# Application of Random Survival Forests and Accelerated Failure Time Shared Frailty Models in Understanding Under-Five Child Mortality in Kenya

**Khaoya Moses Mutaki**

**Submitted to the School of Mathematics in partial fulfillment for a degree of Master of Science in Social Statistics**

**August, 2018**

University of Nairobi

School of Mathematics

ISSN: 2410-1397

Master Project in Mathematics

# Application of Random Survival Forests and Accelerated Failure Time Shared Frailty Models in Understanding Under-Five Child Mortality in Kenya

**Research Report in Mathematics, Number 07, 2018**

Khaoya Moses Mutaki

August 2018

Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Social Statistics

# Application of Random Survival Forests and Accelerated Failure Time Shared Frailty Models in Understanding Under-Five Child Mortality in Kenya

**Research Report in Mathematics, Number 07, 2018**

Khaoya Moses Mutaki

School of Mathematics
College of Biological and Physical sciences
Chiromo, off Riverside Drive
30197-00100 Nairobi, Kenya

Master of Science Project
Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Social Statistics

Prepared for The Director
Graduate School
University of Nairobi

Monitored by School of Mathematics

# Abstract

**Background:** Under-five mortality rates is one of the health indicators of great importance for any country. Kenya is among those nations in the sub-saharan part of Africa which has high under-five deaths, and thus it will be of importance to apply best statistical approaches to establish which factors have influence on child mortality, this will assist to plan for the interventions.

**Approach:** Our study employed use of Random Forest for Survival Regression and Classification to analyze the Kenya Demographic Health Survey (KDHS) 2014 data to do selection of the risks factors for the under-five mortality. Akaike Information Criterion (AIC) statistics was employed to select most appropriate accelerator failure time (AFT)-shared frailty model.

**Results:** The results gotten through fitting the AFT-shared frailty model was that there was presence of unmeasured factors at community cluster while at household cluster there was no evidence suggesting existence of the unmeasured factors. Log-logistic AFT-model showed that the sons who have died, daughters who have died, duration of breastfeeding, and months of breastfeeding were found to be having significant influence on the under-five mortality ($p < 0.05$). Log-logistic AFT model with Gaussian frailty was the most appropriate model for under-five child mortality due it's least Akaike Information Criterion (AIC) statistic.

**Conclusion:** Our study found out that there was presence of unobserved heterogeneity at community clusters, this means that there are other influences that do affect mortality at community clusters which the variables alone in the model cannot explain. On the other hand there was no presence of the unobserved heterogeneity at household clusters, implying that factors influencing under-five deaths in the households can be clarified just by using the covariates in the model without the inclusion of household cluster term.

**Dissertations in Mathematics at the University of Nairobi, Kenya.**
**ISSN 2410-1397: Research Report in Mathematics**

**DISTRIBUTOR: School of Mathematics, University of Nairobi, Kenya**

# Declaration and Approval

I the undersigned declare that this project report is my original work and to the best of my knowledge, it has not been submitted in support of an award of a degree in any other university or institution of learning.

Signature Date

KHAOYA MOSES MUTAKI
Reg No. I56/87323/2016

In my capacity as a supervisor of the candidate, I certify that this report has my approval for submission.

Signature Date

Dr Nelson Owuor Onyango
School of Mathematics,
University of Nairobi,
Box 30197, 00100 Nairobi, Kenya.
E-mail: onyango@uonbi.ac.ke

Signature Date

Dr Rachel Sarguta
School of Mathematics
University of Nairobi,
Box 30197, 00100 Nairobi, Kenya.
E-mail: rsarguta@uonbi.ac.ke

# Dedication

This project is dedicated to my parents Mr. and Mrs. Khaoya, and to my sibling Collins. Most importantly I dedicate this project to the Almighty God for always guiding and strengthening me.

# Contents

# Acknowledgments

Firstly, I give thanks to Almighty God for his protection and strength to complete this research project.
I'd love to acknowledge and thank my supervisors Dr. Nelson Onyango Owuor, and Dr. Rachel Sarguta for the guidance they provided me throughout the preparation of this research project, they have been instrumental.
Lastly, I also would like to acknowledge my SST discussion group whom have been instrumental in making me understand more other concepts in survival analysis. ....

Khaoya Moses Mutaki

Nairobi, 2018.

# 1 Introduction

## 1.1 Survival Analysis

This is a section of study which analyzes period since a definite time of origin up to end of study or an event of interest occurs, this event could be, disease, death, divorce, marriage, etc., in the medical field, the study is typically a cohort follow up study or a scientific trial testing may be the effects of a new drug, but can also be extended to other areas like engineering where it can be referred to as **reliability/ failure time/ duration analysis**, which actually studies the period until a mechanical systems fail, and in the field of sociology it is termed as event history analysis, this is simply analysis which involves rates of how the events take place in a given period e.g period until an individual gets a divorce. The data for survival analysis has two distinct features that other datasets don't have;

- It has **survival times** (period until event occurs, which are usually positive skewed).
- It has **censored** observations. At times, the survival time of subjects involved in the study is never observed during the whole duration of the study, due to; relocating, dying from an unrelated cause, dropping from the study or the study ends before they die/get to experience the event. This means it is impossible to observe their survival times, these category of subjects are typically censored since we have no information about their survival time.

Censoring occurs when an occurrence which is of interest occurs beforehand or afterward the actual observation time.
Censoring can be categorized into three categories, these have been mentioned and explained briefly below :

Right censoring;
There are two categories of right censoring namely:

- Type I right censoring

This category of censoring occurs when the occurrence is seen only when it has occurred especially at the end of the study.

- Type II right censoring

This happens in a case where the study is continued only until $r$-individuals out of a total of $n$-individuals have observed the event of interest.
$r =$ actual event time
$n - r =$ censored time (fixed)

Left censoring
This is when the subject either dies or drops out during the study but the exact time they dropped out is not known.

Interval censoring
Is where the event is observed between two observation periods but we are not certain of the exact time the subject experienced the event.

Assumption made is survival time is independent off censoring.
Major functions that help in summarizing survival data are three in number, and are mentioned and explained below:

Survival function ($S(x)$)
This actually refers to the basic quantile employed for the purpose of defining period to event data.
This would be defined as the chances of an individual living past time $x$.

$$S(x) = Pr[X \geq x]$$

Let us take $X$ which is a variable that is random of continuous form;

$$S(x) = 1 - Pr[X \leq x] = 1 - F(x)$$

Thus

$$S(x) = 1 - \int_{-\infty}^{x} f(u)du$$

$$\frac{dS(x)}{dx} = -\frac{d}{dx}\int_{-\infty}^{x} f(u)du$$

$$-\frac{dS(x)}{dx} = f(x)$$

Two relationships are confirmed

$$S(x) = 1 - F(x)$$

$$f(x) = -\frac{dS(x)}{dx}$$

Hazard function ($h(x)$)
This would be defined as the chances of an incident taking place at a given time $x$ given that the subject has lived upto or beyond period $x$.

$$h(x) = \lim_{\Delta x \to 0} Pr[x \leq X \leq x + \Delta x | T \geq x]$$

With $X$ being a variable that is random and continuous, ignore $\Delta x$ since its assumed to be close to zero, then;

$$h(x) \approx Pr[X = x | X \geq x]$$

$$\frac{Pr[X = x, X \geq x]}{Pr[X \geq x]}$$

$$\frac{Pr[X = x]}{Pr[X \geq x]}$$

$$\frac{f(x)}{S(x)}$$

Since

$$f(x) = -\frac{dS(x)}{dx}$$

Then

$$h(x) = -\frac{dS(x)}{dx} / S(x)$$

$$= -\frac{d}{dx} ln(S(x))$$

Cumulative hazard function
It is gotten via integrating the hazard function, and would be defined as the chances of an incident taking place at time $x$ given survival until time $x$. It's expression can be written as per below;

$$H(x) = \int_{-\infty}^{x} h(u)du$$

$$= \int_{0}^{x} -\frac{d}{du} ln(S(u))du$$

$$= -ln(S(x))$$

$$\rightarrow S(x) = exp(-H(x))$$

$$exp(-\int_{0}^{\infty} h(u)du)$$

**Mean residual life time function**(Median Life Time);
The $p^{th}$ percentile time $t_p$, this is time until the $p\%$ of the population has developed the event of interest is given by;

$$S(t_p) = p$$

Therefore the median survival time is given by;

$$S(t_{median}) = 0.5$$

### 1.1.1 Estimation of S(t)

This is a non-parametric approach which was presented by Kaplan and Meier (1958). It gives a quick and simple survival function estimate while censoring is present. It uses exact failure, it is expressed as;

$$\hat{S}(t) = \prod_{j=1}^{k} \frac{n_j - d_j}{n_j}$$

$\hat{S}(t)$ means Kaplan Meier estimate, where; $n_j$ implies number that is at risk at period $j$, and $d_j$ number of incidents (say deaths) as at period $j$.

The standard error is therefore, expressed as;

$$s.e(\hat{S}(t_j)) = \hat{S}(t_j)\sqrt{\sum_{j=1}^{k} \frac{d_j}{n_j(n_j - d_j)}}$$

Cumulative hazard function estimate will thereby be written as per below;

$$\hat{H}(t) = \sum_{j=1}^{k} \frac{d_j}{n_j}$$

### 1.1.2 Survival regression

#### Cox-proportional hazard model

It is a popularly used model in survival analysis. With this model distribution for baseline hazard function is not specified and this is the reason why it is called a semi-parametric approach.Cox-ph model is a more general model in modelling the hazard and survival function since it doesn't place distributional assumptions on the baseline hazard. The Cox-ph model was first introduced by Cox (1972). This model can be expressed as follows;

$$\lambda_i(t) = \lambda_0(t)exp(\beta_1 X_{i1} + ... + \beta_p X_{ip})$$

$$= \lambda_0(t)exp(\beta^T X_i)$$

Whereby;
$\lambda_0(t)$ is an arbitrary baseline rate.
$X_i$ is the vector of (fixed-effect) covariate.
$\beta$ is the vector of coefficients of regression.
Lets take that event(death) has been observed to have occurred with subject $i$ at a given time $t_i$.The probability that this event occurred will be expressed as:

$$L_i(\beta) = \frac{exp(\beta^T X_i)}{\sum_{j:t_j \geq t_i} exp(\beta^T X_j)}$$

The summation is over the set of subjects $j$ where the event has not occurred before time $t_i$(including subject $i$ itself). $0 < L_i(\beta) \leq 1$, this is the partial likelihood. Treating the subjects to be statistically independent of each other, joint probability of all events is represented by below partial likelihood, where occurrence of the event (death) is indicated with $C_i = 1$:

$$L(\beta) = \prod_{i:C_i=1} L_i(\beta)$$

### 1.1.3 Variable selection

#### Schoenfeld Residuals

For one to be able to do fitting of standard Cox-proportional hazard model, it is important to be aware of one of it's main assumptions. Abeysekera and Sooriyarachchi (2009)Cox (1972) The model takes that hazard of different strata formed by levels of covariates are proportional. One can apply Kaplan-Meier plots to perform test for this assumption but the limitation is that these graphical approaches may not be adequate in situations whereby violation of the proportional hazard assumption is negligible.

