.067-1.211, $p = 0.067$).

CONCLUSION

This paper reviewed different multiplicative multiple events regression models of the time to event survival data namely the mean function regression model, the rate function regression model and the intensity process regression model. The intensity process regression model incorporates the popular models such as Andersen- Gill Model (AGM), Wei, Lin and Weisfeld Model (WLWM) and the Prentice, Williams and Peterson Model (PWPM) following on the layout of the dataset. It was found that data collected at Kigali University Teaching Hospital for 2117 newborns during 365 days of the year 2016 follows the conditions of the WLWM.

The results of the unadjusted WLWM by Breslow, Efron and Cox approaches of ties handling revealed significance on the age of female parents, information on previous abortion, gender of newborn, number of newborns at a time, APGAR, weight of a newborn and the circumference of the head of a newborn. The results of adjusted WLWM by Breslow, Efron and Cox are not critically different. The default approach (Breslow) indicated that the risk of death or attracting a chronic disease or clinical complication of infant is higher in male babies as compared to female babies; it is lower for babies whose parents are from 20 to 34 years old and above 34 years old as compared to babies whose parents are under 20 years old. Babies whose APGAR fall in intervals 4/10 to 7/10 and 7/0 to 10/10 were found to have a better survival outcome than those born with APGAR less than 4/10. Babies with normal weight and overweight were found to have a lower risk as compared to underweight babies. Babies with a normal circumference of head and those with large circumference of head were found to survive better than babies with the relatively small head.

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 *sports activity level*. These variables are not recorded in registry at KUTH.

The future work will consist of testing parametric distribution that could be adapted to the infant mortality at KUTH. The suitable parametric model will be fitted by using appropriate parametric regression model.

ETHICAL APPROVAL AND CONSENT TO PARTICIPATE

We were granted permission to record dataset from registry of the Kigali University Teaching Hospital.

HUMAN AND ANIMAL RIGHTS

No animals / humans were used for the studies that are bases of this research.

CONSENT FOR PUBLICATION

Not applicable.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

This work was supported through the DELTAS Africa Initiative. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [grant 107754/Z/15/Z- DELTAS Africa Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB) programme] and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

REFERENCES

[1] Amorim LD A F, Cai J. Modelling recurrent events: A tutorial for analysis in epidemiology. Int J Epidemiol 2014; 0(0): 1-10. [http://dx.doi.org/10.1093/ije/dyu222] [PMID: 25501468]

[2] Andersen PK, Borgan Ø, Gill RD, Keiding N. Statistical models based on counting processes. Spring-Verlag, New York 1993. [http://dx.doi.org/10.1007/978-1-4612-4348-9]

[3] Andersen PK, Gill RD. Cox's regression model for counting processes: A large sample study. Ann Stat 1992; 10: 1100-20.

[http://dx.doi.org/10.1214/aos/1176345976]

[4] Castañeda J, Gerritse B. Appraisal of several methods to model time to multiple events per subject: Modelling time to hospitalisations and death. Rev Colomb Estad 2010; 33(1): 43-61.

[5] Clayton DG, Cuzick J. Multivariate generalisations of the proportional hazards model (with discussion). J R Stat Soc B 1985; 148: 82-117.

[6] Cook RJ, Lawless JF. Analysis of repeated events. Stat Methods Med Res 2002; 11(2): 141-66.
[http://dx.doi.org/10.1191/0962280202sm278ra] [PMID: 12040694]

[7] Cook RJ, Lawless JF. The statistical analysis of recurrent events. New York: Springer 2007.

[8] Fisher LD. A review of methods for handling multiple endpoints in clinical trials. Proceedings of the American Statistical Association. 43-6.

[9] Hosmer D W, Lemeshow S, May S. Regression modelling of time-to-event data. 2nd. Haboken: John Wiley & Sons Inc. 2008.

[10] Johnson CJ, Boyce MS, Schwartz CC, Haroldson MA. Modelling survival: Application of the Andersen-Gill model to yellowstone grizzly bears. J Wildl Manage 2004; 68: 966-78.
[http://dx.doi.org/10.2193/0022-541X(2004)068[0966:MSAOTA]2.0.CO;2]

[11] Kelly PJ, Lim LL. Survival analysis for recurrent event data: An application to childhood infectious diseases. Stat Med 2000; 19(1): 13-33.
[http://dx.doi.org/10.1002/(SICI)1097-0258(20000115)19:1<13::AID-SIM279>3.0.CO;2-5] [PMID: 10623910]

[12] Lawless JF, Nadeau C. Some simple robust methods for the analysis of recurrent events. Technometrics 2012; 37: 158-68.
[http://dx.doi.org/10.1080/00401706.1995.10484300]

[13] Lin DY, Ying Z. Semiparametric analysis o the additive risk model. Biometrika 1994; 81(1): 61-71.
[http://dx.doi.org/10.1093/biomet/81.1.61]

[14] Lin JS, Wei LJ. Linear regression analysis for multivariate failure time observations. J Am Stat Assoc 1992; 87: 1091-7.
[http://dx.doi.org/10.1080/01621459.1992.10476264]

[15] Louzada F. Intensity models for parametric analysis of recurrent events data. Braz J Probab Stat 2007; 22(1): 23-33.

[16] O'Brien PC. Procedures for comparing samples with multiple endpoints. Biometrics 1984; 40(4): 1079-87.
[http://dx.doi.org/10.2307/2531158] [PMID: 6534410]

[17] Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate incomplete failure time data. Biometrika 1981; 68: 373-9.
[http://dx.doi.org/10.1093/biomet/68.2.373]

[18] Sankaran PG, Anisha P. Shared frailty model for recurrent event data with multiple causes. J Appl Stat 2011; 38(12): 2859-68.
[http://dx.doi.org/10.1080/02664763.2011.570321]

[19] Therneau TM. Extending the cox's model. Proceedings of the First Seattle Symposium in Biostatistics.
[http://dx.doi.org/10.1007/978-1-4684-6316-3_5]

[20] Wei LJ, Glidden DV. An overview of statistical methods for multiple failure time data in clinical trials. Stat Med 1997; 16(8): 833-9.
[http://dx.doi.org/10.1002/(SICI)1097-0258(19970430)16:8<833::AID-SIM538>3.0.CO;2-2] [PMID: 9160483]

[21] Wei LJ, Lin DY, Weissfeld L. Regression analysis of mulyivariate incomplete failure time data by modelling marginal distributions. J Am Stat Assoc 1989; 84(408): 1065-73.
[http://dx.doi.org/10.1080/01621459.1989.10478873]

Paper 3

1874-9445/19

Send Orders for Reprints to reprints@benthamscience.net

136

The Open Public Health Journal

Content list available at: https://openpublichealthjournal.com

RESEARCH ARTICLE

Resampled Cox Proportional Hazards Models for Infant Mortality at the Kigali University Teaching Hospital

Paul Gatabazi*, Sileshi Fanta Melesse and Shaun Ramroop

School of Mathematics, Statistics and Computer Science University of KwaZulu-Natal Pietermaritzburg Private Bag X 01 Scottsville 3209, South Africa

Abstract:

Introduction:

Resampling technique as a way of overcoming instability in Cox Proportional hazard model is used for measuring the risk and related standard error for the infant mortality, given socio-economic and clinical covariates for mother and children at the Kigali University Teaching Hospital in Rwanda.

Methods:

Bootstrap and jackknife Cox proportional hazards models was applied to N=2117 newborn data collected in 2016 at the Kigali University Teaching Hospital in Rwanda.

Results:

The unadjusted models revealed significance of the age of female parents, information on previous abortion, gender of a newborn, number of newborns at a time, APGAR, the weight of a newborn and the circumference of the head of a newborn.

Conclusion:

Statistical analysis supports two major findings: 1) parents under 20 years of age indicate a relatively higher risk of infant death, and 2) abnormality in the newborn's head and weight indicates a relatively higher risk of infant mortality. Recommendations include avoidance of pregnancy until after age 20 and clinically recommended nutrition for the mother during pregnancy to decrease the risk of infant mortality.

Keywords: Infant mortality, Survival analysis, Cox proportional hazards model, Bootstrap, Jackknife, Resampling, Covariate.

Article History	Received: October 21, 2018	Revised: March 06, 2019	Accepted: March 18, 2019

1. INTRODUCTION

The resampling in Cox proportional hazards model consists of conducting the Cox Proportional Hazards Model (CPHM) on a given number of samples obtained after applying a relevant technique of resampling. The popular nonparametric techniques of resampling include *bootstrap method* which is based on random sampling with replacement [1], *jackknife method* which consists of making samples by leaving out one observation a time [1], and *jackknife after bootstrap* [2]. The interest in this study will be on Bootstrap Cox Proportional Hazards Model (BCPHM) and Jackknife Cox Proportional Hazards Model (JCPHM).

* Address correspondence to this author at the School of Mathematics, Statistics and Computer Science University of KwaZulu-Natal Pietermaritzburg Private Bag X 01 Scottsville 3209, South Africa; Tel: +27710513309; E-mail: gatabazi001@gmail.com

Hamada [3] points out the aim of using the resampling technique in CPHM. Firstly the resampling allows the assessment of the stability of the CPHM. The instability may be caused by the correlation of the covariates. Secondly, the resampling may be used when the sample size is relatively small. Model adequacy may be satisfied by selecting variables on which the model is stable rather than testing the proportionality of variables.

BCPHM and JCPHM have been extensively applied to different studies. In [4], bootstrap is applied for estimating the survival function and the hazard rate with respective standard errors. Belašková, Fišerová, and Krupicková [5] published a clinical study which used BCPHM with consideration of right censoring and delayed entries. The study of Belašková *et al.* adapted BCPHM due to the small sample size (N=61). Xu, Sen, and Ying [6] conducted the BCPHM with consideration of

a change-point along the study time with right censored survival data. The study proved the consistency of the model by making a comparison with the model based on data simulation. The JCPHM was adopted by Xiao, Yao-Hua, and Dong-Sheng [7] together with a random weighting which consists of approximating the distribution of the maximum partial likelihood estimates in the CPHM [8 - 10]. Several other manuscripts also discussed the use of the resampled survival analysis including [11 - 17]. In this study, the BCPHM with 1000 bootstrap replicates and the JCPHM were used and compared to the CPHM in modelling the risk of infant death at the Kigali University teaching Hospital from 01-January-2016 to 31-December-2016. The study comprises five sections including the introduction presented in Section 1. Section 2 presents the methods of the study where mathematical formulation of bootstrap and jackknife are reviewed. Section 3 gives the main results. Section 4 discusses the results and Section 5 concludes the paper.

2. METHODOLOGY

2.1. Bootstrap Method

2.1.1. Bootstrap

Assume a sample

$$\mathbf{x} = x_1, x_2, \ldots, x_n,$$

$x_{i,i\in[1,n]}$ are independent and identically distributed with distribution F_θ where θ is the statistical parameter of interest. Consider the distribution function F_{R_n} of a random variable R_n $(\mathbf{x}, F_\theta)$. The bootstrap method as described by Efron and Tibshirani [1], consists of generating

$$\mathbf{x}^* = \mathbf{x}^{*1}, \mathbf{x}^{*2}, \ldots, \mathbf{x}^{*B},$$

where $\mathbf{x}^{*i}_{i,i\in[1,B]}$ are random samples of size n drawn with replacement from the sample $\mathbf{x}$. The varibles $\mathbf{x}^{*i}_{i,i\in[1,B]}$ are independent and identically distributed with distribution $\hat{F}_{\theta,n}$ given $\mathbf{x}$; $\hat{F}_{\theta,n}$ is an estimator of F_θ from $\mathbf{x}$; B is a number of bootstrap samples (replications).

2.1.2. Bootstrap Standard Error

Assume B bootstrap samples $\mathbf{x}^{*1}, \mathbf{x}^{*2}, \ldots, \mathbf{x}^{*B}$. Efron and Tibshirani [1] propose the estimated standard error of the bootstrap statistic of interest $\hat{\theta}$ as:

$$\widehat{se}_B = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left[\hat{\theta}^*(b) - \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^*(b)\right]^2} \quad (1)$$

where $\hat{\theta}$*(b) is an estimate of the statistic of interest from the b^{th} bootstrap sample, b=1,2, ..., B.

2.1.3. Bootstrap Cox Proportional Hazard Model (BCPHM)

Assume a CPHM, h ($t|\mathbf{x}_i$) over the p fixed covariates with values $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$ and the hazard function $h(t)$ when values of all covariates are zeros, that is

$$h(t|\mathbf{x}_i) = h_0(t)\exp(\boldsymbol{\beta}'\mathbf{x}_i) \quad (2)$$

[18], where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$ is a p-dimensional vector of model parameters.

