EVALUATION OF THE PERFORMANCES OF ARTIFICIAL NEURAL NETWORK AND COX PROPORTIONAL HAZARD IN MODELING TIME TO FEMALE GENITAL MUTILATION IN NIGERIA

BY

OGUNTOLA, REGINA OLAIDE

211286

A DISSERTATION SUBMITTED TO THE DEPARTMENT OF EPIDEMIOLOGY AND MEDICAL STATISTICS

FACULTY OF PUBLIC HEALTH, COLLEGE OF MEDICINE,

UNIVERSITY OF IBADAN, NIGERIA

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE DEGREE (M.Sc.) IN BIOSTATISTICS.

JANUARY, 2021

DECLARATION

I the undersigned solemnly declare that the project report title ***Evaluation of the Performances of Artificial Neural Network and Cox Proportional Hazard in Modeling Time to Female Genital Mutilation in Nigeria*** is based on my own work carried out during the course of our study under the supervision of **Dr. Adeniyi Francis Fagbamigbe.**

I assert the statements made and conclusions drawn are an outcome of my research work. I further certify that

I. The work contained in the report is original and has been done by me under the general supervision of my supervisor.

II. The work has not been submitted to any other Institution for any other degree/diploma/certificate in this university or any other University of Nigeria or abroad.

III. I have followed the guidelines provided by the university in writing the report.

IV. Whenever I have used materials (data, theoretical analysis, and text) from other sources, I have given due credit to them in the text of the report and giving their details in the references.

Oguntola Regina Olaide

CERTIFICATION

This Project with the title ***Evaluation of the Performances of Artificial Neural Network and Cox Proportional Hazard in Modeling Time to Female Genital Mutilation in Nigeria*** submitted by Oguntola Regina Olaide with **Matriculation number 211286** have satisfied the regulations governing the award of Master's Degree in Epidemiology and Medical Statistics University of Ibadan, Ibadan, Nigeria.

………………………………..   ………………………

Dr. Adeniyi Francis Fagbamigbe   **Date**

Supervisor

DEDICATION

I would like to dedicate this work to GOD Almighty for giving me the opportunity to do my Masters degree which would not have been possible otherwise. This work is also dedicated to all aspiring students and future researchers.

ACKNOWLEDGEMENTS

The completion of this study could not have been possible without the expertise of Dr A.F Fagbamigbe, my amiable supervisor who has not only been a guide but also a mentor whom I could depend on for advice. On the same note, I will like to appreciate the likes of my amiable Head of Department, Dr Adebowale who is like a father to us during this period, Dr. Akpa, Dr. Akinyemi, Dr. Afolabi, Dr. Yusuf and a host of others within the Faculty of Public Health for your continuous tutelage and mentorship during this program.

I would also like to express my special thanks to my tutors Adedokun Bidemi and Osanyintupin Olawale Dele who have both been so helpful and gave me the support needed to do this Project.

A debt of gratitude is also owed to Fagbenro Olawale for his moral encouragement and financial assistance when needed.

I would like to thank my parents Mr and Mrs Oguntola, my sibling and family at large for their prayers. Without you all, none of this would indeed be possible, not to forget my colleagues and friends for always being there for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Introduction: Female genital mutilation is a crime against womanhood, posing a great health and financial burden to individuals, families and the society. This study evaluated the performances of Artificial Neural Network (ANN) and Cox Proportional Hazard (CPH) in modeling time to Female Genital Mutilation (FGM) in Nigeria.

Methods: Risk factors of FGM from birth to the age or period at which women were exposed to cutting (in years) were modeled using Kaplan Meier method (Non-parametric), Cox proportion hazard model (parametric), and neural network (parametric). Data for this study was extracted from the 2018 Nigeria Demographic Health Survey (NDHS) using multistage stratified random sampling of households of women of ages 15-49. The outcome of interest was whether a respondent has undergone FGM or not as of survey time. The risk factors that were considered in this study were education, religion, residence, ethnicity, daughter circumcision, region, religious belief, opinions about FGM and wealth index.

Results: The findings from the ANN model suggested that type of residence, level of education, opinion of the women on FGM and ethnicity were the most important predictors of FGM. The prevalence of FGM among women with primary and tertiary (higher) qualification was 38.42 and 26.62 respectively. Yoruba ethnic group had the highest prevalence with 51.31. The prevalence rate of FGM among women in urban residence (35.61) was higher compared to rural residence. Southern region had the highest prevalence rate with south east and south west having 47.07 and 46.1 prevalence rate respectively.

Women from urban residence had 13.46% higher hazard of being circumcised compared to women from rural residence. Women with secondary education had 18.66% higher hazard of being circumcised compare to women with tertiary education while women with primary and no education had 73.23% and 50.72% lower hazard of being circumcised compared to women with tertiary education. The comparison between the CPH and ANN model suggested that CPH model was better in terms of classification ability than ANN because it has the higher AUC value of 0.1908 compared to 0.0989.

Conclusion: Artificial Neural Network and Cox Proportional Hazard models are both appropriate for predicting time to FGM and they are recommended for predicting and determining the influence of risk factors on the time to female genital mutilation.

Keywords: Artificial Neural Network, Cox Proportional Hazard Model, Kaplan Meier, Prevalence, Female Genital Mutilation

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

A time-to-event analysis is of fundamental importance in many fields where there is interest in the time until the occurrence of particular events allowing for time-dependent missing outcomes (i.e. "censored" data) (Morris *et al*., 2012). The main characteristic of survival time is that it is censored due to the end of study or withdrawal during the period of study because we cannot follow-up the exact survival time for those who are still alive at the end of the study, or who are lost to follow-up during the period of study.