Kleinbaum and Klein (2002) presented the Goodness of Fit (GOF) testing technique. This method gives a test statistic and p-value for assessing the proportional hazard assumption. This test enables one to make a more reliable decision as opposed to applying graphical approach. Schoenfeld residuals are further discussed by Grambsch and Therneau (1994). Something to note about this particular test is that if proportional hazard assumption do hold for given covariate, then it will simply mean that Schoenfield residuals for that respective covariate does not relate to study time.

### 1.1.4 Frailty

Standard approaches in Survival analysis do require independent event times, given the covariate information. In practice, many studies involve clusters. Examples of clusters can include communities, geographical areas, families e.t.c. Within a cluster, data are typically dependent. This dissertation has focused on the frailty model, introduced in section 3.7, to account for the dependence in clustered data. With frailty model context, the within-cluster association is taken into account by means of a cluster-specific factor, the frailty term. Typical for this model is that the frailty term is treated as a random effect.

Random effects for survival data were first considered in Vaupel et al. (1979) for purpose of improving fit of mortality models at advanced ages. With these early publications, the frailty term acts at the individual level as an unobservable factor in the mortality model and it indicates that **frail** individuals have an increased risk of death. The distribution of the individual-specific frailty term provides a way to model unexplained/unobserved variation in susceptibility to death in a population.

Applications of frailty model in clustered data were firstly discussed in [Clayton et al., 1978] for studies of familial aggregation of disease. Due to e.g environmental and genetic factors, susceptibility to disease do vary from family to family. Therefore, variation in the outcome among relatives tends to be lower than variation in the outcome between non-relatives. In statistical terms, this translates to:

- unobserved heterogeneity between families, and
- association among observations from the same family.

Frailty models with a cluster-specific frailty term, also called shared frailty models, have been developed over the last thirty years to deal with this type of data.

**Cox-Shared frailty model development**

Cases in same cluster here are assumed to be sharing same unseen effects, and this is the reason why it is called shared frailty model. The hazard rate for $j^{th}$ subject in $i^{th}$ cluster is:

$$\lambda_{ij}(t) = \lambda_0(t)e^{(\beta^T X_{ij}+q_i)}$$

Whereby

- $\lambda_0(t)$ is an arbitrary baseline rate
- $X_i$ is fixed-effect covariate vector
- $\beta$ is coefficient of regression vector
- $q_i$ is random effect for given cluster $i$, $i = 1,...,G$

In terms of the frailties $W_1,...,W_G$, given by, $W_i = exp(q_i)$ the frailty model would be expressed as per below:

$$\lambda_{ij}(t) = \lambda_0(t)W_i exp(\beta^T X_{ij})$$

$W_i$ is the frailty term that is unobservable, and it varies through the sample, and it increases risks for subjects in cluster $i$ if $W_i > 1$ and decreases if $W_i < 1$.
This model is represented by the conditional survival function:

$$S_{ij}(t|W_i,X) = exp(-W_i \int_0^t \lambda(u|X)du) = exp(-W_i\Lambda(t|X))$$

Whereby $\Lambda(t|X) = \int_0^t \lambda(u|X)du$. $S_{ij}(t|W_i,X)$ represents the probability of subject $j$ in cluster $i$ surviving upto or beyond time $t$ given $W_i$ and given vector of observable covariates $X$.
Most of the calculations are performed using Laplace transformation. The Laplace transform of the random variable $W$ is given clearly as:

$$L(s) = \int exp(-sw)g(w)dw = E(exp(-sW))$$

Whereby g(w) represents density of W. The range of the integral depends with the distribution in use. The marginal survival function is computed as per below:

$$S(t|X) = \int S(t|W,X)g(w)dw = E(S(t|W,X)) = L(\Lambda(t|X))$$

The $W_i$ are independent and identically distributed following a chosen distribution.

### Frailty distribution

There are many distributions that can be selected for frailty but the mostly used frailty distribution is the gamma distribution. The reason behind this is that it's laplace transform is simple and it's easily obtained.

Many calculations are done by applying Laplace transform. The Laplace transform of the frailty term, say $w$ is described as:

$$L(s) = \int exp(-sw)g(w)dw = E(exp(-sW))$$

Whereby $g(w)$ represents density of $W$. The range of the integral depends with the frailty distribution in use. Distributions that can also be selected for frailty are:

- Positive stable frailty model
- Power variance function frailty
- Normal frailty model
- Inverse Gaussian frailty model.

### Gamma frailty model

Lets take that $W$ follows a gamma distribution; the pdf of a two parameter gamma distribution is expressed as;

$$g(w) = \frac{\beta^{\alpha} w^{\alpha-1} exp(-\beta w)}{\Gamma(\alpha)}, \alpha > 0, \beta > 0, and w > 0$$

Whereby $\alpha$ and $\beta$ are shape and scale parameters correspondingly.
Laplace transformation is given by;

$$L(s) = \int_0^{\infty} exp(-sw)g(w)dw$$

$$= \int_0^{\infty} exp(-sw)\frac{\beta^{\alpha} w^{\alpha-1} exp(-\beta w)}{\Gamma(\alpha)}dw$$

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} exp(-w(s+\beta))w^{\alpha-1}dw$$

let

$$y = w(s+\beta)$$

therefore

$$w = \frac{y}{(s+\beta)}$$

applying change of variable technique we get

$$dw = \frac{dy}{(s+\beta)}$$

substituting $w$ and $dw$ we get

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} exp(-y)(\frac{y}{s+\beta})^{\alpha-1} \frac{dy}{(s+\beta)}$$

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha)(s+\beta)^{\alpha}} \int_0^{\infty} exp(-y)(y)^{\alpha-1} dy$$

since

$$\Gamma(\alpha) = \int_0^{\infty} exp(-y)(y)^{\alpha-1} dy$$

Therefore,

$$L(s) = \frac{\beta^{\alpha}}{\Gamma(\alpha)(s+\beta)^{\alpha}} \times \Gamma(\alpha)$$

$$= \frac{\beta^{\alpha}}{(s+\beta)^{\alpha}}$$

$$= \beta^{\alpha}(s+\beta)^{-\alpha}$$

Mean and variance will therefore be gotten from 1st and 2nd derivatives w.r.t s of transformation.

$$L^1(s) = -\alpha\beta^{\alpha}(s+\beta)^{-\alpha-1}$$

$$L^2(s) = \alpha(\alpha+1)\beta^{\alpha}(s+\beta)^{-\alpha-2}$$

equating $s$ to 0, therefore the mean and variance from laplace becomes;

$$E(W) = (-1)L^1(0) = \frac{\alpha}{\beta}$$

$$Var(W) = L^2(0) - (-L^1(0))^2$$

$$= \frac{\alpha(\alpha+1)}{\beta^2} - (\frac{\alpha}{\beta})^2$$

$$= \frac{\alpha^2+\alpha-\alpha^2}{\beta^2}$$

$$= \frac{\alpha}{\beta^2}$$

In shared gamma frailty, for sake of simplicity we take a one parameter gamma distribution, for the case of one parameter gamma, $\alpha = \beta$, the pdf is expressed as;

$$g(w) = \frac{\alpha^{\alpha} w^{\alpha-1} exp(-\alpha w)}{\Gamma(\alpha)}$$

The mean of $W$ is 1, and the variance is $\frac{1}{\alpha}$. Using the fact that $Var(W) = \theta = \frac{1}{\alpha}$;

$$g(w) = \frac{w^{1/\theta-1} exp(-w/\theta)}{\Gamma(1/\theta)\theta^{1/\theta}}$$

For one parameter gamma, the laplace transformation will be expressed as per below

$$L(s) = \int_0^{\infty} exp(-sw) g(w) dw$$

$$= \int_0^{\infty} \frac{exp(-sw) w^{1/\theta-1} exp(-w/\theta)}{\Gamma(1/\theta)\theta^{1/\theta}} dw$$

$$= \frac{1}{\Gamma(1/\theta)\theta^{1/\theta}} \int_0^{\infty} exp(-sw) w^{1/\theta-1} exp(-w/\theta) dw$$

$$= \frac{1}{\Gamma(1/\theta)\theta^{1/\theta}} \int_0^{\infty} w^{1/\theta-1} exp(-sw - w/\theta) dw$$

$$= \frac{1}{\Gamma(1/\theta)\theta^{1/\theta}} \int_0^{\infty} exp(-w(s+1/\theta)) w^{1/\theta-1} dw$$

let

$$y = w(s+1/\theta)$$

$$w = \frac{y}{(s+1/\theta)}$$

applying change of variable technique we get;

$$dw = \frac{dy}{(s+1/\theta)}$$

$$= \frac{1}{\Gamma(1/\theta)\theta^{1/\theta}} \int_0^{\infty} exp(-y) (\frac{y}{s+1/\theta})^{1/\theta-1} \frac{dy}{(s+1/\theta)}$$

$$= \frac{1}{\Gamma(1/\theta)\theta^{1/\theta}} \int_0^{\infty} exp(-y) \frac{y^{1/\theta-1}}{(s+1/\theta)^{1/\theta}} dy$$

$$= \frac{1}{(s+1/\theta)^{1/\theta}\Gamma(1/\theta)\theta^{1/\theta}} \int_0^{\infty} exp(-y) y^{1/\theta-1} dy$$

using properties of a gamma function;

$$\Gamma(1/\theta) = \int_0^{\infty} exp(-y) y^{1/\theta-1} dy$$

$$= \frac{1}{(s+1/\theta)^{1/\theta}\Gamma(1/\theta)\theta^{1/\theta}} \times \Gamma(1/\theta)$$
$$= \frac{1}{(s+1/\theta)^{1/\theta}\theta^{1/\theta}}$$
$$\frac{\theta^{1/\theta}}{(s\theta+1)^{1/\theta}} \times \frac{1}{\theta^{1/\theta}}$$
$$= (1+s\theta)^{-1/\theta}$$

Mean and variance will therefore be gotten from 1st and 2nd derivatives w.r.t s of laplace transformation

$$L^1(s) = -1(1+s\theta)^{-1/\theta-1}$$
$$L^2(s) = (1+\theta)(1+s\theta)^{-1/\theta-2}$$

letting $s = 0$, the mean together with variance of $W$, which is the frailty term we will have them as;

$$E(W) = (-1)L^1(0) = 1$$
$$Var(W) = L^2(0) - (-L^1(0))^2$$
$$= (1+\theta) - 1$$
$$= \theta$$

Hence the mean and variance of a one parameter gamma distribution is 1, and $\theta$ respectively.