Consider three approaches of approximating the partial likelihood in the presence of tied events namely Breslow [19] approximation of the partial likelihood function given by:

$$L_B(\boldsymbol{\beta}) = \prod_{j=1}^{r}\frac{e^{\boldsymbol{\beta}'\mathbf{s}_j}}{\left[\sum_{l\in\Re(t_{(j)})}e^{\boldsymbol{\beta}'\mathbf{x}_l}\right]^{d_j}}; \quad (3)$$

Efron [20] approximation of the partial likelihood function is given by:

$$L_E(\boldsymbol{\beta}) = \prod_{j=1}^{r}\frac{e^{\boldsymbol{\beta}'\mathbf{s}_j}}{\prod_{k=1}^{d_j}\left[\sum_{l\in\Re(t_{(j)})}e^{\boldsymbol{\beta}'\mathbf{x}_l} - \frac{k-l}{d_j}\sum_{l\in\mathfrak{D}(t_{(j)})}e^{\boldsymbol{\beta}'\mathbf{x}_l}\right]}; \quad (4)$$

and Cox [21] approximation of the partial likelihood function is given by:

$$L_C(\boldsymbol{\beta}) = \prod_{j=1}^{r}\frac{e^{\boldsymbol{\beta}'\mathbf{s}_j}}{\sum_{l\in\Re(t_{(j)},\ d_j)}e^{\boldsymbol{\beta}'\mathbf{s}_l}} \quad (5)$$

where $\Re(t_{(j)},\ d_j)$ is the set of d_j individuals drawn from the risk set $\Re(t_{(j)})$ at time $t_{(j)}$. The inference of model (2) based on bootstrap consists of applying model (2) to each of the B bootstrap samples $\mathbf{x}^{*i}, \forall i \in [1, B]$ of covariates $\mathbf{x}_j, \forall j \in [1, p]$. Bootstrap model parameter estimation uses either Breslow, Efron or Cox approach. The bootstrap standard error is obtained by using Equation (1).

2.2. Jackknife Method

2.2.1. Jackknife

Assume a sample

$$\mathbf{x} = x_1, x_2, \ldots, x_n,$$

where $x_{j,j\in[1,n]}$ are the values of the covariate $\mathbf{x}$. Let θ be a statistic of interest. The jackknife samples consist of leaving out one observation at a time, that is n samples

$\mathbf{x}^{*i} = (x\ , x\ , \ldots, x\quad, x\quad, \ldots, x_n)\ \forall i \in [1, n]$ [1]. The jackknife standard error estimate as proposed [1], is given as:

$$\widehat{se}_{jack} = \sqrt{\frac{n-1}{n}\sum_{i=1}^{n}\left[\hat{\theta}^*(i) - \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}^*(i)\right]^2} \quad (6)$$

where $\hat{\theta}^*(i), i \in [1, n]$ is a statistic of interest for the i^{th} jackknife sample.

2.2.2. Jackknife Cox Proportional Hazard Model (JCPHM)

Model (2) based on jackknife is made by applying it to each of the n jackknife samples $\mathbf{x}^{*i}\ \forall i \in [1, n]$ of covariates $\mathbf{x}_j, \forall j \in [1, p]$. Either Breslow, Efron or Cox approach is used for estimating the jackknife model parameters, with the standard error given by Eq (6).

2.3. Dataset

Table **1** describes the variables of interest and Table **2** summarises the dataset. The full dataset can be obtained from the authors of this article.

Table 1. 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 newborn, 1 = first newborn,
Abortion	Indicator on whether a parent aborted previously	0=not aborted, 1=aborted once, 2= aborted more than once
Child birth	Type of child birth	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 newborn	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 time or not	0=censored, 1=dead
n_events	Indicator on the rank of records per subject	1=first record, 2=second record

The time to event primary dataset of 2117 newborns at the Kigali University Teaching Hospital (KUTH) was recorded from 1st January to 31st December 2016. A complete case analysis is considered where the event is the death of the infant. Eighty-two babies died during the study time, 69 stillborn babies were recorded and 1966 babies were censored. Eleven covariates of interest are demographic covariates that 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 minimum sample size according to Peduzzi *et al*. [22] is $N = \frac{10k}{p}$ where k is the number of predictor variables and p is the number of events. This suggests the minimum sample size at KUTH as:

$$N = \frac{10 \times 11}{0.07132} \approx 1542.$$

Table 2. Summary of newborns under study.

Total Observations	2117
Deaths during the study time	82 (3.873%)
Stillborn babies	69 (3.259%)
Total events	151 (7.132 %)
Censored babies	1966 (92.867%)

3. RESULTS

STATA-15 displays the results in three tables: Table **3** presents estimates of unadjusted CPHM, BCPHM, JCPHM and corresponding adjusted models, by using Breslow estimation method. Both unadjusted and adjusted CPHM, BCPHM and JCPHM by Efron and Cox estimation are also presented in Tables **4** and **5**. The results displayed by the jackknife model are relatively close to that of the Cox proportional hazards model (Table **3**). The standard errors in JCPHM and CPHM are not critically different for all covariates except for the upper levels of covariates *weight*, *head* and *height* where the standard error in JCPHM is more than 40 times that of CPHM. The critical difference in standard error is also observed in BCPHM for the upper levels of covariates *weight*, *head* and *height*, for all levels of covariate *childbirth* and for the covariate *number* where the standard error is relatively higher in BCPHM. Also, BCPHM does not take *age* and *number* as significant covariates unlike the fact of JCPHM and CPHM where these covariates are included in significant covariates. Following suggestions in [23], the χ^2 test statistics suggest a higher performance of the JCPHM as compared to the CPHM and BCPHM since the value of the χ^2 is relatively everywhere lower for the JCPHM.

Table 3. Breslow estimation.

		CPHM					BCPHM					JCPHM				
Covariate (Reference)	**Level**	**HR**	**SE**	**z**	**P>z**	**95% CI**	**HR**	**SE**	**z**	**P>z**	**95% CI**	**HR**	**SE**	**z**	**P>z**	**95% CI**
Age (Under 20 years old)	20 to 34 years old	0.172	0.086	-3.540	p < 0.001	[0.065; 0.456]	0.172	0.254	-1.190	0.234	[0.009; 3.124]	0.172	0.089	-3.400	0.001	[0.062; 0.475]
	35 years old and above	0.216	0.117	-2.840	0.005	[0.075; 0.623]	0.216	0.323	-1.020	0.306	[0.012; 4.058]	0.216	0.124	-2.660	0.008	[0.070; 0.667]
Residence (Rural)	Urban	1.014	0.240	0.060	0.954	[0.637; 1.614]	1.014	0.277	0.050	0.960	[0.594; 1.732]	1.014	0.285	0.050	0.961	[0.585; 1.758]

(Table 3) contd.....

CPHM							BCPHM					JCPHM				
Covariate (Reference)	**Level**	**HR**	**SE**	**z**	**P>z**	**95% CI**	**HR**	**SE**	**z**	**P>z**	**95% CI**	**HR**	**SE**	**z**	**P>z**	**95% CI**
Antecedents (Not 1st newborn)	1st newborn	0.778	0.221	-0.880	0.377	[0.446; 1.358]	0.778	0.223	-0.880	0.381	[0.444; 1.364]	0.778	0.218	-0.900	0.370	[0.449; 1.347]
Abortion (Not aborted)	Aborted once	1.646	0.648	1.270	0.206	[0.761; 3.562]	1.646	0.695	1.180	0.238	[0.720; 3.763]	1.646	0.664	1.230	0.217	[0.746; 3.633]
	Aborted more than once	1.111	0.503	0.230	0.817	[0.457; 2.700]	1.111	2.084	0.060	0.955	[0.028; 43.927]	1.111	0.556	0.210	0.834	[0.416; 2.966]
Childbirth (Ventouse)	Natural	0.593	0.449	-0.690	0.490	[0.135; 2.612]	0.593	3.846	-0.080	0.936	[0.000; 1.963x 10^{5}]	0.593	0.469	-0.660	0.509	[0.126; 2.797]
	Surgery	0.777	0.580	-0.340	0.736	[0.180; 3.358]	0.777	5.021	-0.040	0.969	[0.000; 2.443x 10^{5}]	0.777	0.611	-0.320	0.749	[0.166; 3.630]
Gender (Female)	Male	1.964	0.472	2.810	0.005	[1.227; 3.146]	1.964	0.480	2.760	0.006	[1.217; 3.170]	1.964	0.504	2.630	0.009	[1.188; 3.248]
Number (Singleton)	Multiple	0.306	0.136	-2.660	0.008	[0.128; 0.732]	0.306	0.730	-0.500	0.620	[0.003; 32.826]	0.306	0.136	-2.670	0.008	[0.128; 0.729]
APGAR (Below 4/10)	4/10 to 6/10	0.335	0.133	-2.760	0.006	[0.154; 0.729]	0.335	0.160	-2.290	0.022	[0.131; 0.856]	0.335	0.157	-2.340	0.020	[0.134; 0.839]
	7/10 and above	0.049	0.019	-7.860	p < 0.001	[0.023; 0.103]	0.049	0.020	-7.300	p < 0.001	[0.022; 0.110]	0.049	0.020	-7.380	p < 0.001	[0.022; 0.109]
Weight (Under 2500 g)	2500 g to 4500 g	0.227	0.089	-3.790	p < 0.001	[0.105; 0.489]	0.227	0.102	-3.300	0.001	[0.094; 0.548]	0.227	0.105	-3.210	0.001	[0.091; 0.561]
	Above 4500 g	0.392	0.421	-0.870	0.383	[0.048; 3.213]	0.392	8.103	-0.050	0.964	[0.000; 1.600x 10^{17}]	0.392	17.310	-0.020	0.983	[0.000; 1.740x 10^{37}]
Head (Below 32 cm)	32 cm to 36 cm	0.288	0.111	-3.230	0.001	[0.136; 0.613]	0.288	0.121	-2.960	0.003	[0.127; 0.658]	0.288	0.116	-3.090	0.002	[0.131; 0.635]
	Above 36 cm	0.122	0.128	-2.010	0.045	[0.016; 0.951]	0.122	2.449	-0.100	0.917	[0.000; 1.670x 10^{16}]	0.122	5.426	-0.050	0.962	[0.000; 1.220x 10^{37}]
Height (Below 36 cm)	46 cm to 54 cm	0.567	0.235	-1.370	0.171	[0.251; 1.278]	0.567	0.240	-1.340	0.180	[0.247; 1.300]	0.567	0.247	-1.300	0.193	[0.241; 1.334]
	Above 54 cm	1.020	1.100	0.020	0.986	[0.123; 8.444]	1.020	21.073	0.000	0.999	[0.000; 3.980x 10^{17}]	1.020	44.687	0.000	1.000	[0.000; 2.150x 10^{37}]
Adjusted CPHM							**Adjusted BCPHM**					**Adjusted JCPHM**				
Age (Under 20 years old)	20 to 34 years old	0.215	0.105	-3.150	0.002	[0.083; 0.559]	-	-	-	-	-	0.215	0.104	-3.190	0.001	[0.084; 0.554]
	35 years old and above	0.308	0.159	-2.280	0.023	[0.112; 0.848]	-	-	-	-	-	0.308	0.160	-2.270	0.023	[0.111; 0.852]
Gender (Female)	Male	1.942	0.459	2.810	0.005	[1.222; 3.085]	1.562	0.350	1.990	0.046	[1.007; 2.424]	1.942	0.476	2.700	0.007	[1.200; 3.142]
Number (Singleton)	Multiple	0.264	0.115	-3.060	0.002	[0.112; 0.619]	-	-	-	-	-	0.264	0.117	-3.010	0.003	[0.111; 0.629]
APGAR (Below 4/10)	4/10 to 6/10	0.411	0.154	-2.380	0.017	[0.198; 0.856]	0.695	0.288	-0.880	0.379	[0.308; 1.565]	0.411	0.185	-1.970	0.049	[0.170; 0.995]
	7/10 and above	0.059	0.021	-7.850	p < 0.001	[0.029; 0.119]	0.100	0.039	-5.880	p < 0.001	[0.046; 0.215]	0.059	0.024	-6.810	p < 0.001	[0.026; 0.133]
Weight (Under 2500 g)	2500 g to 4500 g	0.181	0.064	-4.860	p < 0.001	[0.091; 0.361]	0.200	0.084	-3.840	p < 0.001	[0.088; 0.455]	0.181	0.071	-4.390	p < 0.001	[0.084; 0.389]
	Above 4500 g	0.372	0.384	-0.960	0.338	[0.049; 2.809]	0.438	8.985	-0.040	0.968	[0.000; 1.280x 10^{17}]	0.372	16.296	-0.020	0.982	[0.000; 6.880x 10^{36}]
Head (Below 32 cm)	32 cm to 36 cm	0.208	0.068	-4.830	p < 0.001	[0.110; 0.394]	0.216	0.088	-3.760	p < 0.001	[0.097; 0.480]	0.208	0.080	-4.060	p < 0.001	[0.098; 0.444]

(Table 3) contd.....

CPHM							BCPHM					JCPHM				
Covariate (Reference)	**Level**	**HR**	**SE**	**z**	**P>z**	**95% CI**	**HR**	**SE**	**z**	**P>z**	**95% CI**	**HR**	**SE**	**z**	**P>z**	**95% CI**
	Above 36 cm	0.105	0.109	-2.180	0.029	[0.014; 0.797]	0.109	2.234	-0.110	0.914	[0.000; 2.600x 10^{16}]	0.105	4.680	-0.050	0.960	[0.000; 9.160x 10^{36}]
		x^2= 300.360, p < 0.001					x^2 = 296.290, p < 0.001					x^2= 32.310, p < 0.001				

Table 4. Efron estimation.