However, it is known that the survival time of censored individuals is at least longer than the censoring time. There are many different types of censoring such as right-, left-, and interval-censoring. The most popular censoring mechanisms are right censoring, in which the lower limit of the exact survival time is observed, while the upper limit is observed for left censoring, and both lower and upper limits are used for interval-censoring.

Many statistical methods have been developed for estimating survival functions, comparing survival curves between two groups, and the survival data by regression, for association with risk factors, such as demographic and clinical predictors. In survival analysis, nonparametric statistical inference is more extensively used to estimate the survival function, and compare survival curves between two or more groups. Kaplan-Meier (KM) estimators for a survivor function and log-rank test for comparison of survivor functions are derived by a nonparametric approach.

However, if the appropriate distribution for survival data is assumed or pre-specified, the parametric approach is more appropriate. When the association of survival time with various risk factors is the main interest, the most popular model is a Cox regression based on a semi-parametric approach, since the effect of predictors on the hazard rate is parametrically specified, while the baseline hazard function is unspecified. A variety of parametric approaches are also available under the assumed survival distributions, such as an accelerated failure time (AFT) model. Overall, all survival analysis approaches should take into account a censoring mechanism when a statistical inference is made (Lee & Lim, 2019; Uchenna Mberu, 2017).

The Cox proportional hazards (PH) model is traditionally used to predict the clinical outcomes or hazard functions corresponding to specific time units. However, this model has the following major drawbacks: The proportional hazard assumption and linearity of each variable must be satisfied (Babińska *et al*., 2015). These assumptions are difficult to be satisfied using real-world data, and their violation may lead to the creation of a false model. To solve these problems in classical survival analysis, several neural network-based hazard functions and overall survival time prediction models have been developed.

In this study, Cox Proportional Hazards and the Artificial Neural Network (ANN) will be used to model time to Female genital mutilation (FGM). Female genital mutilation (FGM), also known as female genital cutting (FGC) or female circumcision, is defined by the World Health Organization (WHO) as any procedure that involves partial or total removal of the external genitalia and/or injury to the female genital organs whether for cultural or any other non-therapeutic reasons (WHO, 1997). FGM is widely recognized as a human rights violation and is deeply rooted in attitudes and expectations over decades and centuries.

In May 2015, Nigeria's Federal Government passed the Violence Against Persons Prohibition Act 2015 (VAPP), a law prohibiting FGM and other dangerous traditional practices; however, this regulation only extended to Abuja's Federal Capital Territory (FCT). Although the law has been enacted by three more states, Anambra, Ekiti and Oyo, but there is a need for all states to pass similar legislation (*Nigeria28 Too Many*, n.d.).

WHO classifies female genital mutilation into Type I, II, III, and IV. Type I deals with the excision of the prepuce with or without excision of part or all of the clitoris, type II deals with the excision of the clitoris with partial or total excision of the labia minora, type III deals with the excision of part or all of the external genitalia and stitching or narrowing of the vaginal opening (infibulation) while type IV deals with other forms, including pricking, piercing, or incising of the clitoris and/or labia; stretching of the clitoris and/or labia; cauterization by burning of the clitoris and surrounding tissue; scraping of tissue surrounding the opening of the vagina (angurya cuts) or cutting of the vagina (gishiri cuts); and introduction of corrosive substances or herbs into the vagina to cause bleeding or to tighten or narrow the vagina.

It is estimated that 3 million girls in Africa are at risk of FGM annually and more worrisome is that the prevalence remains alarmingly high; in Nigeria, the prevalence ranges from 25- 53.2% with equally high rates (31.3-36.2%) reported among pregnant women(National Population Commission(NPC)[Nigeria] & ICF International, 2019)Although, the national prevalence of FGM in Nigeria stands at 25% on average among adult women, varying prevalence of 49% (South East), 47.5% (South West), 25.8% (South South), 20.7% (North West), 9.9% (North Central) and 2.9% (North East) are also reported across the respective geopolitical zones of the country(National Population Commission(NPC)[Nigeria] & ICF International, 2019).

A 2013 study shows that 19.9million girls and women have undergone FGM but only 62% think it should end. Social acceptance is the most frequently cited reason for supporting the continuation of the practice, the chances that a girl would undergo FGM increases significantly if her mother has undergone FGM. This study aims to compare the performances of Cox Proportional Hazards model and the Artificial Neural Network (ANN) in the estimation of the time to FGM.

1.1.1 Survival Analysis

Duration or survival data cannot generally be analyzed by conventional methods such as linear regression, primarily because some duration (e.g. current age of the girl at circumcision) are usually right-censored.

Another reason why conventional linear regression is not appropriate is that survival times tend to have positively skewed distributions, for instance, most FGM events may occur close to birth (left negatively skewed) or close to age 14 (right positively skewed) (Faraggi & Simon, 1995; Uchenna Mberu, 2017).

Cox's proportional hazards (Cox's PH model) is one of the most widely applied models in survival modeling with a hazard ratio of a group assumed to be constant to a baseline group through the observed time. Violation of this assumption required application of techniques like: stratified model, combining variable subgroups, extending model with time-dependent variables, and piecewise model. Those techniques do not guarantee tenability of Cox's proportional approach in some situations since the determination of some variables must be based on prior information of the phenomena under study. This makes Cox's PH approach unreliable for multivariate data which requires correct functional relationship.

Artificial Neural Network (ANN) will be considered to address these limitations since it requires little assumptions and can be used to directly predict the survival times (Karmaker *et al*., 2011).

1.1.2 Artificial Neural Network (ANN)

The human brain is imitated by an artificial neural network (ANN). A human brain is capable of learning new things, adapting to new and changing environments. The brain has the most incredi ble ability to analyze incomplete and unclear, fuzzy information and make out of it its judgment.