**Positive stable frailty distribution**

Lets take that $W$ follows a stable distribution. Stable distribution do posses this property that with $W_1, ..., W_n$ independently and identically distributed random variables, for every $n$ there is a normalizing constant $c(n)$ so that $F(\sum_{i=1}^{n} W_i) = F(c(n)W_i)$ whereby $F(W)$ refers to the distribution law of $W$. This constant $c(n)$ takes the form $n^{1/\alpha}$ such that $\alpha\varepsilon(0,2]$, $\alpha$ is called the stability parameter Goethals et al. (2008). When $\alpha = 2$, it means that the distribution has a finite variance and it corresponds to Normal distribution. Incase of $\alpha = 2$ we therefore have; $F(\sum_{i=1}^{n} W_i) = F(n^{1/2}W_i)$.For $\alpha < 2$ the distributions will give undefined variance,and undefined mean when $\alpha \leq 1$. Stable distributions on positive half have $\alpha\varepsilon(0,1], (\alpha = 1)$ corresponds to a degenerate distribution. The pdf is expressed as;

$$g(w) = -\frac{1}{\pi w}\sum_{k=1}^{\infty}\frac{\Gamma(k\alpha+1)}{k!}(-w^{-\alpha})^k sin(\alpha k\pi), 0 < \alpha < 1$$

This pdf has undefined mean and variance. We have its Laplace transform written as below;

$$L(s) = exp(-s^{\alpha})$$

One of the key reasons why this distribution was proposed is that it has infinite mean, though it is difficult to work with infinite mean Goethals et al. (2008).

**Power variance function frailty**

Lets take that $W$ follows a power variance function distribution. According to Hougaard (1986) power variance function do act as an extended positive stable distribution. It's p.d.f would be expressed as;

$$g(w) = exp(-\frac{v}{\alpha}(\frac{w}{\mu} + \frac{1}{v-1}))\frac{1}{\pi w}\sum \frac{(v/\alpha)^{kv}(w/v)^{k(v-1)}\Gamma(1-k(v-1))sin(\pi k(v-1))}{k!(v-1)^k}$$

where $\mu > 0$, $\alpha > 0$ and $0 < v \leq 1$. It's laplace transform is expressed as Aalen (1992):

$$L(s) = exp[\frac{v}{\alpha(1-v)}(1-(1+\frac{\alpha\mu s}{v})^{1-v})]$$

The 1st and 2nd derivatives of laplace transform would be written as;

$$L^1(s) = -\mu(1+\frac{\alpha\mu s}{v})^{-v}exp[\frac{v}{\alpha(1-v)}(1-(1+\frac{\alpha\mu s}{v})^{1-v})]$$

$$L^2(s) = -\mu(1+\frac{\alpha\mu s}{v})^{-v} \times -\mu(1+\frac{\alpha\mu s}{v})^{-v}exp[\frac{v}{\alpha(1-v)}(1-(1+\frac{\alpha\mu s}{v})^{1-v})]$$
$$+exp[\frac{v}{\alpha(1-v)}(1-(1+\frac{\alpha\mu s}{v})^{1-v})]\mu v(1+\frac{\alpha\mu s}{v})^{-v-1}\frac{\alpha\mu}{v}$$

$$E(W) = (-1)L^1(0) = \mu$$
$$Var(W) = L^2(0) - (-L^1(0))^2$$
$$= \mu^2 + \mu^2\alpha - \mu^2$$
$$= \mu^2\alpha$$

Let's take $\mu^2\alpha = \theta$.
$W$ is referred to as the frailty term, when the variance of the frailty term $W$, $\theta = 0$ statistically, it implies independence within groups hence no presence of unobserved heterogeneity, but when $\theta > 0$ statistically, it implies there is association within groups hence there is presence of unobserved heterogeneity.

### Normal (Gaussian) frailty model

The Gaussian frailty p.d.f is written as;

$$g(w) = \frac{1}{\sqrt{2\Pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(w-\mu)^2} dw$$

whereby $\mu$ is the distribution's mean, $\sigma$ is the standard deviation,$w\varepsilon(-\infty,\infty)$ and $\sigma^2$ is the variance. More details are given in sub-section 3.7.2

### Inverse Gaussian frailty

The Inverse Gaussian frailty probability density function is expressed as;

$$g(w) = \frac{1}{\sqrt{2\Pi\theta}} w^{-\frac{3}{2}} exp(-\frac{(w-1)^2}{2\theta w})$$

whereby $\theta > 0$, $w > 0$. Laplace transformation is given by;

$$L(s) = exp(\frac{1}{\theta}(1-\sqrt{1+2\theta s}))$$

Whereby $s \geq 0$. To get the mean and variance we need to find 1st and 2nd derivatives of Laplace transform;

$$L^1(s) = -(1+2\theta s)^{-\frac{1}{2}} e^{(\frac{1}{\theta}(1-\sqrt{1+2\theta s}))}$$

$$L^2(s) = (1+2\theta s)^{-1} e^{(\frac{1}{\theta}(1-\sqrt{1+2\theta s}))} + e^{(\frac{1}{\theta}(1-\sqrt{1+2\theta s}))} \theta (1+2\theta s)^{-\frac{3}{2}}$$

$$E(w) = (-1)L^1(0) = 1$$

$$Var(w) = L^2(0) - (-L^1(0))^2$$

$$Var(w) = 1 + \theta - 1 = \theta$$

Hence the mean and variance of inverse gaussian distribution is 1, and $\theta$ respectively.

## 1.2 Background

Based on World Health Organization (WHO), worldwide, the rate of mortality as far as under-five is concerned has declined with a margin 56%, this is from an approximated rate of 93 deaths for every 1000 live births in 1990 to 41 deaths for every 1000 live births in the year 2016. Approximately 20,000 lesser children succumbed to death each day in 2016 compared to 1990 WHO (2017). In Africa back in the year 1970, rate of mortality for children was at 229 for every 1000 live births. By the year 2010, this given rate had gone down by more than half to 111 deaths for every 1000 live births Statista (2017). The under-five child mortality has reduced by 39% in Sub-Saharan Africa between 1990 and 2011. If at all, this kind of trend continues, then 1 in 3 children in the entire world will be born in sub-Saharan Africa, and hence its under-five population will grow quickly USAID (2017).

In the developing nations, the study of under-five child mortality has always been a vital issue in civic health programs. A nation's level of wealth index growth and quality of life are imitated by its under-five child mortality rates. To monitor and assess population and healthiness programs and guidelines, the under-five child mortality rates are used. The rates are likewise valuable in finding promising directions for health and diet programs in a nation. Rate of mortality for infants is at 39 deaths for every 1000 live births, while for the under-five is at 52 deaths for every 1000 live births. With this scenario, almost 1 in every 26 Kenyan children do die before the age of 1, while almost 1 in every 19 do die before the age of 5. Rates of mortality in the early childhood as a whole have gone down between 2003 and 2014 as far as KDHS surveys are concernedKDHS (2014).

To fast-track the attainment of Sustainable Development Goals (SDG's), the Kenyan Government started a Child Survival and Development Strategy which was included in the 2009 budget so as to increase survivorbility of children and also to offer an outline to boost measures for children. This plan is steered by the National Health Sector Strategic Plan II (NHSSP II) and the Vision 2030 Medium Term Plan that purpose to lower the level of imbalance in healthcare services and advance on the child health measures UNDP (2009).

Understanding under-five mortality in Uganda, the determinants/ risk factors were found to be; Mother's educational level, mother's age group, type of residence, education level of mother, education level partner, birth status, gender of child, wealth index, children ever born, birth order, religion, type of toilet facility, mother's occupation, births in the past one year, children below 5 years in household, gender of head of the household, source of drinking water, and age of mother at 1st birth. Nasejje et al. (2015).

The primary objective of this research is to use survival analysis techniques on Kenya Demographic Health Survey data for the year 2014 to identify the factors responsible for the under-five child mortality in Kenya, and to examine the effects of unobserved covariates (frailty) on under-five mortality both at household and community levels.

## 1.3 Statement of Problem

In the demographic health surveys studies as far as under-five mortality is concerned, selection of risks factors based on literature can possibly lead to erroneous choice of determinants. Therefore, use of predictive modelling and machine learning technique (Random Survival Forest) can assist in selection of these risks factors in a reliable manner.

In studying the under-five mortality, the use of the Accelerated Failure Time (AFT) model is to assume that effects of the covariates are either to accelerate or decelerate the survival life time of the under-five children by some constant. But these determinants do not at all times take into consideration the actual differences in the risk particularly in clustered survival data. So, inclusion of the unobserved random factor (frailty term) on the model improves correct measure of the determinants effect, thereby evading the problem of overestimation or underestimation of the model parameters.

## 1.4 OBJECTIVES

### 1.4.1 Overall Objective

The broad objective of this research is to use survival analysis techniques on Kenya Demographic Health Survey data for the year 2014 to identify the factors responsible for the under-five child mortality in Kenya, and to examine the effects of unobserved covariates (frailty) on under-five mortality both at household and community levels.

### 1.4.2 Specific Objectives

- Selection of risks factors of the under-five mortality using Random Survival Forest.
- Doing a comparison between the Accelerator Failure Time (AFT) model and AFT-shared frailty model both at household and community levels to examine the effects of unobserved heterogeneity on under-five mortality.
- Modeling and assessing the covariates that accelerate or decelerate the time until the event death.

## 1.5 Significance of Study

Identifying and estimating the risks factors of the under-five child mortality in a reliable manner will assist the government in reducing the under-five deaths in our communities more effectively in the future, and thereby leading to a healthy nation.

# 2 Literature Review

This chapter include current knowledge, findings in addition to theories and the methods that have been applied before.