CPHM							BCPHM					JCPHM				
Covariate (Reference)	**Level**	**HR**	**SE**	**z**	**P>z**	**95% CI**	**HR**	**SE**	**z**	**P>z**	**95% CI**	**HR**	**SE**	**z**	**P>z**	**95% CI**
Age (Under 20 years old)	20 to 34 years old	0.160	0.079	-3.680	p < 0.001	[0.060; 0.424]	0.160	0.323	-0.910	0.364	[0.003; 8.374]	0.160	0.087	-3.370	0.001	[0.055; 0.464]
	35 years old and above	0.199	0.107	-2.990	0.003	[0.069; 0.573]	0.199	0.406	-0.790	0.429	[0.004; 10.896]	0.199	0.120	-2.680	0.007	[0.061; 0.648]
Residence (Rural)	Urban	1.029	0.246	0.120	0.907	[0.643; 1.645]	1.029	0.307	0.090	0.925	[0.573; 1.847]	1.029	0.314	0.090	0.927	[0.565; 1.871]
Antecedents (Not 1st newborn)	1st newborn	0.723	0.212	-1.110	0.268	[0.407; 1.283]	0.723	0.227	-1.030	0.301	[0.391; 1.337]	0.723	0.233	-1.010	0.314	[0.384; 1.359]
Abortion (Not aborted)	Aborted once	1.588	0.628	1.170	0.242	[0.732; 3.448]	1.588	0.696	1.060	0.291	[0.673; 3.749]	1.588	0.659	1.110	0.265	[0.704; 3.585]
	Aborted more than once	1.147	0.519	0.300	0.762	[0.473; 2.782]	1.147	4.651	0.030	0.973	[0.000; 3.251x 10^{3}]	1.147	0.587	0.270	0.789	[0.420; 3.127]
Childbirth (Ventouse)	Natural	0.532	0.400	-0.840	0.401	[0.122; 2.319]	0.532	3.646	-0.090	0.927	[0.000; 3.605x 10^{5}]	0.532	0.448	-0.750	0.454	[0.102; 2.772]
	Surgery	0.695	0.515	-0.490	0.624	[0.163; 2.969]	0.695	4.766	-0.050	0.958	[0.000; 4.743x 10^{5}]	0.695	0.579	-0.440	0.663	[0.136; 3.558]
Gender (Female)	Male	2.061	0.500	2.980	0.003	[1.282; 3.315]	2.061	0.556	2.680	0.007	[1.215; 3.496]	2.061	0.592	2.520	0.012	[1.173; 3.621]
Number (Singleton)	Multiple	0.243	0.113	-3.040	0.002	[0.098; 0.606]	0.243	0.135	-2.540	0.011	[0.082; 0.724]	0.243	0.141	-2.440	0.015	[0.078; 0.759]
APGAR (Below 4/10)	4/10 to 6/10	0.207	0.084	-3.880	p < 0.001	[0.094; 0.460]	0.207	0.116	-2.820	0.005	[0.070; 0.618]	0.207	0.120	-2.710	0.007	[0.066; 0.648]
	7/10 and above	0.030	0.012	-8.960	p < 0.001	[0.014; 0.065]	0.030	0.015	-7.070	p < 0.001	[0.011; 0.080]	0.030	0.016	-6.750	p < 0.001	[0.011; 0.083]
Weight (Under 2500 g)	2500 g to 4500 g	0.222	0.088	-3.800	p < 0.001	[0.102; 0.483]	0.222	0.105	-3.180	0.001	[0.088; 0.562]	0.222	0.107	-3.110	0.002	[0.086; 0.574]
	Above 4500 g	0.389	0.426	-0.860	0.389	[0.045; 3.338]	0.389	8.081	-0.050	0.964	[0.000; 1.950x 10^{17}]	0.389	17.369	-0.020	0.983	[0.000; 4.530x 10^{37}]
Head (Below 32 cm)	32 cm to 36 cm	0.284	0.110	-3.250	0.001	[0.133; 0.607]	0.284	0.115	-3.100	0.002	[0.129; 0.629]	0.284	0.119	-3.000	0.003	[0.125; 0.647]
	Above 36 cm	0.110	0.117	-2.070	0.038	[0.014; 0.886]	0.110	2.350	-0.100	0.918	[0.000; 1.590x 10^{17}]	0.110	3.679	-0.070	0.947	[0.000; 3.080x 10^{27}]
Height (Below 36 cm)	46 cm to 54 cm	0.569	0.238	-1.350	0.177	[0.251; 1.291]	0.569	0.252	-1.270	0.202	[0.239; 1.354]	0.569	0.273	-1.180	0.240	[0.222; 1.457]
	Above 54 cm	1.010	1.094	0.010	0.993	[0.121; 8.431]	1.010	21.269	0.000	1.000	[0.000; 18.460x 10^{17}]	1.010	44.776	0.000	1.000	[0.000; 5.730x 10^{37}]
Adjusted CPHM							**Adjusted BCPHM**					**Adjusted JCPHM**				
Age (Under 20 years old)	20 to 34 years old	0.201	0.098	-3.280	0.001	[0.077; 0.524]	-	-	-	-	-	0.201	0.102	-3.170	0.002	[0.075; 0.543]

(Table 4) contd.....

CPHM							BCPHM					JCPHM				
Covariate (Reference)	**Level**	**HR**	**SE**	**z**	**P>z**	**95% CI**	**HR**	**SE**	**z**	**P>z**	**95% CI**	**HR**	**SE**	**z**	**P>z**	**95% CI**
	35 years old and above	0.293	0.152	-2.360	0.018	[0.106; 0.811]	-	-	-	-	-	0.293	0.160	-2.250	0.025	[0.101; 0.856]
Gender (Female)	Male	2.071	0.495	3.050	0.002	[1.297; 3.308]	1.562	0.400	1.740	0.081	[0.946; 2.579]	2.071	0.587	2.570	0.010	[1.188; 3.611]
Number (Singleton)	Multiple	0.205	0.092	-3.520	p < 0.001	[0.085; 0.495]	-	-	-	-	-	0.205	0.118	-2.740	0.006	[0.066; 0.637]
APGAR (Below 4/10)	4/10 to 6/10	0.273	0.103	-3.430	0.001	[0.130; 0.573]	0.545	0.273	-1.210	0.226	[0.204; 1.457]	0.273	0.169	-2.100	0.036	[0.081; 0.919]
	7/10 and above	0.038	0.014	-8.980	p < 0.001	[0.019; 0.078]	0.077	0.036	-5.440	p < 0.001	[0.030; 0.193]	0.038	0.023	-5.530	p < 0.001	[0.012; 0.122]
Weight (Under 2500 g)	2500 g to 4500 g	0.179	0.063	-4.890	p < 0.001	[0.090; 0.356]	0.201	0.083	-3.880	0.000	[0.089; 0.452]	0.179	0.071	-4.360	p < 0.001	[0.082; 0.388]
	Above 4500 g	0.379	0.396	-0.930	0.353	[0.049; 2.938]	0.477	9.872	-0.040	0.971	[0.000; 2.040x 10^{17}]	0.379	16.849	-0.020	0.983	[0.000; 2.970x 10^{37}]
Head (Below 32 cm)	32 cm to 36 cm	0.205	0.067	-4.860	p < 0.001	[0.108; 0.388]	0.215	0.090	-3.680	p < 0.001	[0.095; 0.487]	0.205	0.081	-4.030	p < 0.001	[0.095; 0.443]
	Above 36 cm	0.095	0.100	-2.250	0.025	[0.012; 0.740]	0.105	2.180	-0.110	0.914	[0.000; 5.960x 10^{16}]	0.095	4.226	-0.050	0.958	[0.000; 5.340x 10^{36}]
		χ^2= 316.160, p < 0.001					χ^2 = 297.200, p < 0.001					χ^2 = 29.760, p < 0.001				

Table 5. Cox estimation.

CPHM							BCPHM					JCPHM				
Covariate (Reference)	**Level**	**HR**	**SE**	**z**	**P>z**	**95% CI**	**HR**	**SE**	**z**	**P>z**	**95% CI**	**HR**	**SE**	**z**	**P>z**	**95% CI**
Age (Under 20 years old)	20 to 34 years old	0.140	0.075	-3.690	p < 0.001	[0.050; 0.398]	0.140	0.257	-1.070	0.283	[0.004; 5.064]	0.140	0.084	-3.260	0.001	[0.043; 0.457]
	35 years old and above	0.171	0.098	-3.090	0.002	[0.056; 0.523]	0.171	0.313	-0.960	0.335	[0.005; 6.216]	0.171	0.111	-2.710	0.007	[0.048; 0.613]
Residence (Rural)	Urban	1.003	0.258	0.010	0.990	[0.606; 1.660]	1.003	0.347	0.010	0.993	[0.510; 1.974]	1.003	0.342	0.010	0.993	[0.514; 1.956]
Antecedents (Not 1st newborn)	1st newborn	0.726	0.231	-1.010	0.313	[0.389; 1.353]	0.726	0.280	-0.830	0.406	[0.341; 1.545]	0.726	0.268	-0.870	0.386	[0.351; 1.498]
Abortion (Not aborted)	Aborted once	1.671	0.686	1.250	0.211	[0.748; 3.735]	1.671	0.763	1.120	0.261	[0.683; 4.091]	1.671	0.722	1.190	0.234	[0.717; 3.897]
	Aborted more than once	1.388	0.697	0.650	0.514	[0.519; 3.712]	1.388	0.756	0.600	0.548	[0.477; 4.038]	1.388	0.849	0.540	0.593	[0.418; 4.609]
Childbirth (Ventouse)	Natural	0.533	0.422	-0.790	0.427	[0.113; 2.517]	0.533	3.473	-0.100	0.923	[0.000; 1.883x 10^{5}]	0.533	0.449	-0.750	0.456	[0.102; 2.786]
	Surgery	0.759	0.590	-0.360	0.722	[0.166; 3.479]	0.759	4.946	-0.040	0.966	[0.000; 2.683x 10^{5}]	0.759	0.628	-0.330	0.739	[0.150; 3.850]
Gender (Female)	Male	2.195	0.570	3.030	0.002	[1.319; 3.652]	2.195	0.672	2.570	0.010	[1.204; 3.999]	2.195	0.695	2.480	0.013	[1.179; 4.086]
Number (Singleton)	Multiple	0.203	0.110	-2.950	0.003	[0.071; 0.585]	0.203	0.693	-0.470	0.640	[0.000; 162.000]	0.203	0.196	-1.650	0.099	[0.031; 1.353]
APGAR (Below 4/10)	4/10 to 6/10	0.167	0.085	-3.500	p < 0.001	[0.061; 0.455]	0.167	0.602	-0.500	0.620	[0.000; 197.300]	0.167	0.180	-1.660	0.098	[0.020; 1.392]
	7/10 and above	0.022	0.010	-8.140	p < 0.001	[0.009; 0.055]	0.022	0.078	-1.070	0.284	[0.000; 24.091]	0.022	0.021	-3.880	p < 0.001	[0.003; 0.151]
Weight (Under 2500 g)	2500 g to 4500 g	0.221	0.088	-3.790	p < 0.001	[0.101; 0.482]	0.221	0.105	-3.180	0.001	[0.087; 0.560]	0.221	0.105	-3.170	0.002	[0.087; 0.562]

(Table 5) contd.....