An Artificial Neural Network (ANN) is a mathematical model that artificially simulates the functionalities and structures of biological human neural networks. Its main feature is the artificial neurons

The ANN architecture comprises of:

a. Input layer: Receives the input values

b. Hidden layer(s): A set of neurons between the input and output layers. There can be single or multiple layers

c. Output layer: Usually it has one neuron, and its output ranges between 0 and 1, that is, greater than 0 and less than 1. But multiple outputs can also be present

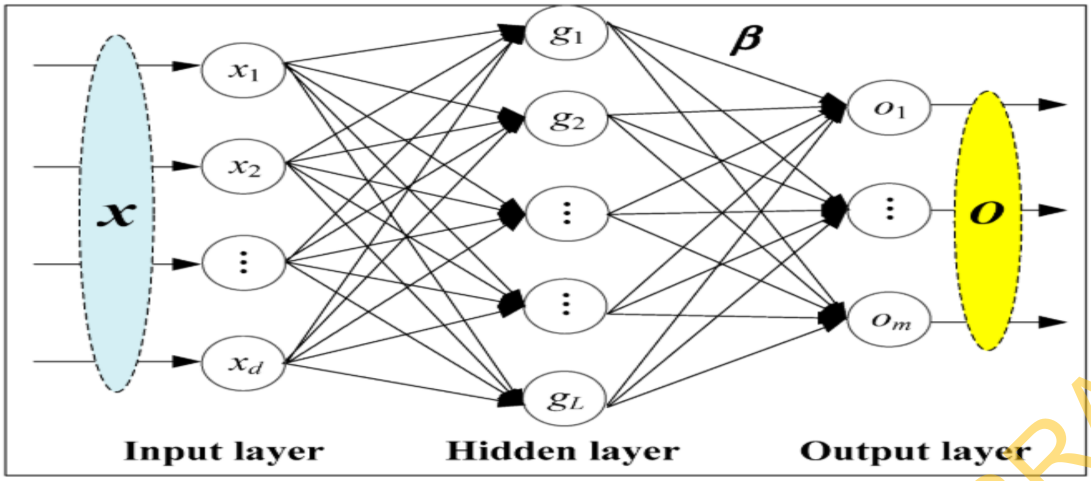

Figure 1.1: Neural Network with one hidden Layer

An ANN **hidden layer** is a layer or set of layers in between input layers and output layers, where artificial neurons take in a set of weighted inputs and through an activation function produces an output.

The processing ability is stored in inter-unit connection strengths, called weights. Input strength depends on the weight value. Weight value can be positive, negative or zero. Negative weight means that the signal is reduced or inhibited. Zero weight means that there is no connection between the two neurons. The weights are adjusted to obtain the required output.

ANN has been successfully used for pattern recognition and survival prediction in several settings. The advantage of a neural network is the ability of the model to capture nonlinearities and complex interactions between factors. Trained on a number of prognostic factors, neural networks have been reported to improve the accuracy of survival prediction.

1.2 Problem Statement

Often, Cox PH model assumptions are violated in the analysis of survival data. The model suggests that each variable has a linear contribution to the model, but often the relationship can be more complicated. Besides, the proportionality of the hazards are often violated. The big concern with statistical modelling in Public Health issues even with the application of some techniques (Cox PH model in survival analysis) to address the violation of assumptions, may still be unreliable. It is, therefore, reasonable to compare the flexibility and robustness of ANN over Cox PH model.

ANN may be an alternative effective methods of analyzing complex data with non-linear covariates, high order interaction among covariates, and time-dependent covariates than the traditional Cox PH model and regression methods (Yang, 2011). ANN offers a more flexible prediction of survival time than traditional methods. ANN model is not without drawbacks. The primary disadvantage of an ANN is its "black box" quality, that is, without extra effort, it is difficult if not impossible to gain insight into a problem based on an ANN model. Regression techniques, for example, allow the user to sequentially eliminate possible explanatory variables that do not contribute to the fit of the model. Similarly, based on the underlying statistical theory, regression techniques allow hypothesis testing regarding both the univariate and multivariate association between each explanatory variable and the outcome of interest. These features are not available for ANN.

1.3 Aim and Objectives of the Study

The aim of this study is to compare the performances of ANN and Cox's proportional hazard models in modeling time to female genital mutilation among girls and women in Nigeria.

The specific objectives of this study are to:

i. Build robust and reliable models for evaluation of time to female genital mutilation among women and girls in Nigeria

ii. Compare the performances CPH and ANN in modeling time to female genital mutilation among women and girls in Nigeria

iii. Identify the prognostic factors of time to FGM among women and girls in Nigeria.

1.4 Justification of the Study

The primary justification is the tendencies of researchers to use CPH for survival analysis without checking its assumption violations, its incapability to handle complex data, hence the need to explore alternative models.

WHO research has found that women who had FGM are slightly more likely to face complications during childbirth, their babies are more likely to die as a result of the procedure. Serious risks during childbirth include the requirement for a caesarean operation, dangerous heavy bleeding, and prolonged hospitalization. FGM has immediate and long-term medical implications.

Morris *et al*. 2012 reveal how circumcision can result in health and sexual problems but revealed that the risk will be minimal if it was not done before the usual circumcision age of 11 years and above. This was due to a change in beliefs on sexuality, marriage, religion and signs of social assimilation. There is a need to model the time to FGM with the most appropriate models.

This study is significant in a number of ways. It adds to existing knowledge and serves as a source of reference material on modeling time to FGM. The Ministry of Health (MOH) may find it useful and beneficial. Information on factors affecting time to FGM will not only help in health facilities but also helps the government to adopt a better and effective policy on how to stop or reduce FGM in the country. Findings of the research will inform health planners to enable them to set up genuine planning schemes for health care services.