### 2.0.1 Frailty models

Getting to understand the risks factors of the under-five mortality in Uganda by employing shared frailty model, it was found out that there was presence of unobserved random factors in household cluster while there was no proof to conclude existence of unobserved random factors at community cluster, gender of head of household , gender of child, and births over the past one year were established to be having significant influence on mortality Nasejje et al. (2015).

Studying childhood mortality in India using shared frailty models, there was presence of unobserved heterogeneity both at individual and community levels, it was established that infant mortality among women married before 18 years of age was thrice compared to female married after 17 years of age. Yadav and Yadav (2016).

Study carried out on under-five mortality using frailty models in Ehiopia, it was found out that there was presence of unmeasured household effects present in the model. It was also established that children who were residing in rural parts of Ethiopia were at higher risk of mortality as opposed to their counterparts residing in urban parts. Also, mothers who were older in age the likelihood of child's death before attaining 5 years was lesser Ayele et al. (2017).

SWAIN and GROVER (2016) used AFT-shared frailty approach for studying HIV/AIDS persons who were on Anti-retroviral therapy. It was found out that there was proof of the existence of heterogeneity among HIV/AIDS persons. It was also established that risks factors weight, gender, manner of transmission and reference haemoglobin were statistically significant for HIV/AIDS persons on ART.

Njagi (2011) used AFT-shared frailty to study the urban rural differentials of infant mortality in Kenya, it was found out that there was presence of unobserved factors, although there was no much difference in results in both models with and without frailty.Pan (2001) did propose accelerated failure time model on presuming frailty on error term which is referred to as accelerated failure time gamma frailty .

### 2.0.2 Random survival forests

Random survival forest refers to combination of tree predictors whereby every individual tree relies on values of random vector that has been sampled in an independent manner and with same distribution for all the trees contained in the forest Breiman (2001) .
From a study carried out on under-5 mortality using UDHS data. Random survival forests, ranking according to variable importance (variables selected based on literature) indicated that; number of children under the age of 5 in household, births over the past 5 years, birth order, wealth index, and the number of children that have ever been born in household had strong influence on mortality rate for children Nasejje and Mwambi (2017)

# 3 Methodology

## 3.1 Data

This 2014 KDHS sample data set was taken from a master sampling frame, Fifth National Sample Survey, and Evaluation Programme (NASSEPV). This is a kind of structure that Kenya National Bureau of Statistics (KNBS) at the moment uses to carry out surveys in households in the country. Kenya as a country is divided into a total of forty-seven counties. In the process concerning this development of NASSEPV, each of these forty-seven counties was stratified in two categories; urban and rural strata. Because both Mombasa and Nairobi counties have just urban zones, the total outcome came to 92 strata. This sample had a total of 40,300 households from 1612 community clusters that were rolled out in entire nation, with 995 clusters from the countryside and 617 from non-rural zones. The samples were chosen in an indepedent manner in every sampling stratum by applying two-stage sample approach. In the 1st stage, the 1612 community clusters were chosen with equal likelihood from NASSEPV frame. Households from operations that were listed were taken to act as sampling frame for 2nd stage of selection, where a total of 25 households were chosen from every cluster KDHS (2014).
This KDHS-2014 dataset includes women of ages between 15 to 49 years. This study includes only children of between 1-59months old, accounting for a total observation of 20354.

## 3.2 Variable Selection

The original data has 1099 variables excluding survival time and event variables, out of these 313 variables had 100% missing data which were deleted. Therefore, random forest for survival regression and classification was applied to the remaining 786 covariates to select those variables that had influence on under-five mortality, ranking the variables according to their importance, splitting rule used was log rank.

## 3.3 Steps to develop algorithm for random survival forests

- $B$ bootstrap samples are taken from original data. Every bootstrap sample does not include on average 37% of the data, this data that is not included is referred to as test data or Out-Of Bag data (OOB data).
- A survival tree is grown for every bootstrap sample. At every node of the tree, $p$ candidate variables are randomly chosen. The node is split by using that candidate variable which maximizes the difference in survival between daughter nodes.
- Tree is grown to fullest size on condition that a terminal node should have minimum of $d_0 > 0$ deaths.
- Compute cumulative hazard function (CHF) for every tree, then average to get ensemble cumulative hazard function (CHF).
- Use the test data to compute prediction error for ensemble cumulative hazard function (CHF).

### 3.3.1 Log-rank split rule

Given that node $u$ can be split into two daughter nodes say $\alpha$ and $\beta$.Best split at node $u$, on variable $x$ at $c*$ which is splitting point is one which would give the largest log-rank statistic between the two daughter nodes. Log-rank statistic for split on variable $x$ at a given covariate value $c*$ can be defined as;

$$i(x,c*) = \frac{\sum_{j=t_1}^{t_N}(d_{\alpha,j} - E(D_{\alpha,j}))}{\sqrt{\sum_{j=t_1}^{t_N} var(D_{\alpha,j})}}$$

whereby $d_{\alpha,j}$ refers to number of events in daughter node $\alpha$ at time point $j$. The expected number of events in daughter node $\alpha$, $E(D_{\alpha,j})$ and its variance are given by;

$$E(D_{\alpha,j}) = d_j(\frac{R_{\alpha,j}}{R_j})$$

$$var(D_{\alpha,j}) = \frac{R_{\alpha,j}}{R_j}(1 - \frac{R_{\alpha,j}}{R_j})(\frac{R_j - d_j}{R_j - 1})d_j$$

Whereby $d_j$ is the combined number of events in daughter nodes $\alpha$ and $\beta$ at time point $j$. $R_{\alpha,j}$ represents number of subjects at risk in node $\alpha$ at time point $j$ and $R_j$ the combined number at risk in daughter nodes $\alpha$ and $\beta$.

### 3.3.2 Computation of ensemble CHF

Finally, the survival tree will reach a saturation point, at this point it's impossible to form new daughters because of the condition that every node should have minimum of $d_0 > 0$ deaths. The nodes which are most extreme in a tree that is saturated are referred to as terminal nodes. We can denote them by $V$. Let us take $(T_{1,a}, \delta_{1,a}), .., (T_{n(a),a}, \delta_{n(a),a})$ to represent study times and event indicator for individuals in terminal node $a \varepsilon V$. The event indicator shows if observation corresponds to an event ($\delta_{j,a} = 1$) at time $T_{j,a}$ or has been censored ($\delta_{j,a} = 0$) at time $T_{j,a}$ . Let $t_{1,a} < t_{2,a} <, ..., < t_{N(a),a}$ represent $N(a)$ distinct event times. Let $d_{l,a}$ and $Y_{l,a}$ be number of dead subjects and subjects at risk at time $t_{l,a}$. Estimate of CHF for $a$ would be expressed as

$$\hat{\Lambda}_a(t) = \sum_{t_{l,a} \leq t} \frac{d_{l,a}}{Y_{l,a}}$$

All individuals in node $a$ have same CHF. Every individual $j$ has a d-dimensional covariate $x_j$. Let $\Lambda(t|x_j)$ be the CHF for $j$.

$$\Lambda(t|x_j) = \hat{\Lambda}_a(t),$$

if $x_j \varepsilon a$, describes the CHF for all subjects and describes the CHF for the tree.
To calculate ensemble CHF, we will add all the CHFs and then divide by $B$ survival trees. Let $\Lambda^*_b(t|x)$ denote CHF $(\Lambda(t|x_j))$ for a tree that is grown from the $b^{th}$ bootstrap sample. Therefore, the bootstrap ensemble CHF for $j$ is given by

$$\Lambda^*_e(t|x_j) = \frac{1}{B} \sum_{b=1}^{B} \Lambda^*_b(t|x_j)$$

### 3.3.3 Calculation of prediction error

To approximate prediction error, we apply Harrell's Concordance index. It is referred to as C-index, it gives probability estimate that, in a randomly chosen pair of subjects, that subject which fails 1st is said it had worst predicted outcome.

**Steps in calculating C-index**

- All possible pairs of subjects are formed throughout data.
- We leave out the pairs where shorter survival times has been censored. Leave out pairs $i$ and $j$ provided that $T_i = T_j$ except if at least one of them is an event (death). Let permissible denote the total number of permissible pairs.
- For every pair that's permissible whereby $T_i \neq T_j$, it is counted 1 if the survival time that is shorter has worse predicted outcome; it is counted 0.5 provided that predicted outcomes are tied. For every pair that's permissible whereby $T_i = T_j$ and both are incidents, it is counted 1 on condition that predicted outcomes are tied; or else it is counted 0.5. For every pair that's permissible whereby $T_i = T_j$, but not that both are incidents, it is counted 1 if the incident has worse predicted outcome; or else it is counted 0.5. We let concordance indicate the sum over all permissible pairs.
- $C$-index, $C$, is explained by $C = Concordance/Permissible$.

**OOB prediction error**

To compute $C$ we require a predicted outcome. We have to use OOB ensemble CHF to describe predicted outcome.
To get OOB ensemble CHF we proceed as follows;
It is worth noting every tree in the forest is grown using a sample that's independent. Interpret $l_{j,b} = 1$ on condition that $j$ is an OOB case for $b$; otherwise, interpret $l_{j,b} = 0$. Let $\Lambda_b^*(t|x)$ indicate CHF $(\Lambda(t|x_j))$ for a tree grown from the $b^{th}$ bootstrap sample. The OOB ensemble CHF for $j$ is

$$\Lambda_e^{**}(t|x_j) = \frac{\sum_{b=1}^{B} l_{j,b}\Lambda_b^*(t|x_j)}{\sum_{b=1}^{B} l_{j,b}}$$

Since this figure is resulting from test data, it can be used to get an OOB estimate for $C$, and, so, an OOB error rate.
Let $t_1^0, ..., t_m^0$ denote pre-chosen unique time points. Ranking two subjects $i$ and $j$, we say j has a worse predicted outcome than $i$ if

$$\sum_{l=1}^{m} \Lambda_e^{**}(t_l^0|x_j) > \sum_{l=1}^{m} \Lambda_e^{**}(t_l^0|x_i)$$

Using this law, calculate $C$ as shown from the above steps. Indicate OOB estimate with $C^{**}$. OOB prediction error, $PE^{**}$, is described as $1 - C^{**}$. It is worth noting that $0 \leq PE^{**} \leq 1$

### 3.3.4 Importance of a variable (VIMP)

To compute importance of a covariate $x$, test data cases are dropped into their training survival tree. Whenever split on $x$ occurs, daughter node is assigned in a random manner. The cumulative hazard function from every such tree is computed and mean is found. Therefore, VIMP for variable $x$ is the error of prediction for the new ensemble gotten by means of randomizing $x$ assignments then subtracting the error of prediction for the original ensemble from it.