CPHM							BCPHM					JCPHM				
Covariate (Reference)	**Level**	**HR**	**SE**	**z**	**P>z**	**95% CI**	**HR**	**SE**	**z**	**P>z**	**95% CI**	**HR**	**SE**	**z**	**P>z**	**95% CI**
	Above 4500 g	0.324	0.362	-1.010	0.313	[0.036; 2.892]	0.324	6.266	-0.060	0.954	[0.000; 9.150x 10^{15}]	0.324	10.526	-0.030	0.972	[0.000; 1.450x 10^{27}]
Head (Below 32 cm)	32 cm to 36 cm	0.285	0.110	-3.240	0.001	[0.133; 0.609]	0.285	0.119	-3.020	0.003	[0.126; 0.644]	0.285	0.117	-3.050	0.002	[0.127; 0.639]
	Above 36 cm	0.106	0.114	-2.090	0.036	[0.013; 0.866]	0.106	2.091	-0.110	0.909	[0.000; 5.660x 10^{15}]	0.106	3.780	-0.060	0.950	[0.000; 1.910x 10^{29}]
Height (Below 36 cm)	46 cm to 54 cm	0.539	0.226	-1.480	0.140	[0.237; 1.225]	0.539	0.236	-1.410	0.158	[0.229; 1.270]	0.539	0.252	-1.320	0.186	[0.216; 1.346]
	Above 54 cm	1.037	1.120	0.030	0.973	[0.125; 8.613]	1.037	20.074	0.000	0.998	[0.000; 3.080x 10^{16}]	1.037	41.221	0.000	0.999	[0.000; 7.290x 10^{33}]
Adjusted CPHM							**Adjusted BCPHM**					**Adjusted JCPHM**				
Age (Under 20 years old)	20 to 34 years old	0.173	0.092	-3.310	0.001	[0.061; 0.488]	-	-	-	-	-	0.181	0.096	-3.230	0.001	[0.064; 0.511]
	35 years old and above	0.250	0.139	-2.490	0.013	[0.084; 0.745]	-	-	-	-	-	0.248	0.139	-2.490	0.013	[0.083; 0.744]
Gender (Female)	Male	2.150	0.550	2.990	0.003	[1.302; 3.549]	2.031	0.473	3.050	0.002	[1.287; 3.205]	1.778	0.506	2.020	0.043	[1.018; 3.106]
Number (Singleton)	Multiple	0.176	0.091	-3.350	0.001	[0.064; 0.486]	-	-	-	-	-	-	-	-	-	-
APGAR (Below 4/10)	4/10 to 6/10	0.249	0.114	-3.030	0.002	[0.101; 0.612]	-	-	-	-	-	0.516	0.330	-1.030	0.301	[0.147; 1.809]
	7/10 and above	0.030	0.013	-8.220	p < 0.001	[0.013; 0.069]	-	-	-	-	-	0.060	0.035	-4.820	p < 0.001	[0.019; 0.188]
Weight (Under 2500 g)	2500 g to 4500 g	0.176	0.062	-4.910	p < 0.001	[0.088; 0.352]	0.149	0.053	-5.380	p < 0.001	[0.075; 0.299]	0.209	0.082	-3.990	p < 0.001	[0.097; 0.451]
	Above 4500 g	0.325	0.347	-1.050	0.293	[0.040; 2.636]	0.367	6.399	-0.060	0.954	[0.000; 2.450x 10^{14}]	0.425	16.781	-0.020	0.983	[0.000; 1.840x 10^{33}]
Head (Below 32 cm)	32 cm to 36 cm	0.196	0.064	-5.020	p < 0.001	[0.103; 0.370]	0.120	0.038	-6.700	p < 0.001	[0.065; 0.224]	0.198	0.077	-4.180	p < 0.001	[0.093; 0.423]
	Above 36 cm	0.090	0.095	-2.290	0.022	[0.011; 0.706]	0.073	1.284	-0.150	0.882	[0.000; 7.170x 10^{13}]	0.098	4.324	-0.050	0.958	[0.000; 3.360x 10^{36}]
		$\mathbf{x}^2$ = 316.430, p < 0.001					$\mathbf{x}^2$ = 210.070, p < 0.001					$\mathbf{x}^2$ = 31.380, p < 0.001				

4. DISCUSSION

The resampling methods adopted in the Cox Proportional Hazard Model (CPHM) include Bootstrap Cox Proportional Hazards Model (BCPHM) and Jackknife Cox Proportional Hazards Model (JCPHM) with three approaches of ties handling. The results by different approaches of ties handling are not critically different as expected. The analysis is then made on the STATA-15 default method [19]. The similarity observed between the results of JCPHM and those of CPHM is relatively stronger than that of BCPHM and CPHM. The similarity between CPHM and JCPHM suggests that the CPHM may be stable. The overall analysis confirms the significant difference of levels of covariates *age, gender*, *number*, *APGAR*, *weight* and *head*. The results show relatively higher risk of babies from under 20 years old parents as compared to the older parents, that is 4.651 times that of babies whose parents' ages range from 20 to 34 years, and 3.247 times that of babies whose parents are 35 years old and above. The risk of male babies is 1.942 times that of female babies. The risk of multiple babies is 0.264 times that of singleton babies. Babies with APGAR below 4/10 are at a relatively higher risk, that is 2.433 times that babies with APGAR ranging from 4/10 to 6/10 and 16.949 times that of babies whose APGAR range from 7/10 to 10/10. The risk of babies whose weight is below 2500 g is 5.525 times that of babies whose weight range from 2500 g to 4500 g and 2.688 times that of babies with weight above 4500 g. The risk for babies born with a circumference of head below 32 cm is 4.808 times that of newborns whose circumference of head ranges from 32 cm to 36 cm, and 9.524 times that of newborns whose circumference of head is above 36 cm.

The results of BCPHM are also close to that of JCPHM and CPHM for all significant covariates but the model shows a relatively high standard error for non-significant levels of covariates. The critical discrepancy between standard errors after resampling for some covariates suggests instability of the CPHM at these specific covariates and this emphasizes their non-significance in the CPHM.

The dataset was recorded for one year. The stability of the adjusted CPHM is justified by the non-critical difference between the adjusted resampled models.

CONCLUSION

This paper reviewed different methods of resampling in Cox Proportional Hazards Model (CPHM) namely the Bootstrap Cox Proportional Hazards Model (BCPHM) and the Jackknife Cox Proportional Hazards Model (JCPHM). The results after resampling are compared to that of the CPHM for three different ties handling methods namely Breslow, Efron and Cox approximation. The test statistics show everywhere a higher performance of the JCPHM as compared to the CPHM and BCPHM.

The results displayed by the JCPHM and CPHM are very close and suggested the significance of the age of female parent, information on previous abortion, the gender of a newborn, the number of newborns at a time, APGAR, the weight of a newborn and the circumference of the head of a newborn. Male babies are at a relatively higher risk as compared to female babies. The risk is higher for babies whose parents are under 20 years old as compared to older parents. Babies born with APGAR less than 4/10 were found to have a higher risk as compared to newborns with APGAR greater than 4/10. Underweight babies were found to have a higher risk as compared to babies with normal weight and overweight. Babies with a normal circumference of the head were found to survive better than those with a relatively big head and relatively small head. Under-height babies were found to have a higher risk as compared to babies born with normal height and over-height newborns. The results of the BCPHM are not far from that of JCPHM and CPHM but the non-significant covariates displayed relatively higher standard error. The overall results for non-significant covariates showed a relatively higher standard error after resampling. Due to a relatively higher risk to death of an infant from under 20 years old parents, the pregnancy of parents belonging in such range of age should be avoided. Also as abnormality lead to a relatively higher risk to infant mortality, clinically recommended nutrition during pregnancy would decrease abnormality of the newborn; this would decrease the infant mortality.

Analysis was limited to one event which is the death of the infant. Resampling with multiple events could improve models where an alternative event is attracting a chronic disease or clinical complication for the infant during the study time.

LIST OF ABBREVIATIONS

APGAR = Appearance, Pulse, Grimace, Activity and Respiration

CPHM = Cox Proportional Hazards Model

BCPHM = Bootstrap Cox Proportional Hazards Model

JCPHM = Jackknife Cox Proportional Hazards Model

KUTH = Kigali University Teaching Hospital

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The study was approved by the Kigali University Teaching Hospital where dataset was taken from the hospital database, with consent that the names of both the parents and the children cannot be published.

HUMAN AND ANIMAL RIGHTS

No animals/ humans were used for the studies that are the basis of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The data supporting the findings of the article is available in the School of Mathematics, Statistics and Computer Science, University of KwaZulu Natal at http://smscs.ukzn.ac.za/ Homepage.aspxL, reference number 00033 260 5610.

FUNDING

This work was funded by the University of KwaZulu Natal and Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB) programme.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

This work was supported through the DELTAS Africa Initiative. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [grant 107754/Z/15/Z- DELTAS Africa Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB) programme] and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

REFERENCES

[1] Efron B, Tibshirani RJ. An introduction to the bootstrap. Chapman & Hall/CRC 1994.

[2] Efron B. Jackknife-After-Bootstrap standard errors and influence functions. Wiley for the Royal Statistical Society 1992; 54(1): 83-127.

[3] Hamada C. Bootstrap Cox regression using SAS procedures. SAS Institute Japan Ltd 1995; 211: 1258-62.

[4] Utzet F, Sánchez A. Some applications of the bootstrap to survival analysis. Anu Psicol 1992; 55: 155-67.

[5] Belašková S, Fišerová E, Krupicková S. Study of bootstrap estimates in Cox regression model with delayed entry. Mathematica 2013; 52(2): 21-30.

[6] Xu G, Sen B, Ying Z. Bootstrapping a change-point Cox model for survival data. Electron J Stat 2014; 8(1): 1345-79. [http://dx.doi.org/10.1214/14-EJS927] [PMID: 25400719]

[7] Xiao L, Yao-Hua W, Dong-Sheng T. Jackknifed random weighting for Cox pro-portional hazards model. Sci China Math 2012; 55(4): 775-86. [http://dx.doi.org/10.1007/s11425-012-4380-4]

[8] Wang Z, Wu Y, Zhao LC. Approximation by randomly weighting method in censored regression model. Sci China Ser A 2009; 52: 567-76.

[http://dx.doi.org/10.1007/s11425-008-0116-x]
[9] Zheng ZG. Random weighting method. Yingyong Shuxue Xuebao 1987; 10: 247-53.
[10] Zheng ZG, Tu D. Random weighting method in regression models. Sci Sinica Series A 1988; 31: 1442-59.
[11] James LF. A study of a class of weighted bootstraps for censored data. Ann Stat 1997; 25: 1595-621.
[http://dx.doi.org/10.1214/aos/1031594733]
[12] Quan H, Tsai WY. Jackknife for the proportional hazards model. J Stat Comput Simul 1992; 43: 163-76.
[http://dx.doi.org/10.1080/00949659208811436]
[13] Sauerbrei W, Schumacher M. A bootstrap resampling procedure for model building: application to the Cox regression model. Stat Med 1992; 11(16): 2093-109.
[http://dx.doi.org/10.1002/sim.4780111607] [PMID: 1293671]
[14] Akritas MG. Bootstrapping the Kaplan-Meier estimator. J Am Stat Assoc 1986; 81: 1032-8.
[15] Efron B. Censored data and the bootstrap. J Am Stat Assoc 1981; 76: 312-9.
[http://dx.doi.org/10.1080/01621459.1981.10477650]
[16] Hjort N. Bootstrapping Cox's regression model. Technical Report, Stanford University, Dept. of Statistics 1985; p. 241.
[17] Kim J. Conditional bootstrap methods for censored data. PhD dissertation, Florida State University, Dept, of Statistics 1990.
[18] Collet D. Modelling survival data in medical research. 2nd ed. London: Chapman & Hall 2003.
[19] Breslow N. Covariance analysis of censored survival data. Biometrics 1974; 30(1): 89-99.
[http://dx.doi.org/10.2307/2529620] [PMID: 4813387]
[20] Efron B. The efficiency of Cox's likelihood function for censored data. J Am Stat Assoc 1977; 72: 557-65.
[http://dx.doi.org/10.1080/01621459.1977.10480613]
[21] Cox DR. Regression models and life-tables (with discussion). J R Stat Soc B 1972; 34(2): 187-220.
[http://dx.doi.org/10.1111/j.2517-6161.1972.tb00899.x]
[22] Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996; 49(12): 1373-9.
[http://dx.doi.org/10.1016/S0895-4356(96)00236-3] [PMID: 8970487]
[23] Parzen M, Lipsitz SR. A global goodness-of-fit statistic for Cox regression models. Biometrics 1999; 55(2): 580-4.
[http://dx.doi.org/10.1111/j.0006-341X.1999.00580.x] [PMID: 11318217]

Paper 4

Gatabazi et al. BMC Pediatrics (2020) 20:62
https://doi.org/10.1186/s12887-020-1945-1

BMC Pediatrics

RESEARCH ARTICLE Open Access

Comparison of three classes of Marginal Risk Set Model in predicting infant mortality among newborn babies at Kigali University Teaching Hospital, Rwanda, 2016

Paul Gatabazi*, Sileshi Fanta Melesse and Shaun Ramroop

Abstract

Background: The Infant Mortality Rate (IMR) in Sub-Saharan Africa (SSA) remains the highest relatively to the rest of the world. In the past decade, the policy on reducing infant mortality in SSA was reinforced and both infant mortality and parental death decreased critically for some countries of SSA. The analysis of risk to death or attracting chronic disease may be done for helping medical practitioners and decision makers and for better preventing the infant mortality.

Methods: This study uses popular statistical methods of re-sampling and one selected model of multiple events analysis for measuring the survival outcomes for the infants born in 2016 at Kigali University Teaching Hospital (KUTH) in Rwanda, a country of SSA, amidst maternal and child's socio-economic and clinical covariates. Dataset comprises the newborns with correct information on the covariates of interest. The Bootstrap Marginal Risk Set Model (BMRSM) and Jackknife Marginal Risk Set Model (JMRSM) for the available maternal and child's socio-economic and clinical covariates were conducted and then compared to the outcome with Marginal Risk Set Model (MRSM). That was for measuring stability of the MRSM.

Results: The 2117 newborns had the correct information on all the covariates, 82 babies died along the study time, 69 stillborn babies were observed while 1966 were censored. Both BMRSM JMRSM and MRSM displayed the close results for significant covariates. The BMRSM displayed in some instance, relatively higher standard errors for non-significant covariates and this emphasized their insignificance in MRSM. The models revealed that female babies survive better than male babies. The risk is higher for babies whose parents are under 20 years old parents as compared to other parents' age groups, the risk decreases as the APGAR increases, is lower for underweight babies than babies with normal weight and overweight and is lower for babies with normal circumference of head as compared to those with relatively small head.

Conclusion: The results of JMRSM were closer to MRSM than that of BMRSM. Newborns of mothers aged less than 20 years were at relatively higher risk of dying than those who their mothers were aged 20 years and above. Being abnormal in weight and head increased the risk of infant mortality. Avoidance of teenage pregnancy and provision of clinical care including an adequate dietary intake during pregnancy would reduce the IMR in Kigali.