CHAPTER TWO

LITERATURE REVIEW

2.1 Comparison of Artificial Neural Network and Cox Regression Model

Sabouri *et al*. determined factors related to Colorectal Cancer (CRC) patients' survival using ANN and Cox regression (Sabouri *et al*., 2020; Uchenna Mberu, 2017). In their historical cohort, information of patients who were diagnosed with CRC in Omid Hospital of Mashhad were collected. A total of 157 subjects were investigated from 2006 to 2011 and were followed up until 2016. In ANN, data were divided into two groups of training and testing, and the best neural network architecture was determined based on the area under the ROC curve (AUC). Cox regression model was also fitted and the accuracy of these two models in survival prediction was compared by AUC. The study showed that the mean and standard deviation of the age was 56.4 ± 14.6 years. The three-, five- and seven-year survival rates of patients were 0.67, 0.62, and 0.58, respectively. Using test dataset, the area under curve was estimated 0.759 for the chosen model in ANN and 0.544 for Cox regression model. They concluded that ANN was an appropriate approach for predicting CRC patients' survival which was superior to Cox regression. Thus, it was recommended for predicting and also determining the influence of risk factors on patients' survival.

Sadiq *et al*. conducted a research to extend the ANN model to Relative Survival and compare the predictive accuracy to that of Cox(Sadiq *et al*., 2016). Two data sets were used for illustration, real-life dataset and artificial dataset. The real-life dataset was collected in the in the study carried out at the University Clinical Centre in Ljublana and contains 1040 Acute Myocardial Infarction patients diagnosed between 1982 and 1986 and followed up until 1997. The data were randomly divided into two, training and validation groups. The population data was the

Slovenian population Census table since the study was carried out in Slovenia. Then, Cox and three layers Artificial Neural Network model with backpropagation algorithm were used to analyze the data. To compare the prediction of both models, the Area under the Receiver Operating Characteristics (ROC) Curve (AUC) was used. The artificial dataset comprises of three different simulation schemes, based on Monte Carlo simulation. Varying sample sizes of 50, 200 and 1000 random observations were generated for each schema. To assess the accuracy of predictions, part of the data set was allocated to the training group and the remainder allocated to the testingvalidation. Three datas*et al*locations were used 80:20, 90:10 and 95:5. The performance function for the assessment were the misclassification error rates (MER), sensitivity (SEN) and specificity (SPE). It was deduced that ANN model had more predictive capability than the Cox model and prediction ability increases with decrease in percentage holdout.

Khosravi *et al*. attempted to model the survival of patients (two years old and above) after liver transplantation using neural network and Cox Proportional Hazards (Cox PH) regression models (Khosravi *et al*., 2015). The event was defined as death due to complications of liver transplantation. They carried out a historical cohort study where the clinical findings of 1168 patients who underwent liver transplant surgery (from March 2008 to march 2013) at Shiraz Namazee Hospital Organ Transplantation Center, Shiraz, Southern Iran, were used. They model the one to five years survival of such patients using Cox PH regression model accompanied by three layers feed forward artificial neural network (ANN) method on the data separately and their prediction accuracy was compared using the area under the receiver operating characteristic curve (ROC). They also used Kaplan-Meier method to estimate the survival probabilities in different years. The result of their study showed that the estimated survival probability of one to five years for the patients were 91%, 89%, 85%, 84%, and 83%, respectively. The areas under the ROC were 86.4% and 80.7% for ANN and Cox PH models, respectively. In addition, the accuracy of prediction rate for ANN and Cox PH methods was equally 92.73%.

Zhu *et al*. carried out a study on the comparison between artificial neural network and cox regression model in predicting the survival rate of gastric cancer patients (ZHU *et al*., 2013). In their study, they determine the prognostic factors and their significance in gastric cancer (GC) patients, using the artificial neural network (ANN) and Cox regression hazard (CPH) models. A retrospective analysis was undertaken, including 289 patients with GC who had undergone gastrectomy between 2006 and 2007. According to the CPH analysis, disease stage, peritoneal dissemination, radical surgery and body mass index (BMI) were selected as the significant variables. According to the ANN model, disease stage, radical surgery, serum CA19-9 levels, peritoneal dissemination and BMI were selected as the significant variables. The true prediction of the ANN was 85.3% and of the CPH model 81.9%. They concluded that the ANN model was a more powerful tool in determining the significant prognostic variables for GC patients, compared to the CPH model.

Hosseini *et al.*, compared Cox regression and Artificial Neural Network (ANN) models to predict survival in acute leukemia patients (Uchenna Mberu, 2017)(Hosseini Teshnizi S., Tzazhibi M., 2013). The information on 197 patients with acute leukemia in Sayyed-O-Shohada Hospital was collected using a checklist. Cox regression model was fitted to the observations. They selected an efficient ANN to compare with Cox regression model, the number of hidden layer neurons was changed. The prediction accuracy of the two models were compared using receiver operating characteristic (ROC) curve and kappa. The result of their analysis suggested that out of 9 ANN models with one hidden layer and 4 to 12 neurons, an ANN with 5 neurons in hidden layer was a superior model compared with Cox regression model. The areas under ROC curve for ANN model and Cox model were estimated to be 0.0709 and 0.458, respectively. The accuracies of prediction of survival for ANN model and Cox model were estimated as 78.9% and 50.3%, respectively. They concluded that due to the high predicting accuracy of ANN models, the use of different models of ANN and their development in various fields of medical science are recommended.