## 3.4 Estimation of $S(t)$

We have used Kaplan-Meier estimate to explain how risk of dying for the under-five children is distributed across groups of the chosen determinants based on Random Forest for Survival, Regression, and Classification (RF-SRC) on KDHS 2014 data.

$$\hat{S}(t) = \prod_{j=1}^{k} \frac{n_j - d_j}{n_j}$$

$\hat{S}(t)$ is the Kaplan Meier estimate, where; $n_j$ represents the number of children who are at risk of dying at beginning of period $j$, and $d_j$ is deaths at the beginning of period $j$.

### 3.4.1 How to derive estimation of $S(t)$

We can derive estimator of the Kaplan-Meier from maximum likelihood estimation of the hazard function. Given that $d_j$ is deaths at beginning of period $j$, $n_j$ be children who are at risk of dying at beginning of period $j$, discrete hazard rate $\lambda_j$ can be described as the probability of a child with an event (death) at the beginning of period $j$. Then the survival rate can be described as:

$$S(t) = \prod_{j=1}^{k} (1 - \lambda_j)$$

and therefore, likelihood function for the hazard function upto period $j$ is;

$$L(\lambda_j : j \leq k | d_j : j \leq k, n_j : j \leq k) = \prod_{j=1}^{k} \lambda_j^{d_j} (1 - \lambda_j)^{n_j - d_j}$$

Thus the loglikelihood will be;

$$log(L) = \sum_{j=1}^{k} (d_j log(\lambda_j) + (n_j - d_j) log(1 - \lambda_j))$$

To obtain the maximum of log likelihood with respect to $\lambda_j$ we get:

$$\frac{\partial log(L)}{\partial \lambda_j} = \frac{d_j}{\hat{\lambda}_j} - \frac{n_j - d_j}{1 - \hat{\lambda}_j} = 0 \Rightarrow \hat{\lambda}_j = \frac{d_j}{n_j}$$

whereby hat denotes the maximum likelihood estimation. Given this result we can therefore, write:

$$\hat{S}(t) = \prod_{j=1}^{k}(1-\hat{\lambda}_j)$$

$$\hat{S}(t) = \prod_{j=1}^{k}(1-\frac{d_j}{n_j})$$

## 3.5 Survival regression

### 3.5.1 Accelerated failure time (AFT) model

Is an approach which needs distributional assumptions. It assumes that the effects of predictors are either to accelerate or decelerate survival life time by some constant. Given the values of the covariates $X$. The pdf is given as ;

$$f(t) = (\sigma t)^{-1} f_o(\frac{logt - log\psi(X)}{\sigma}) \quad \textbf{(1)}$$

whereby $\sigma$ is referred to as scale parameter, and $\psi(X)$ is some function of covariates.

$$\psi(X) = exp(\mu + X'\beta) \quad \textbf{(2)}$$

Therefore, corresponding Accelerator Failure Time (AFT) model expression in regression form is;

$$logT = \mu + X'\beta + \sigma\varepsilon \quad \textbf{(3)}$$

whereby $\mu$ is an intercept, $\varepsilon$ is a variable which is random and has a density function $f_o(\varepsilon)$ and the corresponding baseline survival function $S_o(\varepsilon)$. AFT models do allow a wide range of parametric forms for the density function. The survival function of the AFT models is given as;

$$S(t) = S_o^*[(\frac{t}{\psi(X)})^{\frac{1}{\sigma}}] = S_o(\frac{logt - log\psi(X)}{\sigma}) \quad \textbf{(4)}$$

whereby $S_o^*$ is the baseline survival function. Since $\psi(X) = exp(\mu + X'\beta)$, the survival function can therefore, be rewritten in the form;

$$S(t) = S_o(\frac{logt - \mu - X'\beta}{\sigma}) \quad \textbf{(5)}$$

### 3.5.2 Inference for AFT models

For random lifetime $T_i$ of the subjects $i = 1, .., n$, the likelihood function under model (5) is expressed as;

$$L(\beta, \sigma) = \prod_{i=1}^{n} (\frac{1}{\sigma} f_o(\frac{logt_i - \mu - X'\beta}{\sigma}))^{\delta_i} S_o(\frac{logt_i - \mu - X'\beta}{\sigma})^{1-\delta_i} \quad \textbf{(6)}$$

Using $\varepsilon_i = \frac{logt_i - \mu - X'\beta}{\sigma}$, the log-likelihood function would take the form;

$$l(\beta, \sigma) = -rlog\sigma + \sum_{i=1}^{n} [\delta_i log f_o(\varepsilon_i) + (1 - \delta_i) log S_o(\varepsilon_i)] \quad \textbf{(7)}$$

where $r = \sum \delta_i$ refers to the number of events. Let $X_i' = (X_{i1}, ..., X_{ij}, ..., X_{ip})$ denote the set of covariates under which the $i^{th}$ subject responds. The first partial derivatives of $l(\beta, \sigma)$ are given by;

$$\frac{\partial l}{\partial \beta_j} = -\frac{1}{\sigma} \sum_{i=1}^{n} [\delta_i \frac{\partial log f_o(\varepsilon_i)}{\partial \varepsilon_i} + (1 - \delta_i) \frac{\partial S_o(\varepsilon_i)}{\partial \varepsilon_i}] X_{ij} \quad \textbf{(8)}$$

$$\frac{\partial l}{\partial \sigma} = -\frac{r}{\sigma} - \frac{1}{\sigma} \sum_{i=1}^{n} [\delta_i \varepsilon_i \frac{\partial log f_o(\varepsilon_i)}{\partial \varepsilon_i} + (1 - \delta_i) \varepsilon_i \frac{\partial log S_o(\varepsilon_i)}{\partial \varepsilon_i}] \quad \textbf{(9)}$$

Maximum likelihood estimators $\hat{\beta}$ and $\hat{\sigma}$ are gotten through solving the equations $\frac{\partial l}{\partial \beta} = 0$ and $\frac{\partial l}{\partial \sigma} = 0$

The distributions mostly used in AFT models have been mentioned and explained in details below:

### 3.5.3 Exponential distribution

It has only one parameter, it's pdf is expressed by;

$$f(t) = \lambda e^{-\lambda t}, t > 0, \lambda > 0$$

The survival function, $S(t)$ which refers to as the chances of an individual living upto or past time $t$ can be obtained from;

$$\begin{aligned} S(t) &= -\int_0^t \lambda e^{-\lambda u} du \\ &= e^{-\lambda t} \end{aligned}$$

Cumulative distribution function, $F(t)$ is expressed as;

$$F(t) = 1 - S(t) = 1 - e^{-\lambda t}$$

Hazard function,$h(t)$ is given by;

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

Cumulative hazard function, $H(t)$ is given by;

$$H(t) = \int_0^t \lambda du = \lambda t$$

### 3.5.4 Weibull distribution

This distribution has two parameters, it's pdf can be given by;

$$f(t) = \lambda k t^{k-1} e^{-\lambda t^k}, \lambda > 0, k > 0$$

It's survival function, $S(t)$ would be expressed b;

$$S(t) = e^{-\lambda t^k}$$

The cumulative distribution function, $F(t)$ can be gotten from;

$$F(t) = 1 - S(t) = 1 - e^{-\lambda t^k}$$

The hazard function,$h(t)$ would be given by;

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda k t^{k-1} e^{-\lambda t^k}}{e^{-\lambda t^k}} = \lambda k t^{k-1}$$

Cumulative hazard function, $H(t)$ would be given by;

$$H(t) = \int_0^t \lambda k u^{k-1} du = \frac{\lambda k u^k}{k} |_0^t = \lambda t^k$$

### 3.5.5 Lognormal distribution

A variable which is random, say $T$, is said to be distributed lognormally on condition that $Y = ln(T)$ is distributed normally with natural logarithm. General formular for pdf of the lognormal distribution is;

$$f(t) = \frac{e^{-((ln((t-\theta)/m))^2/(2\sigma^2))}}{(t-\theta)\sigma\sqrt{2\pi}}$$

where $t > \theta; m, \sigma > 0$, $\sigma$ refers to shape parameter , $\theta$ represents parameter for location and $m$ refers to parameter for scale (and is also median for the distribution)
When $t = \theta$, therefore, $f(t) = 0$. In a scenario whereby $\theta = 0$, and $m = 1$ we will have distribution which is standard. In a scenario whereby $\theta = 0$ we will have a two-parameter distribution.
Therefore, pdf for standard form of this distribution is;

$$f(t) = \frac{e^{-((lnt)^2/2\sigma^2)}}{t\sigma\sqrt{2\pi}}$$

where $t > 0$, $\sigma > 0$, because the general form of the pdf can be written with regard to standard form, all succeeding formulas here are written specifically for standard form.
Cumulative distribution function, $F(t)$ is expressed as

$$F(t) = \Phi(\frac{ln(t)}{\sigma})$$

where $t \geq 0$; $\sigma > 0$. $\Phi$ refers to the cumulative distribution function of normal distribution.
Survival function would be written as;

$$S(t) = 1 - F(t) = 1 - \Phi(\frac{ln(t)}{\sigma})$$

Hazard function is expressed as;

$$h(t) = \frac{(\frac{1}{t\sigma})\phi(\frac{lnt}{\sigma})}{\Phi(\frac{-lnt}{\sigma})}$$

where $t > 0$, $\sigma > 0$. $\phi$ refers to pdf of normal distribution.
The cumulative hazard function would be written as;

$$H(t) = -ln(1 - \Phi(\frac{ln(t)}{\sigma}))$$

where $t \geq 0$, $\sigma > 0$

### 3.5.6 Log-logistic distribution

It is also a parametric model which is applied in survival analysis for those events whose rates do increase initially and decrease later, eg mortality rate from a disease say cancer following treatment or diagnosis.
This distribution is a probability distribution of a random variable whose logarithm has a logistic distribution. It's probability density function is expressed as;

$$f(t) = \frac{(\beta/\alpha)(t/\alpha)^{\beta-1}}{(1+(t/\alpha)^{\beta})^2}$$

where $t > 0$, $\alpha > 0$, $\beta > 0$. $\alpha$ and $\beta$ represent scale and shape parameters respectively. Cumulative distribution function (CDF), $F(t)$, would be written as;

$$F(t) = \frac{1}{1+(t/\alpha)^{-\beta}}$$

It's survival function, $S(t)$ is;

$$S(t) = 1 - F(t) = [1+(t/\alpha)^{\beta}]^{-1}$$

It's hazard function, $h(t)$ is expressed as;

$$h(t) = \frac{f(t)}{S(t)} = \frac{(\beta/\alpha)(t/\alpha)^{\beta-1}}{1+(t/\alpha)^{\beta}}$$

## 3.6 Best fitting model selection

According to Akaike (1987), the factual truth of meanness regarding building a statistical model do dictate that increase in number of the parameters must be stopped the moment it has been realized any more increase doesn't give a significant upgrade of the fit of the model to the data. He suggested employment of Akaike Information Criterion (AIC) which can be expressed as follows;

$$AIC = -2Loglikelihoog + 2numberofparameters$$

As a quantification of Goodness of Fit (GOF) of the model described by parameters approximated by the maximum likelihood approach. It is worth noting that the Akaike Information Criterion (AIC) value will always increase if an unnecessary variable has been included in the model. Therefore, this means that the smaller the AIC the better the model.