Keywords: Infant mortality, Survival analysis, Marginal risk set model, Re-sampling, Covariate, Rwanda

* Correspondence: Gatabazi001@gmail.com
Department of Statistics, University of Kwazulu Natal, Pietermaritzburg, Private Bag X 01, Scottsville 3209, South Africa

Background

The discrepancy in IMR and low life expectancy of the SSA versus the other parts of the world attracts several researchers. The report of the World Bank in 2011 pointed that the IMR was 75/1000 in SSA versus 11/1000 in developed countries [1]. The same report pointed that half of the ten million children who die every year is in SSA. The World Bank dataset from 1960 to 2005 suggests that low life expectancy at birth in SSA is relatively higher in Middle Africa as compared to other sub-regional disparities of SSA [2]. The World Bank records of 2017 indicated that the IMR was 51.50/1000 in SSA [3]. Central African Republic had the highest IMR of 87.60/1000, the lowest IMR were found in Mauritius (11.60/1000), the IMR in Rwanda was 28.90/1000. Several studies on factors that could lower the infant mortality have been done and recommendations were suggested but the IMR remains a problem in SSA.

The multiple events model for infant mortality at the Kigali University Teaching Hospital analysed in [4] leaves a question on whether the adopted model is stable. The main causes of instability may be the correlation of the covariates or relatively small sample size [5]. One of the ways of assessing instability in survival regression models is a use of re-sampling techniques [6]. The analysis in [4] is a none re-sampled model that used the primary dataset of the year 2016. Two observable events per subject are death and the occurrence of at least one of the common conditions that may also cause the long-term death to infants. It was found that the Marginal Risk Set Model (MRSM) also known as the Wei, Lin and Weissfeld Model (WLWM) fit the data well. The WLWM is among the multiplicative methods for analysing ordered events found in [7]. Other multiplicative models include the Andersen-Gill Model (AGM) and the Prentice, Williams and Peterson Model (PWPM) [8].

The present study uses two popular nonparametric methods of re-sampling namely *bootstrap* which is based on the random samples with replacement [9], and *jackknife* method that is based on sampling by leaving out one observation at time [9]. The size of the sample in [4] is 2117 and the record is effective in the year 2016. The long-term results could be assumed according to the stability potentially observed after re-sampling. Several manuscripts on re-sampling in survival analysis are limited on the re-sampled Cox proportional hazards model and on estimating standard errors of the survival and hazard functions such as in [6, 10–13] where bootstrap is involved [13–16]; in which the jackknife is implicated or [17–22] where hazard and survival functions with their respective standard errors are of interest. The present study analyses the bootstrap-based MRSM with 1000 replicates and the jackknife-based MRSM. The results are then compared to that of the MRSM.

Methods

Dataset

The time to event data of 2117 newborns at the KUTH is recorded from the 1st January to the 31st December 2016. At KUTH, all newborns are recorded in registries with all details of parents and clinical outcomes of each newborn. The information in registry provides references on card indexes that provide information on clinical behavior of babies after leaving the hospital. KUTH as a site of interest in this study is a central Hospital where most of complicated childbirths countrywide are transferred. In 2016, KUTH recorded relatively high incidence of stillborn cases (69 stillborn babies or 3.259%) and relatively high infant mortality rate (3.873%). Table 1 summarises the information on newborns at KUTH along the study time.

The study is interested on subjects with a correct information on the covariates of interests. The two events per subject are observed namely the death and the incidence of at least one chronic disease or complication such as *severe oliguria*, *severe prematurity*, *very low birth weight*, *macrosomia*, *severe respiratory distress*, *gastroparesis*, *hemolytic*, *trisomy*, *asphyxia* and *laparoschisis*. Apart from the *event* status and the *time* to event, 11 covariates are recorded and subdivided in demographic covariates which include the *age* and the place of *residence* for parents; clinical covariates for female parents that include obstetric *antecedents*, type of *childbirth* and previous *abortion*. Clinical covariates for babies include *APGAR*; *gender*, *number* of births at a time, *weight*, circumference of the *head*, and *height*. Table 2 gives a description of the variables of interest.

Statistical methods

Marginal risk set model

Assume that $h(t|\mathbf{x}_i)$ is the hazard function of the survival time T given the p fixed covariates $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$. Let $h_0(t)$ be the hazard function when $\mathbf{x}_i = (0, 0, \ldots, 0)$ for all i, then

$$h(t|\mathbf{x}_i) = h_0(t)\ \exp\left(\boldsymbol{\beta}'\mathbf{x}_i\right) \quad (1)$$

Table 1 Summary on newborns under study

Total observations	2117
Deaths during the study time	82 (3.873%)
Stillborn babies	69 (3.259%)
Total events	151 (7.132%)
Censored babies	1966 (92.867%)

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 Gender	Type of childbirth Gender of a newborn	0 = born using ventouse, 1 = born naturally, 2 = born after surgery 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 time or not	0 = censored, 1 = dead
n events	Indicator on the rank of records per subject	1 = first record, 2 = second record

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$ is a p-dimensional vector of model parameters [23]. Define an indicator function as.

$\delta_{ij}(t) = 1$ if individual i is at risk of the j^{th} event and $\delta_{ij}(t) = 0$ otherwise.

The marginal risk set model (MRSM) or the Wei Lin and Weisfeld Model (WLWM) assumes that events are unordered where each event has its own stratum and each data point appears in all strata [4, 24]. In other words, the k^{th} time interval per subject is in the k^{th} stratum, $k = 1, 2, \ldots, n$.

The hazard function for the j^{th} event for the individual i is given by

$$h(t|\mathbf{x}_i) = \delta_{i\,j}(t) h_{0\,j}(t)\ \exp\left(\boldsymbol{\beta}'_j\,\mathbf{x}_i\right) \tag{2}$$

Maximum likelihood and parameter estimation

Let]0, τ_i [be the interval of time in which the individual i is observed with n_i the number of events of the individual i along]0, τ_i [and Assume that two events cannot occur simultaneously in continuous time. The probability density function for the outcome n_i along]0, τ_i [is given by.

$$L(\boldsymbol{\Phi}) = \prod_{i=1}^{n} L_i(\phi)$$

where

$$L_i(\phi) = \prod_{j=1}^{n_i} h(t|x_i) e^{-\int_0^{\tau_i} \delta_{ij}(v) h(v|x_i) dv}. \tag{3}$$

In (3), individual i has n_i events with $n_i \geq 0$ at times $t_{i1} \leq \mathrm{t}_{i2} \leq \cdots \leq \mathrm{t}_{ini}$.

The appropriate partial likelihood functions for tied time to event data is well described in [24] and in [25] and include Breslow's, Efron's and Cox's techniques. The maximum likelihood estimates are given by a system

$$\left\{ \begin{array}{l} \dfrac{\partial\ \ln L(\Phi)}{\partial \alpha} \\ \dfrac{\partial\ \ln L(\Phi)}{\partial \beta} \end{array} \right. \tag{4}$$

where α is known as the baseline parameter vector while β is a vector of model parameters. The Newton-Raphson method is one of numerical methods used for solving system (4). The adequacy checking of the likelihood estimates is done by finding the elements $\mathfrak{I}_{\alpha\alpha}$, $\mathfrak{I}_{\alpha\beta}$, $\mathfrak{I}_{\beta\alpha}$ and $\mathfrak{I}_{\beta\beta}$ of the information matrix $\mathfrak{I}$ and assume that as $n \to \infty$, $\hat{\Phi} - \Phi \mapsto N(0, \mathfrak{I}^{-1}(\hat{\Phi}))$ [4, 26].

In MRSM, n is assumed to be the maximum number of events per subject while τ_k, $k = 1, 2, \ldots n$ are times to events per subject along the study time with range [0, T]. The study time is partitioned into $n + 1$ intervals of the form

$$0\text{–}\tau_1, 0\text{–}\tau_2, \ldots, 0\text{–}\tau_n, 0\text{–}T. \tag{5}$$

STATA 15 provides results of the MRSM by applying the Cox Proportional Hazards Model (CPHM) to the

dataset in the setup (5). The test of proportional hazards assumption is done by checking patterns of survival functions per groups of each covariate. Figure 1 presents the patterns of survival functions per groups of each covariate using Kaplan-Meier estimation. The patterns are approximately parallel for the covariates of interest. This allows a construction of the MRSM for all the covariates.

Re-sampled MRSM

The Bootstrap Marginal Risk set Model (BMRSM) is the inference of model (2) based on bootstrap samples (see Appendix). The BMRSM consists of applying model (2) to each of the B bootstrap samples $\mathbf{x}_i^{*k}$, $\forall k \in [1, B]$ of covariates $\mathbf{x}_i$, $\forall i \in [1, p]$. Bootstrap model parameter estimation in presence of tied events uses either Breslow, Efron or Cox approach. The bootstrap standard error is obtained by using Eq. (6) of the Appendix.

As for the BMRSM, the Jackknife Marginal Risk Model (JMRSM) consists of applying model (2) to each of the n jackknife samples $\mathbf{x}_i^{*k}$ of covariates $\mathbf{x}_i$, $i \in [1, p]$ with a use of Breslow, Efron or Cox approach for estimating the jackknife model parameters. The Jackknife standard error is given by Eq. (7) found in the Appendix.

Results

Using Breslow estimation [27], Table 3 presents unadjusted MRSM, BMRSM, JMRSM and corresponding adjusted models. Unadjusted and adjusted MRSM, BMRSM and JMRSM are also presented in Tables 4 and 5 for Efron [28] and Cox estimation [29].

The results of the unadjusted JMRSM are relatively close to that of the unadjusted MRSM (Table 3). The standard errors in JMRSM and MRSM are close for all covariates. The standard errors in BMRSM and MRSM are also close for covariates except for all levels of covariates *childbirth* where the standard error in BMRSM is about 4 times that of MRSM and the upper levels of covariates *weight*, *head* and *height* where the standard error in BMRSM is about 20 times that of MRSM. Significance difference in levels of covariates is found at the same covariates for both MRSM, BMRSM and JMRSM except at the upper