Amir *et al*. conducted a research to compare Cox regression and artificial neural network models in prediction of kidney transplant survival(Amir *et al.,* 2013). The present multi-center retrospective study was conducted on the medical records of 756 kidney transplant recipients undergoing kidney operations at two treatment centers from 2001 through 2012. The data were randomly divided into two educational and experimental (validation) groups. Then, Kaplan-Meier, Cox proportional hazard, and three-layer artificial neural network models were used for analyzing the data. They compared the prediction of both models by applying the area under the curve in the characteristic function. Post-operative creatinine and relative family are among the factors of influencing kidney transplant survival. Moreover, the survival estimates of the transplanted kidney for periods of six months, one year, three years, and five years were 89, 87.4, 80, and 75 percent, respectively. ROC areas under the curve, for multi-layer perceptron neural network model and Cox regression, were 81.3% and 71%, respectively.

Puddu and Menotti compared Cox versus Neural Network models in predicting 45-year all-cause mortality (45-ACM) by 18 risk factors selected a priori: age; father life status; mother life status; family history of cardiovascular diseases; job-related physical activity; cigarette smoking; body mass index (linear and quadratic terms); arm circumference; mean blood pressure; heart rate; forced expiratory volume; serum cholesterol; corneal arcus; diagnoses of cardiovascular diseases, cancer and diabetes; minor ECG abnormalities at rest(Puddu & Menotti, 2012). Two Italian rural cohorts of the Seven Countries Study, made up of men aged 40 to 59 years, enrolled and first examined in 1960 in Italy. Cox models were estimated by forcing all factors; a forward-stepwise; and a backward-stepwise procedure. Observed cases of deaths and of survivors were computed in decile classes of estimated risk. Forced and stepwise NN were run and compared by C-statistics (ROC analysis) with the Cox models. Out of 1591 men, 1447 died. Model global accuracies were extremely high by all methods (ROCs > 0.810) but there was no clear-cut superiority of any model to predict 45-ACM. The highest ROCs (> 0.838) were observed by NN. There were inter-model variations to select predictive covariates: whereas all models concurred to define the role of 10 covariates (mainly cardiovascular risk factors), family history, heart rate and minor ECG abnormalities were not contributors by Cox models but were so by forced NN. Forced expiratory volume and arm circumference (two protectors), were not selected by stepwise NN but were so by the Cox models (Puddu & Menotti, 2012).

2.2 Practice of Female Genital Mutilation

Female Circumcision, sometimes also referred to as female genital mutilation has attracted increasing global attentions for more than 40 years. Governmental and nongovernmental organizations (NGOs) such as world health organization (WHO), United Nations Children's Fund (UNICEF), and the United Nations Population Fund (UNFPA) have took advantage of several United Nations world conferences to establish a strong global consensus against the practices of Female genital mutilation. Human rights of women and children in recent years also. In recent years raised additional calls for the practice to be stopped. More than 200 million girls are living with the results and practice of Female Genital Mutilation and 30 million more are at the risk of being mutilated in the next decade; according to a report by the United Nations International Children Fund (UNICEF, 2013).

Although FGM is considered as an old and unsafe cultural and traditional practice in most part of the world, it is still being practiced in over 28 countries in Africa and few scattered communities in Asia. The prevalent is mostly in Nigeria, Egypt, Mali, Eritrea, Sudan, Central African

Republic, and northern part of Ghana. Somalia and Djibouti have the highest prevalence rates world-wide(Okeke *et al*., 2012).

FGM is widespread in Africa with differing prevalence in different countries, although the incidence is decreasing because it is considered a human right problem with tremendous support for its removal by mainly nongovernmental organizations. In many African countries, it is mostly underreported particularly where it has been declared illegal. FGM is often carried out by a non-medical practitioner with the intention of performing religious or cultural rituals and at times, with the subsequent acute, intermediate and late complications for economic benefits. It is sometimes achieved by medical practitioners when it is genuinely assumed that the risks involved with the profession are minimized by its medicalization. Compared to male circumcision and voluntary modifications to female external genitals such as cutting and tattooing, FGM/C vulnerability is intensified as practices. The magnitude of the physical and psychosocial consequences of FGM outweighs the presumed benefits of the procedures highlighting the need for all stakeholders and in all sectors to improve the multiple preventive measures (Odukogbe *et al*., 2017).

2.3 Modeling of Female Genital Mutilation

Imran *et al*. carried out a study on the economic status, a salient motivator for medicalization of FGM in sub-Saharan Africa: Myth or reality from 13 national demographic health surveys (Morhason-Bello *et al*., 2020). In their study, they hypothesized that the risk of mediatized FGM by girls/women was associated with socioeconomic status (SES) their household belongs. They used 2010–2019 Demographic and Health surveys data from 13 countries in SSA. They analyzed information on 214,707 women (Level 1) nested within 7299 neighborhoods (Level 2) from the 13 countries (Level 3). They fitted 5 multivariable binomial multilevel logistic regression models using the MLWin 3.03 module in Stata. The estimation algorithms that was adopted in their study was the first order marginal quasi-likelihood linearization using the iterative generalized least squares. The study showed that the odds of FGM medicalization increased with the wealth status of the household of the woman, with 29%, 45%- and 75%-times higher odds in the middle, richer and richest household wealth quintiles, respectively than those from the poorest households ($p < 0.05$). The more educated a woman and the better a woman's community SES was, the higher her odds of reporting medicalization of FGM. Rural community was associated with higher odds of mediatized FGM than urban settings. They concluded that mediatized FGM was common among women from a high socioeconomic, educational background and rural settings of SSA.

Acar, identified identify how parameters such region, religion, educational level and occupation of the parents affect the practice of FGM(Acar, 2019). Data used for his study was sourced from the Egypt Demographic and Health surveys from 2005 to 2014, whereby ever-married women were asked about the FGM status of their daughters. The dataset consisted of both right-censored and interval-censored observations. First, the survival function was estimated through nonparametric maximum likelihood estimation. Next, we fitted both semi-parametric Proportional Hazards (PH) and parametric Accelerated Failure Time (AFT) models to evaluate the effect of the parameters of interest.