## 3.7 Fraily model

Shared frailty models have been applied , this has been done both at household and community levels of clustering separately.Here we have analyzed factors that do affect the under-five mortality in Kenya while taking into consideration unobserved heterogeneity in the data using Accelerator Failure Time (AFT) shared frailty models. The frailty term has taken into account the scenario where some of the children might be exposed to the risk of death before the age of five than the others. Because of the unmeasured or unobserved factors some children are more likely to die than others. The frailty term captures total effects of all factors that influence the child's risk of death that are not included in the AFT model. The model that has been presented here takes into consideration both the observed and unobserved effects. This model has been based on the AFT model. Frailty term gets into the AFT model as random effects. Estimated variance of the frailty effects is used to test if frailty term is significant. When the variance of the frailty term is zero it means that there is no presence of unobserved heterogeneity among groups hence there is independence within groups, while a large variance implies presence of unobserved heterogeneity among groups hence there is association within groups.

### 3.7.1 AFT models with shared frailty

Shared frailty models are suitable when subjects within a cluster share a common unobserved heterogeneity. Off-late AFT models with shared frailty have received some attention. Given $q$-dimensional vector of random effects $w_i$ event times within cluster are assumed to be independent. AFT models with shared frailty is expressed in the following form

$$logT_{ij} = \mu + X'_{ij}\beta + w_i + \sigma\varepsilon_{ij} \tag{10}$$

whereby $\mu$ is an intercept, $\beta$ is a vector of regression coefficients, $X_{ij}$ is the vector of fixed-effect covariate, $\sigma$ is a scale parameter, $\varepsilon'_{ij}s$ are independent and identically distributed random errors, and $w'_is$ are the frailty terms which are assumed to be independent and identically distributed with density function $f(w_i)$. Here, frailty could be an unobserved covariate which is additive on the log failure time scale describing some reduced or increased event times for different clusters. All subjects in a given cluster share a common frailty.

AFT models with shared frailty do specify a direct linear relationship between the log of failure time and the covariates. The survival function for an AFT model at time $t$ is expressed as;

$$S(t) = S^*_o[(\frac{t}{\psi_{ij}})^{\frac{1}{\sigma}}] = S_o(\frac{logt - log\psi_{ij}}{\sigma}) \tag{11}$$

whereby $\sigma$ refers to the scale parameter, $S^*_o$ is a survival function defined on $(0,\infty)$, and $S_o$ is the baseline survival function that satisfies the relationship $S^*_o(\omega) = S_o(log\omega)$, $\psi_{ij}$ is

some function of the covariates. Where $\psi_{ij} = exp(\mu + X'_{ij}\beta + w_i)$. Conditional survival function is given by;

$$S_{ij}(t|w_i) = S_o(\frac{logt - \mu - X'_{ij}\beta - w_i}{\sigma}|w_i) \tag{12}$$

where $S_o(.)$ is the survival function of $\varepsilon_{ij}$ and $\mu$ is an intercept, $\beta$ is vector of regression coefficients, $X_{ij}$ is a vector of fixed-effect covariate of the $j^{th}$ subject in the $i^{th}$ cluster. Here we have assumed that the frailty term, $w_i$ follows Gaussian distribution with mean and variance of $\mu$ and $\theta$ respectively. With $\varepsilon_{ij} = \frac{logT_{ij} - \mu - X'_{ij}\beta - w_i}{\sigma}$, the conditional survival and hazard functions are expressed as;

$$S_{ij}(t|w_i) = S_o(\varepsilon_{ij}|w_i) \tag{13}$$

$$h_{ij}(t|w_i) = \frac{1}{\sigma t}h_o(\varepsilon_{ij}|w_i) \tag{14}$$

respectively, whereby $h_o(.)$ is the hazard function of $\varepsilon_{ij}$.
Let $G$ be the number of clusters, $i = 1, ..., G$, and $n_i$ be the number of subjects within the $i^{th}$ cluster. The conditional likelihood for the observed data is;

$$L_c = \prod_{i=1}^{G}\prod_{j=1}^{n_i}[\frac{1}{\sigma t_{ij}}h_o(\varepsilon_{ij}|w_i)]^{\delta_{ij}}S_o(\varepsilon_{ij}|w_i) \tag{15}$$

Integrating unobserved frailties ($w_i$) out, all the clusters will have marginal likelihood function which is given by;

$$L_m = \prod_{i=1}^{G}\int\prod_{j=1}^{n_i}[\frac{1}{\sigma t_{ij}}h_o(\varepsilon_{ij}|w_i)]^{\delta_{ij}}S_o(\varepsilon_{ij}|w_i)f(w_i)dw_i \tag{16}$$

Estimates of the parameters $(\sigma, \beta, \theta)$ can be found by maximizing the likelihood function (16).

### 3.7.2 Gaussian frailty

The Gaussian frailty probability density function is given by;

$$f(w) = \frac{1}{\sqrt{2\Pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(w-\mu)^2}dw$$

whereby $\mu$ is the mean of the distribution, $\sigma$ is the standard deviation,$w\varepsilon(-\infty, \infty)$ and $\sigma^2$ is the variance. Laplace transformation is given by;

$$L(s) = \int_{-\infty}^{\infty}e^{-sw}f(w)dw$$

$$= \int_{-\infty}^{\infty} e^{-sw} \frac{1}{\sqrt{2\Pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(w-\mu)^2} dw$$

let;

$$w - \mu = u$$

$$w = u + \mu$$

using change of variable technique;

$$dw = du$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\Pi\sigma^2}} e^{-s(u+\mu)} e^{-\frac{u^2}{2\sigma^2}} du$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\Pi\sigma^2}} e^{-su-s\mu} e^{-\frac{u^2}{2\sigma^2}} du$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\Pi\sigma^2}} e^{-su} e^{-s\mu} e^{-\frac{u^2}{2\sigma^2}} du$$

$$= e^{-s\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\Pi\sigma^2}} e^{-su} e^{-\frac{u^2}{2\sigma^2}} du$$

multiplying and dividing by $e^{\frac{s^2\sigma^2}{2}}$ we have;

$$= e^{-s\mu} e^{\frac{s^2\sigma^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\Pi\sigma^2}} e^{-su} e^{-\frac{s^2\sigma^2}{2}} e^{-\frac{u^2}{2\sigma^2}} du$$

$$-\frac{s^2\sigma^2}{2} - \frac{u^2}{2\sigma^2} = \frac{-\sigma^2 s^2 \sigma^2 - u^2}{2\sigma^2}$$

$$\frac{-\sigma^4 s^2 - u^2}{2\sigma^2}$$

$$= e^{-s\mu} e^{\frac{s^2\sigma^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\Pi\sigma^2}} e^{-su} e^{\frac{-\sigma^4 s^2 - u^2}{2\sigma^2}} du$$

$$-\frac{su}{1} + \frac{-\sigma^4 s^2 - u^2}{2\sigma^2} = \frac{-2\sigma^2 su - \sigma^4 s^2 - u^2}{2\sigma^2}$$

$$-\frac{1}{2\sigma^2}(u^2 + 2\sigma^2 su + \sigma^4 s^2)$$

$$-\frac{1}{2\sigma^2}(u + \sigma^2 s)^2$$

$$= e^{-s\mu + \frac{s^2\sigma^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\Pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(u+\sigma^2 s)^2} du$$

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\Pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(u+\sigma^2 s)^2} du = 1, N\left(\sigma^2 s, \sigma^2\right)$$

Therefore, the Laplace transform is given by:

$$L(s) = e^{-s\mu + \frac{s^2\sigma^2}{2}}$$

Mean and variance can therefore be obtained from the first and second derivatives of the laplace transformation.

$$L^1(s) = (-\mu + s\sigma^2)e^{-s\mu+\frac{s^2\sigma^2}{2}}$$

$$L^2(s) = (-\mu + s\sigma^2)(-\mu + s\sigma^2)e^{-s\mu+\frac{s^2\sigma^2}{2}} + \sigma^2 e^{-s\mu+\frac{s^2\sigma^2}{2}}$$

equating $s$ to 0, therefore the mean and variance from laplace becomes;

$$E(W) = (-1)L^1(0) = \mu \tag{17}$$

$$Var(W) = L^2(0) - (-L^1(0))^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2 \tag{18}$$

### 3.7.3 Chi-square test

This is a test which is performed when one wants to do an assessment of the goodness of fit between a set of observed and expected values in a theoretical manner.
This test has been applied in this study to establish whether or not there is an association within clusters. Presence of association within groups implies that there's existence of the unseen random factors, while no association implies that there's no existence of the unseen random factors.
The null and alternative hypothesis are stated as;

$H_0$ : No association within clusters
$H_1$ : There is association within clusters

The chi-square statistic is given by

$$\chi = \sum \frac{(O_i - E_i)^2}{E_i}$$

Whereby, $O_i$: refers to data which is observed, and $E_i$ refers to expected values.
Degree of freedom $(n-1)$ is used to read corresponding critical value at 5% level of significance on a chi-square table. The critical and statistic values are therefore compared, If the statistic value is greater than the critical value then it means the $p - value < 0.05$ hence the $H_0$ is rejected and conclude that there is association within clusters, therefore, there is presence of the unobserved heterogeneity. On the other hand if the statistic figure is less than the critical figure then it means $p - value > 0.05$ hence the $H_0$ is not rejected and conclude that there is no association within clusters, therefore, there is no presence of the unobserved heterogeneity.