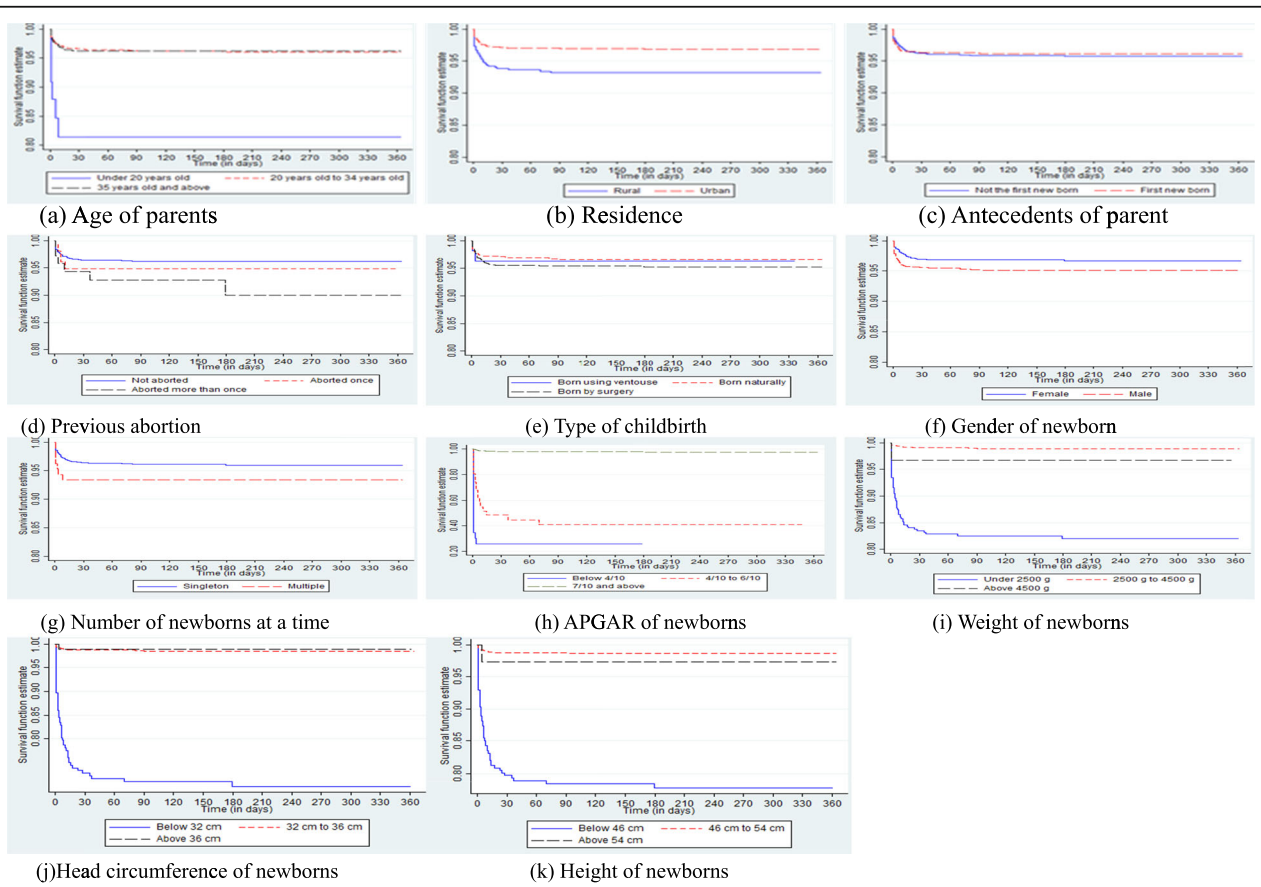


Fig. 1 Plots of the survival function per groups of covariates

Table 3 Breslow estimation

MRSM						BMRSM				JMRSM			
Covariate (reference)	Level	HR	SE	$P > z$	95% CI	HR	SE	$P > z$	95% CI	HR	SE	$P > z$	95% CI
Age (Under 20 years old)	20 to 34 years old	0.277	0.100	$p < 0.001$	[0.137; 0.560]	0.277	0.088	$p < 0.001$	[0.149; 0.515]	0.277	0.081	$p < 0.001$	[0.155; 0.493]
	35 years old and above	0.395	0.157	0.020	[0.181; 0.863]	0.395	0.132	0.005	[0.205; 0.761]	0.395	0.127	0.004	[0.210; 0.741]
Residence (Rural)	Urban	0.847	0.139	0.309	[0.614; 1.167]	0.847	0.148	0.341	[0.601; 1.193]	0.847	0.158	0.372	[0.587; 1.220]
Antecedents (Not 1st newborn)	1st newborn	0.806	0.157	0.270	[0.550; 1.182]	0.806	0.138	0.207	[0.577; 1.126]	0.806	0.134	0.193	[0.582; 1.116]
Abortion (Not aborted)	Aborted once	1.405	0.398	0.231	[0.806; 2.448]	1.405	0.459	0.298	[0.741; 2.664]	1.405	0.471	0.311	[0.728; 2.710]
	Aborted more than once	0.479	0.161	0.028	[0.248; 0.925]	0.479	0.280	0.208	[0.152; 1.507]	0.479	0.360	0.328	[0.110; 2.094]
Childbirth (Ventouse)	Natural	0.873	0.491	0.808	[0.290; 2.627]	0.873	1.973	0.952	[0.010; 73.427]	0.873	0.329	0.718	[0.416; 1.829]
	Surgery	1.115	0.613	0.843	[0.380; 3.274]	1.115	2.517	0.962	[0.013; 93.040]	1.115	0.372	0.744	[0.580; 2.143]
Gender (Female)	Male	1.740	0.296	0.001	[1.247; 2.429]	1.740	0.324	0.003	[1.209; 2.505]	1.740	0.337	0.004	[1.191; 2.544]
Number (Singleton)	Multiple	0.409	0.131	0.005	[0.218; 0.766]	0.409	0.107	0.001	[0.245; 0.682]	0.409	0.100	$p < 0.001$	[0.252; 0.661]
APGAR (Below 4/10)	4/10 to 6/10	0.377	0.112	0.001	[0.211; 0.673]	0.377	0.127	0.004	[0.195; 0.729]	0.377	0.139	0.008	[0.182; 0.778]
	7/10 and above	0.130	0.036	$p < 0.001$	[0.076; 0.222]	0.130	0.033	$p < 0.001$	[0.079; 0.212]	0.130	0.031	$p < 0.001$	[0.081; 0.208]
Weight (Under 2500 g)	2500 g to 4500 g	0.250	0.068	$p < 0.001$	[0.146; 0.427]	0.250	0.064	$p < 0.001$	[0.151; 0.412]	0.250	0.063	$p < 0.001$	[0.153; 0.408]
	Above 4500 g	0.442	0.285	0.206	[0.125; 1.565]	0.442	4.002	0.928	[0.000; 2.290×10^{7}]	0.442	0.508	0.478	[0.046; 4.222
Head (Below 32 cm)	32 cm to 36 cm	0.456	0.128	0.005	[0.263; 0.789]	0.456	0.1150	0.002	[0.277; 0.749]	0.456	0.117	0.002	[0.275; 0.753]
	Above 36 cm	0.290	0.219	0.102	[0.066; 1.278]	0.290	4.156	0.931	[0.000; 4.470×10^{11}]	0.290	0.284	0.206	[0.043; 1.971]
Height (Below 36 cm)	46 cm to 54 cm	0.894	0.276	0.716	[0.488; 1.637]	0.894	0.241	0.677	[0.527; 1.516]	0.894	0.253	0.692	[0.513; 1.557]
	Above 54 cm	1.670	1.264	0.498	[0.379; 7.361]	1.670	22.884	0.970	[0.000; 7.73×10^{11}]	1.670	1.612	0.596	[0.251; 11.093]
	Adjusted MRSM					Adjusted BMRSM				Adjusted JMRSM			
Covariate (reference)	Level	HR	SE	$P > z$	95% CI	HR	SE	$P > z$	95% CI	HR	SE	$P > z$	95% CI
Age (Under 20 years old)	20 to 34 years old	0.307	0.107	0.001	[0.155; 0.609]	0.309	0.089	$p < 0.001$	[0.176; 0.543]	0.309	0.083	$p < 0.001$	[0.182; 0.523]
	35 years old and above	0.472	0.179	0.047	[0.225; 0.992]	0.489	0.145	0.016	[0.274; 0.874]	0.489	0.137	0.011	[0.282; 0.848]
Abortion (Not aborted)	Aborted once	1.482	0.406	0.152	[0.866; 2.537]	–	–	–	–	–	–	–	–
	Aborted more than once	0.541	0.175	0.057	[0.287; 1.019]	1.607	-0.304	- 0.012	- [1.109; 2.328]	–	–	–	–
Gender (Female)	Male	1.672	0.280	0.002	[1.204; 2.321]	0.417	0.106	0.001	[0.254; 0.686]	1.607	0.316	0.016	[1.093; 2.363]
Number (Singleton)	Multiple	0.401	0.128	0.004	[0.214; 0.750]	0.412	0.137	0.008	[0.215; 0.791]	0.417	0.103	$p < 0.001$	[0.258; 0.677]

Table 3 Breslow estimation *(Continued)*

MRSM						BMRSM				JMRSM			
APGAR (Below 4/10)	4/10 to 6/10	0.414	0.119	0.002	[0.236; 0.726]	0.150	0.034	$p < 0.001$	[0.096; 0.234]	0.412	0.142	0.010	[0.210; 0.809]
	7/10 and above	0.144	0.038	$p < 0.001$	[0.086; 0.242]	0.240	0.057	$p < 0.001$	[0.151; 0.381]	0.150	0.033	$p < 0.001$	[0.098; 0.232]
Weight (Under 2500 g)	2500 g to 4500 g	0.238	0.060	$p < 0.001$	[0.144; 0.391	0.478	4.519	0.938	[0.000; 5.32×10^{7}]	0.240	0.057	$p < 0.001$	[0.151; 0.381]
	Above 4500 g	0.447	0.284	0.205	[0.129; 1.550	0.439	0.103	$p < 0.001$	[0.277; 0.696]	0.478	0.419	0.400	[0.086; 2.669]
Head (Below 32 cm)	32 cm to 36 cm	0.420	0.100	$p < 0.001$	[0.264; 0.669]	0.303	4.200	0.931	[0.000; 1.970×10^{11}]	0.439	0.107	0.001	[0.273; 0.707]
	Above 36 cm	0.284	0.210	0.089	[0.067; 1.211]					0.303	0.298	0.225	[0.044; 2.084]
		$\chi^2 = 213.161$, $p < 0.001$				$\chi^2 = 203.14$, $p < 0.001$				$\chi^2 = 22.310$, $p < 0.001$			

level of the covariate *abortion* where significance is suggested by the MRSM. Following the recommendations of Parzen and Lipsitz [30], the χ^2 test statistics suggest a higher performance of the JCPHM as compared to the CPHM and BCPHM since the χ^2 is relatively everywhere lower for the JCPHM..

Discussion

The overall results of MRSM, BMRSM and JMRSM by different approaches of ties handling (Tables 3, 4 and 5) are not critically different as expected. The STATA default method (Breslow) is then of interest in the analysis. The JMRSM is adopted for checking stability since the results are closer to that of MRSM than that of BMRSM. The similarity between MRSM and JMRSM suggests that the MRSM may be stable. The global analysis upholds the significance difference of all levels of covariates *age*, *gender*, *number* and *APGAR* and intermediate levels of covariates *weight* and *head*.

The re-sampled adjusted models by Breslow technique of handling tied events suggest that the risk of death or attracting a chronic disease of babies whose parents' age range from 20 to 34 years old is lower than that of babies whose parents are under 20 years old and that of babies whose parents are 35 years and above. Basinga et al. [31] argue that the unintended pregnancy induces abortion in Rwanda, their study suggests a relatively higher rate of teenage unintended pregnancies as compared to the other age ranges, this contributes on the first hand, to the increase of infant mortality rate. On the second hand, the study by Olausson et al. [32] confirms a relatively higher risk for teenage pregnancies due to biological immaturity. As for the advanced maternal age, Lampinen et al. [33] point that it is associated with relatively poorer outcomes to pregnancies due to the observed higher incidence of chronic medical conditions among older women.

The results show that the risk for male babies is higher than that of female babies. This complies with the usual better survival outcome of the females as reports several manuscripts such as [34] or [35]. Multiple babies survive better than singleton babies; this is however against the results from studies conducted in Sub-Saharan Africa by Monden and Smits [36] and Pongou et al. [37]. This may be due to the small number of multiple newborns recorded at KUTH along the year 2016. The survival outcomes of babies whose APGAR is below 4/10 are poorer than that of babies with higher APGAR score. Babies whose weight range from 2500 g to 4500 g survive better than those whose weight is below 2500 g and those whose weight is above 4500 g while babies whose circumference of head range from 32 cm to 36 cm survive better than those whose circumference of head is below 32 cm. The results of APGAR, weight and circumference of the head comply with the recommendations of the clinical medicine and related manuscripts such as [38] for example.

The study shows that the BMRSM is close to JMRSM and MRSM for all significant covariate but the BMRSM shows relatively higher standard errors for some non-significant covariates. The discrepancy between standard errors after re-sampling for covariates such as *childbirth*, *weight*, *head* and *height* suggests the instability of the MRSM at these specific covariates and this emphasizes their non-significance in the MRSM.

The present analysis is limited on eleven covariates. Unavailable covariates concerning parents that could improve models are, for example, demographic

Table 4 Efron estimation

MRSM						BMRSM				JMRSM			
Covariate (reference)	Level	HR	SE	$P > z$	95% CI	HR	SE	$P > z$	95% CI	HR	SE	$P > z$	95% CI
Age (Under 20 years old)	20 to 34 years old	0.230	0.083	$p < 0.001$	[0.114; 0.466]	0.230	0.086	$p < 0.001$	[0.111; 0.478]	0.230	0.083	$p < 0.001$	[0.114; 0.466]
	35 years old and above	0.324	0.129	0.005	[0.149; 0.706]	0.324	0.128	0.004	[0.149; 0.703]	0.324	0.125	0.004	[0.152; 0.691]
Residence (Rural)	Urban	0.831	0.137	0.261	[0.602; 1.147]	0.831	0.160	0.337	[0.570; 1.212]	0.831	0.174	0.376	[0.552; 1.252]
Antecedents (Not 1st newborn)	1st newborn	0.756	0.149	0.156	[0.513; 1.113]	0.756	0.149	0.155	[0.514; 1.112]	0.756	0.143	0.140	[0.521; 1.096]
Abortion (Not aborted)	Aborted once	1.393	0.396	0.244	[0.798; 2.430]	1.393	0.470	0.326	[0.719; 2.699]	1.393	0.522	0.377	[0.668; 2.904]
	Aborted more than once	0.452	0.154	0.020	[0.232; 0.880]	0.452	0.322	0.265	[0.112; 1.826]	0.452	0.391	0.359	[0.083; 2.465]
Childbirth (Ventouse)	Natural	0.736	0.408	0.580	[0.249; 2.179]	0.736	1.482	0.879	[0.014; 38.109]	0.736	0.336	0.502	[0.301; 1.801]
	Surgery	0.921	0.499	0.880	[0.319; 2.661]	0.921	1.858	0.968	[0.018; 47.963]	0.921	0.388	0.846	[0.403; 2.104]
Gender (Female)	Male	1.823	0.312	$p < 0.001$	[1.304; 2.549]	1.823	0.361	0.002	[1.238; 2.687]	1.823	0.400	0.006	[1.186; 2.804]
Number (Singleton)	Multiple	0.324	0.106	0.001	[0.170; 0.617]	0.324	0.100	$p < 0.001$	[0.177; 0.591]	0.324	0.096	$p < 0.001$	[0.181; 0.578]
APGAR (Below 4/10)	4/10 to 6/10	0.214	0.065	$p < 0.001$	[0.118; 0.387]	0.214	0.080	$p < 0.001$	[0.102; 0.447]	0.214	0.093	$p < 0.001$	[0.091; 0.501]
	7/10 and above	0.070	0.020	$p < 0.001$	[0.041; 0.121]	0.070	0.019	$p < 0.001$	[0.041; 0.120	0.070	0.019	$p < 0.001$	[0.041; 0.119]
Weight (Under 2500 g)	2500 g to 4500 g	0.231	0.063	$p < 0.001$	[0.135; 0.395]	0.231	0.064	$p < 0.001$	[0.134; 0.396]	0.231	0.062	$p < 0.001$	[0.136; 0.391]
	Above 4500 g	0.412	0.269	0.174	[0.115; 1.479]	0.412	3.892	0.925	[0.000; 4.57×10^{7}]	0.412	0.485	0.451	[0.041; 4.149]
Head (Below 32 cm)	32 cm to 36 cm	0.422	0.119	0.002	[0.243; 0.734]	0.422	0.115	0.002	[0.247; 0.720]	0.422	0.118	0.002	[0.244; 0.729]
	Above 36 cm	0.246	0.187	0.065	[0.055; 1.093]	0.246	3.784	0.927	[0.000; 3.030×10^{12}]	0.246	0.251	0.169	[0.033; 1.819]
Height (Below 36 cm)	46 cm to 54 cm	0.917	0.285	0.781	[0.499; 1.687]	0.917	0.290	0.784	[0.494; 1.704]	0.917	0.294	0.788	[0.489; 1.721]
	Above 54 cm	1.692	1.283	0.488	[0.383; 7.476]	1.692	24.567	0.971	[0.000; 3.890×10^{12}]	1.692	1.700	0.601	[0.236; 12.140]
	Adjusted MRSM					Adjusted BMRSM				Adjusted JMRSM			
Covariate (reference)	Level	HR	SE	$P > z$	95% CI	HR	SE	$P > z$	95% CI	HR	SE	$P > z$	95% CI
Age (Under 20 years old)	20 to 34 years old	0.262	0.092	$p < 0.001$	[0.132; 0.522]	0.265	0.088	$p < 0.001$	[0.138; 0.509]	0.265	0.088	$p < 0.001$	[0.138; 0.508]
	35 years old and above	0.407	0.155	0.018	[0.193; 0.859]	0.421	0.151	0.016	[0.208; 0.850]	0.421	0.146	0.013	[0.213; 0.833]
Abortion (Not aborted)	Aborted once	1.487	0.408	0.149	[0.868; 2.546]	–	–	–	–	–	–	–	–
	Aborted more than once	0.520	0.170	0.046	[0.274; 0.987]	- 1.684	- 0.336	- 0.009	- [1.138; 2.490]	–	–	–	–
Gender (Female)	Male	1.764	0.297	0.001	[1.268; 2.453]	0.322	0.097	$p < 0.001$	[0.178; 0.583]	1.684	0.367	0.017	[1.098; 2.582]
Number (Singleton)	Multiple	0.308	0.101	$p < 0.001$	[0.162; 0.586]	0.246	0.093	$p < 0.001$	[0.117; 0.515]	0.322	0.101	$p < 0.001$	[0.175; 0.594]