In addition, we assessed the extent of bias that would be elicited if the event times were imputed rather than using the actual interval-censored event times. Finally, a mixture cure model was fit to account for the presence of a population of girls insusceptible to FGM. The result from his study was independently of the type of model that was fitted, all the covariates of interest were shown to be significantly associated with FGM. The results showed a remarkable increase in the hazard of FGM for girls from Egypt. In contrast, highly educated or Christian mothers seemed to substantially decrease the hazard of FGM. Right-point imputation of the event times showed to cause the least bias in the nonparametric estimator of the survival function as well as in the estimated coefficients of the semi-parametric PH model, while midpoint imputation seemed preferable for the parametric AFT model.

An AFT mixture cure model that assumes a log-logistic distribution for the latency part and a complementary log-log function for the incidence part were found to best fit the data. The educational level of the mother was significantly associated with the proportion of girls who are insusceptible, but does not affect survival for girls who are susceptible. In contrast, occupation of the mother was not significant in modeling the probability of susceptibility, but was significant in the latency part of model. All other covariates were retained in both parts of the model. He concluded that region, religion, educational level and occupation of the parents are important factors in modeling the survival function of FGM. When imputed event times were used instead of the interval-censored event times, the extent of bias seemed to be rather limited and dependent of the type of imputation as well as the type of model that was fitted. However, he recommended that further analyses are necessary to better assess the adequacy of the models that were fitted and to address other potential sources of bias, especially the presence of left truncation.

Ngianga-Bakwin *et al.*, investigated the roles of normative influences and related risk factors (e.g., geographic location) on the persistence of FGM among 0–14 years old girls in Kenya (Ngianga-Bakwin Kandala, Chibuzor Christopher, 2019). Their key objective was to identify and map hotspots (high risk regions). They fitted spatial and spatio-temporal models in a Bayesian hierarchical regression framework on two datasets extracted from successive Kenya Demographic and Health Surveys (KDHS) from 1998 to 2014. The models were implemented in R statistical software using Markov Chain Monte Carlo (MCMC) techniques for parameters estimation, while model fit and assessment employed deviance information criterion (DIC) and effective sample size (ESS). The results from their study showed that daughters of cut women were highly likely to be cut. Also, the likelihood of a girl being cut increased with the proportion of women in the community who were cut, who supported FGM continuation, and who believed FGM was a religious obligation. Other key risk factors included living in the northeastern region; belonging to the Kisii or Somali ethnic groups and being of Muslim background. Their findings offered a clearer picture of the dynamics of FGM in Kenya and will aid targeted interventions through bespoke policymaking and implementations.

Gayawan *et al*, aimed to explain the spatial variations in the risk factors for female cutting in Nigeria. Data used for their study were sourced from the geo-referenced 2013 Nigeria Demographic and Health Survey(Gayawan & Lateef, 2019). The objective of their study was to estimate the spatially varying relationships of the factors influencing women's attitude towards female cutting in Nigeria and to identify how the variables exact influence across the states using geographically weighted logistic regression analysis, a technique that allows for spatially varying relationships among variables to be established. The results showed that women's higher educational level and higher household wealth lowered the desire for continuation of FGM everywhere in the country, but the effects of most other variables varied in direction, strength and magnitude. The findings suggested that the use of local approaches to address the factors that encourage the continuation of female cutting in Nigeria.

Osuorah determined the socio-demographic factors that predict circumcision of the girl child. Multivariate logistic regression was used to determine predictors of girl child circumcision(Osuorah, 2015). The result of their study showed that the overall prevalence of the girl child circumcision in Nigeria is 23.9%. When stratified by region, the prevalence was 36.8%, 15.6% and 9.1% in the North-west, North-central and North-east region respectively and

30.8%, 22.8% and 8.0% in the South-west, South-east and South-south region respectively. Within states, the prevalence of female circumcision ranged from 0% in Katsina to 57.9% in Jigawa state.

Prevalence was almost twice among Muslim households than other religions. Majority (91.8%) of the girl child circumcision was done before the first birthday (0-11 months) and 84.6% was performed by a traditional circumciser. Type I (44.1%) and II (47.9%) were the commonest type of genital mutilation performed. In the final regression model, only mothers opinion about female circumcision, whether she was circumcised and region of residence significantly determined the likelihood of a girl child been circumcised. They concluded that the appropriate and targeted education of mothers and circumcisers focusing on the undesirable effect of female circumcision will significantly reduce the practice of female genital mutilation across Nigeria.

Achia carried out a study to detect identify geographical area where practice of FGM/cutting continues to increase and where serious intervention strategies were needed. 2008 Kenya Demographic and Health Survey (KDHS) data with multistage stratified random sampling plan to select women of reproductive age (15–49 years)(Achia, 2014). Questions about their FGM status and their support for the continuation of FGM were asked. Significances of clustering of FGM practice in Kenya was tested using a spatial scan statistical analysis, while the risk were mapped and modelled using hierarchical spatial model under the Integrated Nested Laplace approximation approach using the INLA library in R.

Hierarchical spatial models with spatially structured random effects provided the best fit to data for the considered response variables (Deviance Information Criterion (DIC)). According to the study, FGM has prevalence rate of 28% at both the North-Eastern and South-Western regions of the country ($p < 0.001$) and about 10% of the study participant reveals interest in continuation of

FGM. Education, region, marital status, age, socioeconomic status, rural–urban classification, religion, and media exposure were significantly associated with FGM. Current FGM/C status of a woman was also found to be a significant predictor of support for the continuation of practice of FGM/C. In the bid to eradicate the practice of FGM, Achia concluded that deep religious and cultural beliefs need to be addressed and appropriate government intervention must also be put in place(Achia, 2014).