# 4 Results

### 4.0.1 Variable selection using Random Survival Forest

**Table 1. Characteristics of the fitted Random Forest for Survival Regression and Classification**

| | |
|---|---|
| Sample size | 20354 |
| Number of deaths | 428 |
| Was data imputed | Yes |
| Number of trees | 1000 |
| Forest terminal node size | 3 |
| Average number of terminal nodes | 290.723 |
| Number of variables tried at each split | 29 |
| Total number of variables | 786 |
| Analysis | Random Survival Forest |
| Family | Survival |
| Splitting rule | Log-rank *random* |
| Number of random split points | 10 |
| Error rate | 0.31% |

From table 1 we have fitted random forest for survival regression and classification model on data set using log-rank split rule. It is worth noting that this model is built using the 786 covariates. To identify the most important covariates that influence under-five mortality in Kenya, permutation importance was applied to measure importance of variable Ishwaran et al. (2008)Strobl et al. (2008). Results for ranking variables according to their level of influence on under-five mortality are summarized in table 2 .

**Table 2. Variable importance (VIMP) for 786 covariates basing on an overall of 1000 trees using log-rank split criteria**

| Variable | Importance |
|---|---|
| Sons who have died | 0.0167 |
| Daughters who have died | 0.0112 |
| Living children+current pregnancy | 0.0036 |
| Sex of child | 0.0034 |
| Duration of breastfeeding | 0.0028 |
| Number of living children | 0.0021 |
| Months of breastfeeding | 0.002 |

### 4.0.2 Variable importance;

Variable importance is applied since we need to threshold covariates, since a covariate whose level of importance is less than 0.002 is likely regarded as a noise Ishwaran et al. (2008). Using this rule the variables found to be having high influence on under-five mortality based on their level of importance are displayed in table 2, and these are the variables that we have chosen to use as our explanatory variables .

**Table 3. Demographic and socioeconomic characteristics by child survival**

| Variable | Mortality,N(%) | Variable | Mortality,N(%) |
|---|---|---|---|
| Sons who have died | | Living children+ | |
| None | 18237(0.8) | None | 26(100) |
| 1son | 1775(13.1) | 1child | 3051(2.6) |
| 2sons | 275(12.4) | 2children | 4618(2.2) |
| 3sons | 58(12.1) | 3children | 3876(1.8) |
| 4sons | 7(14.3) | 4children | 2983(1.8) |
| 5sons | 1(0) | 5children | 2075(2.2) |
| 6sons | 1(0) | 6children and above | 3725(1.4) |
| Sex of child | | Daughters who've died | |
| Male | 10302(2.3) | None | 18503(1.1) |
| Female | 10052(1.9) | 1daughter | 1580(11.6) |
| Births in last five years | | 2daughters | 212(17) |
| One | 9531(1.2) | 3daughters | 44(20.5) |
| Two | 8707(2.4) | 4daughters | 14(21.4) |
| Three | 1999(4.6) | 5daughters | 1(0) |
| Four | 107(7.5) | Region | |
| Five | 10(10) | Coast | 2565(2.2) |
| Contraceptives use | | North eastern | 1556(1.7) |
| Currently using | 9760(1.6) | Eastern | 2930(1.6) |
| Used since last birth | 6233(2.7) | Central | 1371(1.7) |
| Used before last birth | 1091(2.8) | Rift valley | 6651(1.5) |
| Never used | 3270(2.1) | Western | 1926(2.8) |
| Currently breastfeeding | | Nyanza | 2845(3.7) |
| No | 14841(2.4) | Nairobi | 510(3.1) |
| Yes | 5513(1.2) | | |

Table 3 displays how deaths have been distributed for children under-5 years of age at every level that has been included during analysis. It shows that among mothers who had male children, out of 10302 children born,2.3% died before their fifth birthday then tread by mothers who had female children having 1.9% of deaths. This table has done a summary on how births and deaths of children have been distributed for rest of the variables that were considered during the study.

### 4.0.3 Non-parametric exploratory analysis methods:

Given that $t_j$ represents the time to death of the under-five child (child's age) in the Kenya Demographic Health Survey data set 2014, the non-parametric method which is mostly graphical was used to give a description of how the risk of death for the children under-five years of age is distributed across the strata of one of the chosen covariates. It is worth noting that what we meant by time to death is actually years of life until death for the child under five years of age. This time has been recorded in months for analysis. Figure 1 below is a Kaplan-Meier plot for the chosen factor based on Random Survival Forest, affecting under-five child survival in Kenya from the KDHS-2014 data set.

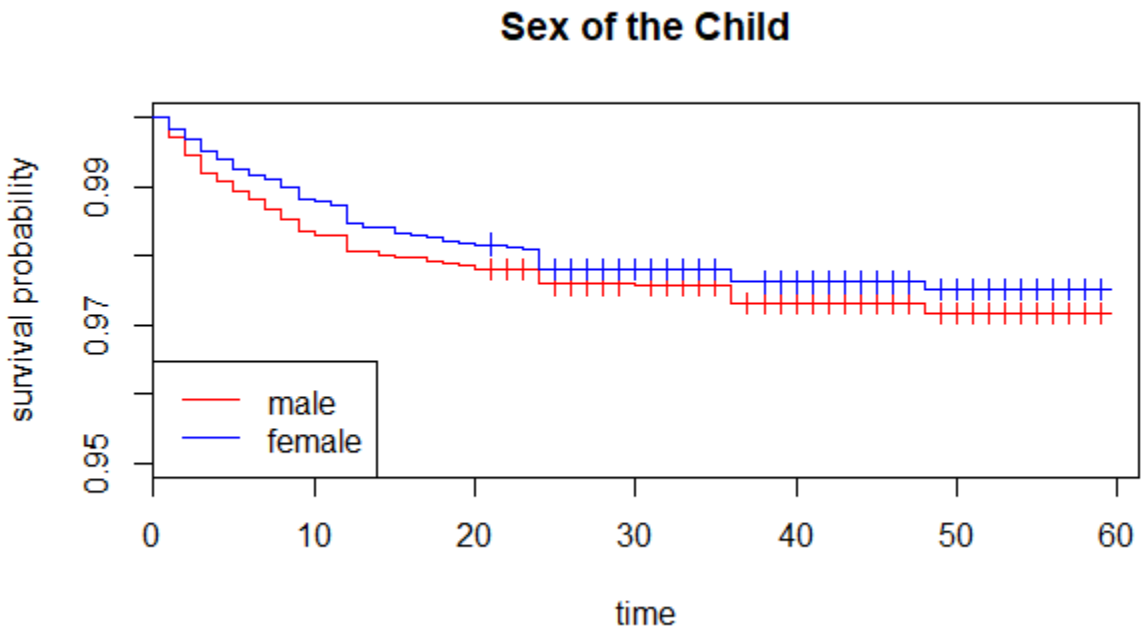


**Figure 1. Survival probability across gender of the child**

From figure 1, it is clear from the Kaplan-Meier plot that female children have a better survival probability than the male children. The reason behind this could be due to the fact that girls have a strong biological advantage in early childhood, another reason why it is this way is that most tribes in Kenya have custom of seeing a female child as the origin of wealth via bride price.

### 4.0.4 Modeling for determinants using AFT model variants:

Test for checking for the parametric model assumptions was performed for the four AFT-models; Exponential, Weibull, Lognormal, and Log-logistic distributions. Checking to see if a model satisfies the parametric model assumptions, plots are done for each model to see whether most of the data points lie within the straight line or 95% confidence limits, if they do then it implies that the assumptions have been met Mulera.B (2017). We found out that all these models met the assumptions. We have therefore, reported results from Log-logistic AFT model since it had the least Akaike Information Criterion (AIC) statistic.

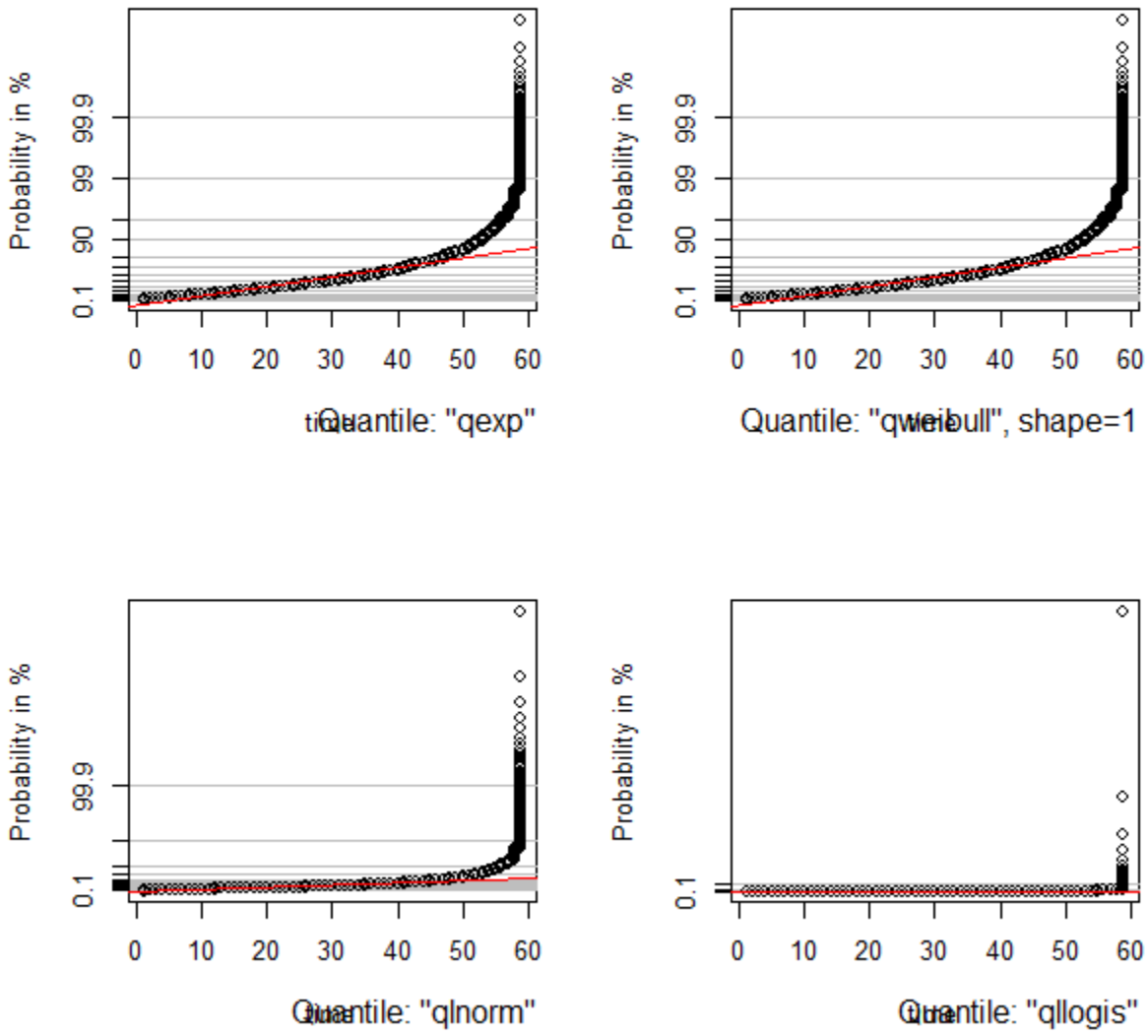


**Figure 2. Checking for parametric model assumptions for Exponential, Weibull, Lognormal and Log-logistic AFT-models**

**Table 4. AIC figures of AFT-shared frailty models (Community and household)**

| | No frailty | | Community | Household |
|---|---|---|---|---|
| Baseline distributions | AIC | Frailty distribution | AIC | AIC |
| Exponential | 2630.088 | Gaussian | 2487.14 | 2471.722 |
| Weibull | 2566.057 | Gaussian | 2568.274 | 2696.672 |
| Log-logistic | 2462.352 | Gaussian | 2485.497 | 2461.255 |
| Lognormal | 2499.53 | Gaussian | 2508.123 | 2501.959 |