Table 4 Efron estimation *(Continued)*

MRSM						BMRSM				JMRSM			
APGAR (Below 4/10)	4/10 to 6/10	0.249	0.073	$p < 0.001$	[0.140; 0.442]	0.085	0.021	$p < 0.001$	[0.052; 0.138]	0.246	0.100	0.001	[0.110; 0.546]
	7/10 and above	0.081	0.022	$p < 0.001$	[0.048; 0.137]	0.225	0.057	$p < 0.001$	[0.137; 0.369]	0.085	0.021	$p < 0.001$	[0.052; 0.138]
Weight (Under 2500 g)	2500 g to 4500 g	0.222	0.057	$p < 0.001$	[0.135; 0.366]	0.487	5.083	0.945	[0.000; 3.730×10^{8}]	0.225	0.056	$p < 0.001$	[0.138; 0.367]
	Above 4500 g	0.430	0.276	0.189	[0.122; 1.512]	0.403	0.105	$p < 0.001$	[0.242; 0.671]	0.487	0.453	0.440	[0.078; 3.023]
Head (Below 32 cm)	32 cm to 36 cm	0.388	0.093	$p < 0.001$	[0.243; 0.622]	0.252	3.678	0.925	[0.000; 6.680×10^{11}]	0.403	0.108	0.001	[0.238; 0.683]
	Above 36 cm	0.235	0.175	0.052	[0.054; 1.014]					0.252	0.259	0.180	[0.034; 1.889]
		$\chi^2 = 203.061$, $p < 0.001$				$\chi^2 = 172.14$, $p < 0.001$				$\chi^2 = 21.514$, $p < 0.001$			

covariates such as the parent's education level, employment and income; behavioral covariates namely smoking habit, alcohol consumption and dietary and physiotherapeutic variables such as sports activity level. These variables are not recorded in registry at KUTH.

Conclusion

Marginal Risk Set Model (MRSM) and related re-sampling using Bootstrap (BMRSM) and Jackknife (JMRSM) are described and compared with a use of the dataset on infant mortality. The JMRSM and MRSM displayed relatively close results. The risk 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 4/10 and those whose APGAR range between 4/10 and 6/10. The risk is lower for underweight babies as compared to babies with normal weight and overweight. The survival outcomes for babies with normal circumference of head were found to be better than those with relatively small head. The study suggests that pregnancy of under 20 years old parents should be avoided, also appropriate clinical ways of keeping pregnancy against any cause of infant abnormality could help in lowering infant mortality.

Appendix

Bootstrap and Jackknife re-sampling methods

Bootstrap

Consider the p fixed covariates $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in})$ in Eq. (2) where $x_{i\ j,i\in[1,p]}$ are independent and identically distributed possibly with distribution F_θ where θ is the statistical parameter of interest. Consider the distribution function F_{Rn} of a random variable $R_n(\mathbf{x}, F_\theta)$. A bootstrap method as described in [9], consists of generating samples.

$\mathbf{x}_i^* = \mathbf{x}_i^{*1}, \mathbf{x}_i^{*2}, \ldots, \mathbf{x}_i^{*B}$,

where $\mathbf{x}_i^{*k}$, k ∈ [1, B] are random samples of size n drawn with replacement from the sample $\mathbf{x}_i$.

The variables of $\mathbf{x}_i^{*k}$ are independent and identically distributed with distribution $\hat{F}_{\theta,n}$, given $\mathbf{x}$; $\hat{F}_{\theta,n}$ is an estimator of F_θ from $\mathbf{x}_i$; B is a number of bootstrap samples also known as replications.

The estimated standard error of the bootstrap statistic of interest is given in Efron and Tibshirani [9] as

$$\hat{se}_B = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left[\hat{\theta}^*(b) - \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^*(b)\right]^2} \tag{6}$$

where $\hat{\theta}^*(b)$ is an estimate of the statistic of interest from the b^{th} bootstrap sample,

$b = 1, 2, \ldots B$

Jackknife

Consider the p fixed covariates $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in})$ in Eq. (2).

Let θ be a statistic of interest. The jackknife samples consist of leaving out one observation at a time, that is n samples.

$\mathbf{x}_i^* = (x_{i1}, x_{i2}, \ldots, x_{i\ k-1}, x_{i\ k+1}, \ldots, x_{in})\ \forall\ k \in [1, n]$ [9].

The jackknife standard error estimate as propose [9], is

$$\hat{se}_{jack} = \sqrt{\frac{n-1}{n}\sum_{i=1}^{n}\left[\hat{\theta}^*(i) - \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}^*(i)\right]^2} \tag{7}$$

where $\theta^*(i)$, $i \in [1, n]$ is a statistic of interest for the i^{th} jackknife sample.

Table 5 Cox estimation

MRSM						BMRSM				JMRSM			
Covariate (reference)	Level	HR	SE	$P > z$	95% CI	HR	SE	$P > z$	95% CI	HR	SE	$P > z$	95% CI
Age (Under 20 years old)	20 to 34 years old	0.193	0.085	$p < 0.001$	[0.081; 0.458]	0.193	0.094	0.001	[0.074; 0.502]	0.193	0.088	$p < 0.001$	[0.079; 0.472]
	35 years old and above	0.267	0.128 $p < 0.001$	0.006	[0.104; 0.682]	0.267	0.131	0.007	[0.102; 0.697]	0.267	0.124	0.004	[0.107; 0.662]
Residence (Rural)	Urban	0.766	0.150	0.175	[0.521; 1.126]	0.766	0.221	0.356	[0.435; 1.349]	0.766	0.221	0.356	[0.435; 1.350]
Antecedents (Not 1st newborn)	1st newborn	0.763	0.185	0.264	[0.475; 1.226]	0.763	0.219	0.345	[0.435; 1.338]	0.763	0.194	0.289	[0.463; 1.258]
Abortion (Not aborted)	Aborted once	1.404	0.453	0.293	[0.746; 2.643]	1.404	0.627	0.448	[0.585; 3.369]	1.404	0.593	0.422	[0.613; 3.215]
	Aborted more than once	0.378	0.152	0.015	[0.172; 0.830]	0.378	0.336	0.274	[0.066; 2.155]	0.378	0.446	0.409	[0.038; 3.814]
Childbirth (Ventouse)	Natural	0.732	0.481	0.635	[0.202; 2.653]	0.732	0.369	0.537	[0.273; 1.968]	0.732	0.365	0.532	[0.276; 1.945]
	Surgery	1.016	0.654	0.980	[0.288; 3.590]	1.016	0.480	0.973	[0.403; 2.565]	1.016	0.455	0.971	[0.423; 2.443]
Gender (Female)	Male	1.991	0.405	0.001	[1.336; 2.966]	1.991	0.534	0.010	[1.177; 3.368]	1.991	0.601	0.023	[1.101; 3.599]
Number (Singleton)	Multiple	0.218	0.111	0.003	[0.080; 0.589]	0.218	0.155	0.033	[0.054; 0.882]	0.218	0.131	0.011	[0.067; 0.709]
APGAR (Below 4/10)	4/10 to 6/10	0.080	0.042	$p < 0.001$	[0.029; 0.224]	0.080	0.056	$p < 0.001$	[0.020; 0.319]	0.080	0.052	$p < 0.001$	[0.022; 0.287]
	7/10 and above	0.021	0.011	$p < 0.001$	[0.008; 0.056]	0.021	0.014	$p < 0.001$	[0.006; 0.076]	0.021	0.011	$p < 0.001$	[0.008; 0.061]
Weight (Under 2500 g)	2500 g to 4500 g	0.236	0.070	$p < 0.001$	[0.131; 0.423]	0.236	0.077	$p < 0.001$	[0.124; 0.448]	0.236	0.068	$p < 0.001$	[0.134; 0.415]
	Above 4500 g	0.378	0.257	0.153	[0.100; 1.436]	0.378	4.696	0.938	[0.000; 1.410×10^{10}]	0.378	0.473	0.437	[0.033; 4.386]
Head (Below 32 cm)	32 cm to 36 cm	0.391	0.119	0.002	[0.216; 0.708]	0.391	0.101	$p < 0.001$	[0.236; 0.649]	0.391	0.115	0.001	[0.219; 0.698]
	Above 36 cm	0.212	0.171	0.055	[0.043; 1.033]	0.212	3.376	0.922	[0.000; 7.780×10^{12}]	0.212	0.238	0.167	[0.023; 1.913]
Height (Below 36 cm)	46 cm to 54 cm	0.828	0.283	0.582	[0.423; 1.620]	0.828	0.254	0.539	[0.454; 1.512]	0.828	0.284	0.582	[0.423; 1.622]
	Above 54 cm	1.706	1.351	0.500	[0.361; 8.060]	1.706	28.569	0.975	[0.000; 3.090×10^{14}]	1.706	1.747	0.602	[0.229; 12.707]
	Adjusted MRSM					Adjusted BMRSM				Adjusted JMRSM			
Covariate (reference)	Level	HR	SE	$P > z$	95% CI	HR	SE	$P > z$	95% CI	HR	SE	$P > z$	95% CI
Age (Under 20 years old)	20 to 34 years old	0.218	0.094	$p < 0.001$	[0.094; 0.509]	0.219	0.078	$p < 0.001$	[0.109; 0.439]	0.219	0.087	$p < 0.001$	[0.101; 0.476]
	35 years old and above	0.341	0.157	0.019	[0.138; 0.841]	0.352	0.133	0.006	[0.167; 0.738]	0.352	0.141	0.009	[0.160; 0.771]
Abortion (Not aborted)	Aborted once	1.479	0.459	0.208	[0.804; 2.719]	–	–	–	–	–	–	–	–
	Aborted more than once	0.424	0.161	0.024	[0.201; 0.892]	- 1.833	- 0.544	- 0.041	- [1.025; 3.278]	–	–	–	–
Gender (Female)	Male	1.886	0.374	0.001	[1.278; 2.783]	0.227	0.136	0.013	[0.070; 0.732]	1.833	0.528	0.036	[1.042; 3.225]
Number (Singleton)	Multiple	0.214	0.108	0.002	[0.079; 0.576]	0.091	0.053	$p < 0.001$	[0.029; 0.286]	0.227	0.135	0.013	[0.070; 0.730]

Table 5 Cox estimation *(Continued)*

MRSM						BMRSM				JMRSM			
APGAR (Below 4/10)	4/10 to 6/10	0.098	0.050	$p < 0.001$	[0.036; 0.267]	0.026	0.013	$p < 0.001$	[0.010; 0.067]	0.091	0.062	$p < 0.001$	[0.024; 0.345]
	7/10 and above	0.026	0.012	$p < 0.001$	[0.010; 0.066]	0.215	0.060	$p < 0.001$	[0.125; 0.371]	0.026	0.013	$p < 0.001$	[0.010; 0.069]
Weight (Under 2500 g)	2500 g to 4500 g	0.213	0.057	$p < 0.001$	[0.125; 0.361]	0.398	4.183	0.930	[0.000; 3.590×10^{8}]	0.215	0.057	$p < 0.001$	[0.128; 0.362]
	Above 4500 g	0.364	0.245	0.134	[0.097; 1.364]	0.374	0.102	$p < 0.001$	[0.219; 0.640]	0.398	0.385	0.340	[0.060; 2.650]
Head (Below 32 cm)	32 cm to 36 cm	0.349	0.090	$p < 0.001$	[0.211; 0.579]	0.222	3.684	0.928	[0.000; 7.970×10^{13}]	0.374	0.105	$p < 0.001$	[0.216; 0.648]
	Above 36 cm	0.199	0.160	0.044	[0.042; 0.957]					0.222	0.253	0.186	[0.024; 2.067]
		$\chi^2 = 200.400$, $p < 0.001$				$\chi^2 = 190.114$, $p < 0.001$				$\chi^2 = 23.710$, $p < 0.001$			

Abbreviations

AGM: Andersen-Gill Model; APGAR: Appearance, Pulse, Grimace, Activity and Respiration; BCPHM: Bootstrap Cox Proportional Hazards Model; BMRSM: Bootstrap Marginal Risk Set Model; CPHM: Cox Proportional Hazards Model; CPHM: Jackknife Cox Proportional Hazards Model; IMR: Infant Mortality Rate; JMRSM: Jackknife Marginal Risk Set Model; KUTH: Kigali University Teaching Hospital; MRSM: Marginal Risk Set Model; PWPM: Prentice, Williams and Peterson Model; SSA: Sub-Saharan Africa; WLWM: Wei, Lin and Weissfeld Model

Acknowledgements

We thank DELTAS Africa Initiative and the University of KwaZulu-Natal for jointly supporting this work.