Karmaker *et al*. carried out a study on the economic and social factors related with female genital mutilation in Burkina Faso using 12,049 Demographic Health Survey (DHS) women data representing underlying populations of the different regions within the country (Uchenna Mberu, 2017)(Karmaker *et al*., 2011). The data was gotten using multistage stratified random sampling of households of women of ages 15 – 49. The study uses t-test and C2, and Mann Whitney U-Test to explore potential prevalence and risk factors of FGM, while Logistics regression was used to model association between demographic, social, and economic risk factors with FGM. The primary aim of the study was to examine the factors associated with 'woman having gone' through FGM and 'one or more of the daughters having FGM. The univariate analysis of the study reveals significant relationship between the two outcomes and some factors such as: age, religion, wealth, ethnicity, literacy, years of education, household affluence, region and who had responsibility for health care decisions in the household had (RHCD) with p-value < 0.01. The multivariate version of the study stratified by religion confirmed that factors like education were associated with a reduced likelihood of FGM (only significant for Christian women.). They concluded that ethnic group, religions and high level of education can reduce FGM rate in younger women.

CHAPTER THREE

RESEARCH METHODOLOGY

This chapter intends to explain the methodology that would be used to achieve the objective of this research. The study design, study setting and sampling technique would be outlined, followed by a description of measurement tool used.

3.1 Study Setting

The study is carried out in Nigeria. Nigeria is the most populous country in Africa and it lies on the west coast of Africa with six geo-political regions which are: North Central, North East, North West, South East, South South and South West.

3.2 Data Collection

Data set for this study is extracted from the 2018 Nigeria Demographic Health Survey (NDHS). The 2018 NDHS collected information on fertility, awareness and use of family planning strategies, breastfeeding activities, nutritional status of women and children, maternal and child health, mortality of adults and children, empowerment of women, domestic violence, genital reduction of women, malaria prevalence, awareness and actions regarding HIV / AIDS and other sexually transmitted infections disability, and other health-related issues such as smoking.

3.3 Study Design

This quantitative cross-sectional study was conducted to evaluate the performances of artificial neural network and cox proportional hazard in modeling time to female genital mutilation in Nigeria.

According to Creswell (2003), a quantitative design provides a sample of a population in order to numerically describe the population's attitudes, patterns and/or opinions(Ishtiaq, 2019). Surveys provide an inexpensive, quick, efficient, and accurate means to evaluate information about a population. This research design was appropriate for obtaining relevant information.

3.4 Sample Design

The 2018 NDHS study was a stratified survey, selected in two phases. Stratification was accomplished by division into urban and rural areas of each of the 36 states and the Federal Capital Territory. In total, 74 strata were identified for sampling. The samples were chosen separately through a two-stage filtering in each stratum. The 2018 NDHS sample was a two-stage stratified cluster sample selected from the sampling frame, sampling weights were calculated based on sampling probabilities separately for each sampling stage and for each cluster.

3.5 Study Population

The data for the study was gotten using multistage stratified random sampling of households of women of ages 15-49. Population at risk is all women of ages 15 to 49. The outcome of interest is whether a respondent has undergone FGM or not as of survey time.

3.6 Data Analysis

3.6.1 Censoring

Censoring in a study is when there is incomplete information about a study participant, observation or value of a measurement in clinical trials, its when the event does not happen while the subject is being monitored or because they drop out of the trial.

An event time is censored if there is some amount of follow-up on a subject, but the event is not observed because of loss-to-follow-up, death from a cause other than the trial endpoint, study termination, and other reasons unrelated to the endpoint of interest.

3.6.2 Types of Censoring applicable in the Study

The dataset contains both right-censored and independent, non-informative censorships. A data is right censored if the time for the case is not identified before the subject left the study or before the investigation ended and thus $X = \min(T, C_r)$, where T represents the time of events and C_r represents the right censoring time. In addition to X, usually a failure indicator δ is also observed which equals 1 if the event observed during the study $(T \leq C_r)$ and 0 in case of right censoring $(T > C_r)$.

In our sample, the collection of right-censored observations comprises all women who did not undergo FGM before the survey. There is interval censorship when the exact time of the occurrence is not detected but is known to fall within a specified interval: $T \in (L, R)$ where L represents the lower limit and R the upper limit of the interval.

In this research, women who have been circumcised are censored at intervals as no specific event time has been established but only the age in years. To use the calculation of maximum likelihood, all present censoring forms should be included in the likelihood function. Therefore, the likelihood function can be written as:

$$L(\Theta; x, \delta) = \prod_{i \in R} S(C_r) \prod_{i \in I} [S(L_i) - S(R_i)] \quad (4.1)$$

Where R is the set of right-censored observations and I is the set of interval-censored observations.

Right censorship is well applied in standard statistical software and is often appropriately presented. Instead of carefully analyzing data at intervals, single imputation techniques such as right-point and mid-point imputation are commonly applied and accompanied by classical techniques to model survival time.

Truncation is a second feature which occurred in our data on survival. Left truncation happens when a case is observed only if the event was not encountered before a given time

Girls who entered the study would have had to survive at least until the time of the test and joined the study at different ages. Therefore, the participation of each respondent is conditional upon being alive at the time of the survey and the likelihood function should be modified to:

$$L(\Theta; x, \delta) = \frac{\prod_{i\in R} S(C_r) \prod_{i\in I}[S(L_i) - S(R_i)]}{\prod_{i=1}^{n} P(T_{death} > a)} \quad (4.2)$$

Where T_{death} represent the time of death and a the time of survey.