Table 4 shows results from the AIC figures of AFT-shared frailty models. We have assumed Exponential, Weibull, Log-logistic and lognormal distributions for baseline; and the Gaussian is treated as frailty distribution. The AIC values of the different AFT,and AFT models with Gaussian shared frailty model are displayed in table 4 for community and household clusters respectively. The AIC values of Log-logistic AFT, and Log-logistic AFT with Gaussian frailty model have been seen as the least compared to the rest of the models in all cases, denoting that it is the best model in terms of efficiency.

Table results of Log-logistic AFT and Gaussian shared frailty model with Log-logistic baseline distribution has been displayed in table 5, which was seen as the most efficient model for the data. Approximated coefficients, p-value, parameter estimates of baseline distributions and frailty variance have been displayed in table 5. The Log-logistic AFT-model shows that the sons who have died, daughters who have died, duration of breastfeeding and months of breastfeeding were found to be having significant influence on the under-five mortality $(p < 0.05)$. Increase in the number of sons and daughters who have died in the households reduced the risk of death. Sex of child, number of living children, and living children plus current pregnancy were found not to be significant factors for mortality.

On applying chi-square test whose null hypothesis states that variance of community frailty term is 0 $(\theta = 0)$, chi-square test statistics gave $p - value$ of 0.028. Using 5% significance level, it means there's proof of existence of unseen random factors at community cluster. This means that there are other influences that do affect mortality at community cluster that cannot be interpreted by observed covariates which are presented in the model. The origin of the unobserved random factors at community cluster can be due to limit access to food and other elements that may not be easily measured at community cluster. More focus should be given in this area as far as further research is concerned for the purposes of explaining reasons for unobserved random factors at community cluster .

In the scenario involving household frailty term, the $p-value$ is 0.28. At 5% significance level it means there's no proof of existence of unobserved random factors at household cluster. Therefore, implying that factors influencing under-five deaths in the households can be clarified just by using the covariates in the model without the inclusion of household cluster term. In this scenario one can apply Log-logistic AFT-model without frailty since the outcome proposes that there will be no difference when it comes to making conclusions which will be drawn about the data.

**Table 5. AFT model with shared frailty for the under-five mortality**

| | Loglogistic(no frailty) | | Loglogistic(G)(community) | | Loglogistic(G)(household) | |
|---|---|---|---|---|---|---|
| Parameters | Coef | P-value | Coef | P-value | Coef | P-value |
| Intercept | 4.6843 | 0.000 | 4.0527 | 0.000 | 4.3908 | 0.000 |
| Sons who have died | -1.9436 | 0.000 | -1.4136 | 0.000 | -1.6941 | 0.000 |
| Daughters who've died | -1.6536 | 0.000 | -1.2341 | 0.000 | -1.4572 | 0.000 |
| Gender of child | | | | | | |
| Male | Ref | | Ref | | Ref | |
| Female | 0.3298 | 0.132 | 0.2739 | 0.067 | 0.3015 | 0.110 |
| Number of living children | 0.4986 | 0.125 | 0.3587 | 0.110 | 0.4347 | 0.120 |
| Living children+ | 0.0516 | 0.873 | 0.0388 | 0.860 | 0.0439 | 0.870 |
| Duration of breastfeeding | 0.2275 | 0.000 | 0.1713 | 0.000 | 0.2011 | 0.000 |
| Months of breastfeeding | -0.0402 | 0.002 | -0.0317 | 0.001 | -0.0364 | 0.001 |
| Frailty | | | | 0.028 | | 0.2800 |
| Variance | | | 0.717 | | 0.337 | |

# 5 Discussion

The study of under-five child mortality has always been one of the most crucial researches in the middle income countries including Kenya because of its high rate. Good news is that Kenya has witnessed a significant decline in under-five child mortality recently KDHS (2014). In this paper, an effort has been made to determine the possible determinants of mortality for children under-5 years of age in Kenya by using random survival forests, and Accelerated Failure Time (AFT)-shared frailty models.

Random survival forests has progressively become well liked other way for analyzing time to incident data Segal and Bloch (1989). This approach has been recognized as satisfactory for analyzing survival data. Covariates such as number of living children, living children plus current pregnancy, and sex of child showed up as determinants which were crucial in describing mortality for children in Kenya. Though, these determinants were not significant for rate of mortality for the under-5 children when we used AFT model. Is more fascinating to take note that random survival forests gives extra particulars regarding importance of variables.

Our study considered two clusters, community and household levels. Gaussian shared frailty with baseline distribution as Log-logistic was used to approximate effect of risk factors on child survivorbility. The output from this model was compared to the model without frailty, and the results were that there was no presence of the unobserved heterogeneity at household clusters, meaning the model with no frailty can be used to explain the risk factors of the under-five mortality without the household frailty term. On the other hand our study found out that there was presence of unobserved heterogeneity at community clusters, this means that there are other influences that do affect mortality at community clusters which the variables alone in the model cannot interpret . Our study disagreed with Nasejje et al. (2015) which found out that there was presence of the unobserved random factors at household level. Sons who have died and daughters who have died are the determinants which are related with decreased risks of death in Kenya.

## 5.1 Conclusion

Our study found out that there was presence of unobserved heterogeneity at community clusters, this means that there are other influences that do affect mortality at community clusters which the variables alone in the model cannot interpret or explain. On the other hand there was no presence of the unobserved heterogeneity at household clusters, implying that factors influencing under-five deaths in the households can be clarified just by using the covariates in the model without the inclusion of household cluster term.

## 5.2 Future Research

Future research should be done to identify appropriate approaches of how to handle missing data without reducing the number of observations.

## 5.3 Study limitations

Data from demographic health survey are always cross-sectional, thus, vulnerable to issues to do with high degree of missingness due to respondents not being able to remember events in the past, and also some of the variables which can assist during analysis process may not get captured in the surveyNasejje et al. (2015). The high degree of missingness was noticeable in 2014 KDHS data set.

## References


Aalen, O. O. (1992). Modelling heterogeneity in survival analysis by the compound poisson distribution. *The Annals of Applied Probability*, pages 951–972.

Abeysekera, W. and Sooriyarachchi, R. (2009). Use of schoenfeld's global test to test the proportional hazards assumption in the cox proportional hazards model: an application to a clinical study.

Akaike, H. (1987). Factor analysis and aic. In *Selected Papers of Hirotugu Akaike*, pages 371–386. Springer.

Ayele, D. G., Zewotir, T. T., and Mwambi, H. (2017). Survival analysis of under-five mortality using cox and frailty models in ethiopia. *Journal of Health, Population and Nutrition*, 36(1):25.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

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

Goethals, K., Janssen, P., and Duchateau, L. (2008). Frailty models and copulas: similarities and differences. *Journal of Applied Statistics*, 35(9):1071–1079.

Grambsch, P. M. and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3):515–526.

Hougaard, P. (1986). A class of multivanate failure time distributions. *Biometrika*, 73(3):671–678.

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., and Lauer, M. S. (2008). Random survival forests. *The annals of applied statistics*, pages 841–860.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481.

KDHS (2014). Infant and under-five mortality rate. Technical report, Demographic Health Surveys.

Kleinbaum, D. G. and Klein, M. (2002). Modeling strategy for assessing interaction and confounding. *Logistic regression: A self-learning text*, pages 191–226.

Mulera.B (2017). Survival analysis of cardiovascular diseases with an extension to competing risks. Master's thesis, Minessota State University, Mankato.

Nasejje, J. B. and Mwambi, H. (2017). Application of random survival forests in understanding the determinants of under-five child mortality in uganda in the presence of covariates that satisfy the proportional and non-proportional hazards assumption. *BMC research notes*, 10(1):459.

Nasejje, J. B., Mwambi, H. G., and Achia, T. N. (2015). Understanding the determinants of under-five child mortality in uganda including the estimation of unobserved household and community effects using both frequentist and bayesian survival analysis approaches. *BMC public health*, 15(1):1003.

Njagi, P. M. (2011). Urban rural differentials of infant mortality in kenya using a shared frailty model. Master's thesis, Institute of Tropical and Infectious Diseases.

Pan, W. (2001). Using frailties in the accelerated failure time model. *Lifetime Data Analysis*, 7(1):55–64.

Segal, M. R. and Bloch, D. A. (1989). A comparison of estimated proportional hazards models and regression trees. *Statistics in medicine*, 8(5):539–550.

Statista (2017). Mortality rate of children under 5 in africa. Technical report, Statista.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., and Zeileis, A. (2008). Conditional variable importance for random forests. *BMC bioinformatics*, 9(1):307.

SWAIN, P. K. and GROVER, G. (2016). Accelerated failure time shared frailty models: Application to hiv/aids patients on anti-retroviral therapy in delhi, india. *Turkiye Klinikleri J Biostat*, 8(1):13–20.

UNDP (2009). Sustainable development goals. Technical report, UNDP.

USAID (2017). Under-five mortality in subsaharan africa. Technical report, United States Agency For International Development.

Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–454.

WHO (2017). Statistics on under-five child mortality. Technical report, World Health Organization.

Yadav, A. K. and Yadav, R. J. (2016). A study on childhood mortality using shared frailty modeling approach. *Journal of Applied Sciences*, 16(1):11–17.