Authors' contributions

SF-M participated in the study design and data acquisition and reviewed the study, S-R participated in the study conceptualization and reviewed the study, P-G collected the dataset, analysed the dataset and reported the text. All authors reviewed and approved the final manuscript.

Funding

Writing the manuscript is founded by the University of KwaZulu-Natal. The study design, data collection, data analysis and interpretation are funded by the Wellcome Trust, grant 107754/Z/15/Z-DELTAS Africa via Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB).

Availability of data and materials

The dataset used is confidential. Some information on it is available from authors on reasonable request.

Ethics approval and consent to participate

The present study was approved by the Kigali University Teaching Hospital where dataset was taken from the hospital database, with consent that the names of both parents and children cannot be published.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 7 January 2019 Accepted: 24 January 2020
Published online: 10 February 2020

References

1. Ester PV, Torres A, Freire JM, Hernández V, Gil A. Factors associated to infant mortality in Sub-Saharan Africa. J Public Health Afr. 2011;2(e27):112–6.
2. Adetunji J, Bos ER. Disease and Mortality in Sub-Saharan Africa. The International Bank for Reconstruction and Development/ the World Bank. 2nd ed; 2006. URL http://www.ncbi.nlm.nih.gov/books/NBK2292/
3. Gatabazi P, Melesse SF, Ramroop S. Infant mortality at the Kigali University Teaching Hospital: Application of Aalen additive hazards model and comparison with other classical survival models. Afr Popul Stud. 2019;33(2): 4834–51.
4. Gatabazi P, Melesse SF, Ramroop S. Multiple events model for the infant at Kigali University teaching hospital. Open Public Health J. 2018;11:464–73.
5. Hamada C. Bootstrap Cox regression using SAS procedures. SAS Institute Japan Ltd. 1995;211:1258–62.
6. Utzet F, Sánchez A. Some applications of the bootstrap to survival analysis. Anuario de Psicologia. 1992;55:155–67.
7. Cook RJ, Lawless JF. The statistical analysis of recurrent events. New York: Springer; 2007.
8. Wei LJ, Glidden DV. An overview of statistical methods for multiple failure time data in clinical trials. Stat Med. 1997;16:833–9.
9. B Efron and R J Tibshirani. An introduction to the bootstrap. Chapman & Hall/CRC, 1994.
10. Bělašková SB, Fišerová E, Krupičková S. Study of bootstrap estimates in Cox regression model with delayed entry. Mathematica. 2013;52(2):21–30.
11. Xu G, Sen B, Ying Z. Bootstrapping a change-point Cox model for survival data. Electron J Stat. 2014;8(1):1345–79.
12. Xiao L, Yao-Hua W, Dong-Sheng T. Jackknifed random weighting for Cox proportional hazards model. Sci China Math. 2012;55(4):775–86.
13. Gatabazi P, Melesse SF, Ramroop S. Re-sampled Cox proportional hazard model for the infant mortality at the Kigali University teaching hospital teaching hospital. Open Public Health J. 2019;12:136–44.
14. Wang Z, Wu Y, Zhao LC. Approximation by randomly weighting method in censored regression model. Sci China Ser A. 2009;52:567–76.
15. Zheng ZG. Random weighting method. Acta Math Appl Sin. 1987;10: 247–53.
16. Zheng ZG, Tu D. Random weighting method in regression models. Sci Sinica Series A. 1988;31:1442–59.
17. James LF. A study of a class of weighted bootstraps for censored data. Ann Stat. 1997;25:1595–621.
18. Quan H, Tsai WY. Jackknife for the proportional hazards model. J Stat Comput Simul. 1992;43:163–76.
19. Sauerbrei W, Schumacher M. A bootstrap resampling procedure for model building: application to the Cox regression model. Stat Med. 1992;11:2093–109.
20. Efron B. Censored data and the bootstrap. J Am Stat Assoc. 1981;76:312–9.
21. Hjort N. Bootstrapping Cox's regression model. Technical Report, Stanford University, Dept. of Statistics, vol. 241; 1985.
22. Kim J. Conditional bootstrap methods for censored data. PhD dissertation, Florida State University, Dept, of Statistics; 1990.
23. Collet D. Modeling survival data in medical research. 2nd ed. London: Chapman & Hall; 2003.
24. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. J Am Stat Assoc. 1989; 84(408):1065–73.
25. Gatabazi P, Kabera G. Survival analysis and its stochastic process approach with application to diabetes data; 2015.

26. Sankaran PG, Anisha P. Shared frailty model for recurrent event data with multiple causes. J Appl Stat. 2011;38(12):2859–68.
27. Breslow NE. Covariance analysis of censored survival data. Biometrics. 1974; 30:89–99.
28. Efron B. The efficiency of Cox's likelihood function for censored data. J Am Stat Assoc. 1977;72:557–65.
29. Cox DR. Regression models and life-tables (with discussion). J R Stat Soc. 1972;34(2):187–220 Series B.
30. Parzen M, Lipsitz SR. A global goodness-of-fit statistic for Cox regression models. Biometrics. 1999;55:580–4.
31. Basinga P, Moore AM, Singh S, Remez L, Birungi F, Nyirazinyoye L. Unintended pregnancy and induced abortion in Rwanda: causes and consequences. New York: Guttmacher Institute; 2012.
32. Olausson PO, Cnattinius S, Haglund B. Teenage pregnancies and risk of late fetal death and infant mortality. Br J Obstet Gynaecol. 1999;106:116–21.
33. Lampinen R, Julkunen KV, Kankkunen P. A review of pregnancy in women over 35 years of age. Open Nurs J. 2009;3:33–8.
34. Pongou R. Why is infant mortality higher in boys than in girls? A new hypothesis based on preconception environment and evidence from a large sample of twins. Demography. 2013;50:421–44.
35. Zarulli V, Jones JAB, Oksuzyanc A, Lindahl-Jacobsen R, Christensen K, Vaupel JW. Women live longer than men even during severe famines and epidemics. Proc Natl Acad Sci U S A. 2018;115(4):E832–40.
36. Monden CWS, Smits J. Mortality among twins and singletons in Sub-Saharan Africa between 1995 and 2014: a pooled analysis of data from 90 demographic and health surveys in 30 countries. Proc Natl Acad Sci U S A. 2017;5(7):E673–9.
37. Pongou R, Shapiro D, Tenikue M. Mortality convergence of twins and singletons in sub-Saharan Africa. Demogr Res. 2019;41(36):1047–58.
38. Janssen PA, Thiessen P, Klein MC, Whitfield MF, Ying C, Cullis-Kuhl SC. Standards for the measurement of birth weight, length and head circumference at term in neonates of European, Chinese and South Asian ancestry. Open Med. 2007;1(2):74–88.

Publisher's Note

Appendix C

STATA code

```
/**Generating age groups**/

gen agegroup=0 if age<20

replace agegroup=1 if age>=20 & age<35

replace agegroup=2 if age>=35

label var agegroup "Age groups"

label values agegroup age

label define age 0 "<20" 1 "[20,35)" 2 "35+"

/** Setting data as time to event survival data **/

stset time, failure(event)

/**Life table estimate of the survival function**/

ltable time event, survival

/**Graph of the life table survival function with 95% confidence limits**/

ltable time event, graph notable survival ci plotopts(recast(line) ///

lcolor(blue)) ciopts(recast(rline) lcolor(black) lpattern (dash)) ///
```

ytitle(Survival function estimate) xtitle(Time (in days))

/**Kaplan-Meier estimate of the survival function**/

sts list

/**Graph of Kaplan-Meier estimate of the survival function with 95% confidence limits**/

sts graph, ci plotopts(recast(line) lcolor(blue)) ciopts(recast(rline) lcolor(black) ///
lpattern(dash)) ytitle(Survival function estimate) ylabel(6) xtitle(Time (in days))

/**Log-rank test for comparison of survival function for variables age, residence, antecedents, abortion, childbirth, gender, number, APGAR, weight, head and height, respectively**/

sts test age, logrank

sts test residence, logrank

sts test antecedents, logrank

sts test abortion, logrank

sts test childbirth, logrank

sts test gender, logrank

sts test number, logrank

sts test apgar, logrank

sts test weight, logrank

sts test head, logrank

sts test height, logrank /**Wilcoxon test for for comparison of survival function for variables age, residence, antecedents, abortion, childbirth, gender, number, APGAR, weight, head and height, respectively**/

sts test age, wilcoxon

sts test residence, wilcoxon

sts test antecedents, wilcoxon

sts test abortion, wilcoxon

sts test childbirth, wilcoxon

sts test gender, wilcoxon

sts test number, wilcoxon

sts test apgar, wilcoxon

sts test weight, wilcoxon

sts test head, wilcoxon

sts test height, wilcoxon

/**Cox proportional hazards model for original dataset**/

xi: stcox i.age i.residence i.antecedents i.abortion i.childbirth i.gender i.number i.apgar i.weight i.head i.height

/**Stepwise Cox proportional hazard model**/

xi: stepwise stcox i.age i.residence i.antecedents i.abortion i.childbirth i.gender i.number i.apgar i.weight i.head i.height, pr(0.05)

/**Marginal risk set mode**/

stset tstop, fail(event) exit(time.) enter(tstart)///
stcox i.age i.residence i.antecedents i.abortion i.childbirth i.gender i.number i.apgar i.weight i.head i.height,///
strata(rec)

/**Aalen additive hazards regression models with 4 tests based on weights**/

xi: stlh i.age i.residence i.antecedents i.abortion i.childbirth i.gender i.number i.apgar i.weight i.head i.height, ///

xlabel(0,500,1000,1500,2000) l1title("Cumulative parameter function") testwt(1 2 3 4) ///
b2title("Time (in days)")

/**Bootstrap Cox Proportional Hazards model**/

/**Breslow estimation**/

bootstrap, reps(1000): stcox i.age i.residence i.antecedents i.abortion i.childbirth i.gender i.number i.apgar i.weight i.head i.height

/**Efron estimation**/

bootstrap, reps(1000): stcox i.age i.residence i.antecedents i.abortion i.childbirth i.gender i.number i.apgar i.weight i.head i.height, efron

/**Cox estimation**/

bootstrap, reps(1000): stcox i.age i.residence i.antecedents i.abortion i.childbirth i.gender i.number i.apgar i.weight i.head i.height, exactp

/**Jackknife Cox Proportional Hazards model**/

/**Breslow estimation**/

jackknife: stcox i.age i.residence i.antecedents i.abortion i.childbirth i.gender i.number i.apgar i.weight i.head i.height

/**Efron estimation**/

jackknife: stcox i.age i.residence i.antecedents i.abortion i.childbirth i.gender i.number i.apgar i.weight i.head i.height, efron

/**Cox estimation**/

jackknife: stcox i.age i.residence i.antecedents i.abortion i.childbirth

i.gender i.number i.apgar i.weight i.head i.height, exactp

/**Bootstrap Marginal risk set mode**/

stset tstop, fail(event) exit(time.) enter(tstart)///

bootstrap, reps(1000): stcox i.age i.residence i.antecedents i.abortion i.childbirth i.gender i.number i.apgar i.weight i.head i.height,///

strata(rec)

/**Jackknife Marginal risk set mode**/

stset tstop, fail(event) exit(time.) enter(tstart)///

jackknife: stcox i.age i.residence i.antecedents i.abortion i.childbirth i.gender i.number i.apgar i.weight i.head i.height,///

strata(rec)