3.6.3 Variables considered in the Study

For this study, Women aged 15 to 49 years were analyzed. The response variable in this study is time to circumcision of women which is measured in years. It is measured as the length of time from birth to age at which the women have been circumcised. The mothers gave information of their FGM status, whether or not they have been circumcised, also at what age are they circumcised (in years). For women that did not undergo FGM, the value is missing. The possible covariates for FGM includes education (which is categorize into no education, primary, secondary and higher), religion (which is categorize into Christian, Islam, traditional and other), residence (which is categorize into urban and rural), ethnicity (which is categorize into Yoruba, Igbo, Hausa and others), daughter circumcision (which is categorize into No and Yes), region (which is categorize into North east, North west, North central, South east, South south and South west), religious belief (which is categorize into No and Yes), opinions about FGM from the respondents (which is categorize into continued, stopped, depends and don't know) and wealth index (which is categorize into poorest, poorer, middle, richer and richest).

3.7 Statistical Analysis

In this study, survival analysis was carried out on the data to model time to FGM. The Kaplan-Meier approach was used to describe the survival functions of the circumcised women and the Log-rank tests was used to compare the survival curves among groups. The time to FGM was modelled by using Cox Proportional Hazard Model and Artificial Neural Network model.

3.7.1 The Kaplan Meier Product Limit Method

In Kaplan Meier product limit method, survival probabilities can been obtained as:

$$\hat{S} = \prod_{k}^{j=1} \left(\frac{n_j - d_j}{n_j}\right), k \le n, t_j \le t < t_{j+1} \qquad (4.3)$$

Where; d_j = the number of failure in t_j, n_j is the number of incident cases at risk in t_j , k is the number of sequential observations, n is the total number of incident cases.

3.7.2 The Log Rank Test

The log rank test is a hypothesis test to compare the survival distributions of two samples. It is appropriate to use when the data are right skewed and censored.

Hypothesis:

H_0: No difference among survival curves

H_1: There are differences among survival curves

The log rank statistic for two groups is

$$\frac{(O_2 - E_2)^2}{\text{var}(O_2 - E_2)} \sim \chi^2_{G-1} \qquad (4.4)$$

The expected frequency is calculated as

Eij = proportion in risk set × failures over both groups

$$E_{1j} = \left(\frac{n_{1j}}{n_{1j}+n_{2j}}\right) * (m_{1j}+m_{2j}) \quad (4.5)$$

$$E_{2j} = \left(\frac{n_{2j}}{n_{1j}+n_{2j}}\right) * (m_{1j}+m_{2j}) \quad (4.6)$$

3.7.3 Cox Proportional Hazard (PH) Model

The Cox Proportional Hazards model is given as

$$h(t \mid x) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p) = h_0(t)\exp(\beta' x) \qquad (4.7)$$

where $h_0(t)$ is called the baseline hazard function which is the hazard function for an individual for whom all the variables included in the model are zero., $x = (x_1, x_2, \ldots, x_p)'$ is the value of the vector of explanatory variables for a particular individual, and $\beta' = (\beta_1, \beta_2, \ldots, \beta_p)$ is a vector of regression coefficients.

The corresponding survival functions are related as follows

$$S(t \mid x) = S_0(t)\exp\left(\sum_{i=1}^{p} \beta_i x_i\right) \qquad (4.8)$$

This model, also known as the Cox regression model, makes no assumptions about the form of $h_0(t)$ (non-parametric part of model) but assumes parametric form for the effect of the predictors on the hazard (parametric part of model). The model is therefore referred to as a semi-parametric model.

The hazard ratio of two individuals with different covariates x and x^* is

$$\hat{HR} = \frac{h(t)\exp(\hat{\beta}' x)}{\hat{\beta}' x^*} = \exp\left(\sum \beta' (\hat{x} - x^*)\right) \qquad (4.9)$$

This hazard ratio is time-independent, which is why this is called the proportional hazards model.

3.7.4 Artificial Neural Network (ANN) Model

Artificial Neural networks are machine learning methods evolved from the idea of simulating the human brain. They are networks of simple processing elements, namely 'neurons'. The neurons of a neural network are ordered to an input layer, one or more intermediate hidden layer and an output layer. The output of the network corresponds to the variable that is required to be predicted. The most commonly utilized form of neural networks is the Multi-layer perceptron (MLP). MLPs are general-purpose flexible models. Giving enough hidden neurons and enough data, a MLP can approximate any function to any desired degree. A two-layer MLP is shown in Fig.4.1.

ANN can be expressed mathematically:

$$u_k = \sum_{j=1}^{m} w_{kj} x_j \qquad (4.10)$$

$$y_k = \psi(u_k + b_k)$$

(4.11)

Where $x_1, x_2, \dots x_m$ features are (input signals): $w_{k1}, w_{k2}, \dots w_{km}$ are synaptic weight, u_k is the linear combination of input signals, $\psi(u_k + b_k)$ is sigmoid function (activation function) which produce output y_k.

3.7.4.1 Neural network model as extension of the cox proportional hazard model

A neural network extension of the Cox proportional hazard model firstly proposed in (Faraggi and Simon, 1995). This model generalizes Cox model to allow nonlinear function in place of the usual linear combination of covariates and provides the ability to model complexities and interactions in the input data that simple cox model would miss (Xiang *et al*. (2000)). The model has a single hidden layer with logistic activation function and an output layer with linear activation function. Mathematically the cox model is written as:

$$h(t, X_i) = h_0(t)\exp(\beta^T X_i) \quad (4.12)$$

Here the linear function $\beta^T X_i$ will be replaced by $\psi(u_k + b_k)$ the output of a neural network. So the proportional hazard model will be:

$$h(t, X_i) = h_0(t)\exp[\psi(u_k + b_k)] \quad (4.13)$$

By utilizing the Newton-Raphson method to maximize the partial likelihood function, the maximum likelihood estimates of the k Parameters are obtained. This model preserves all advantages of the Cox proportional hazard model and permits standard statistical methods for evaluating covariates. Also the model allows for modeling time-varying effects.

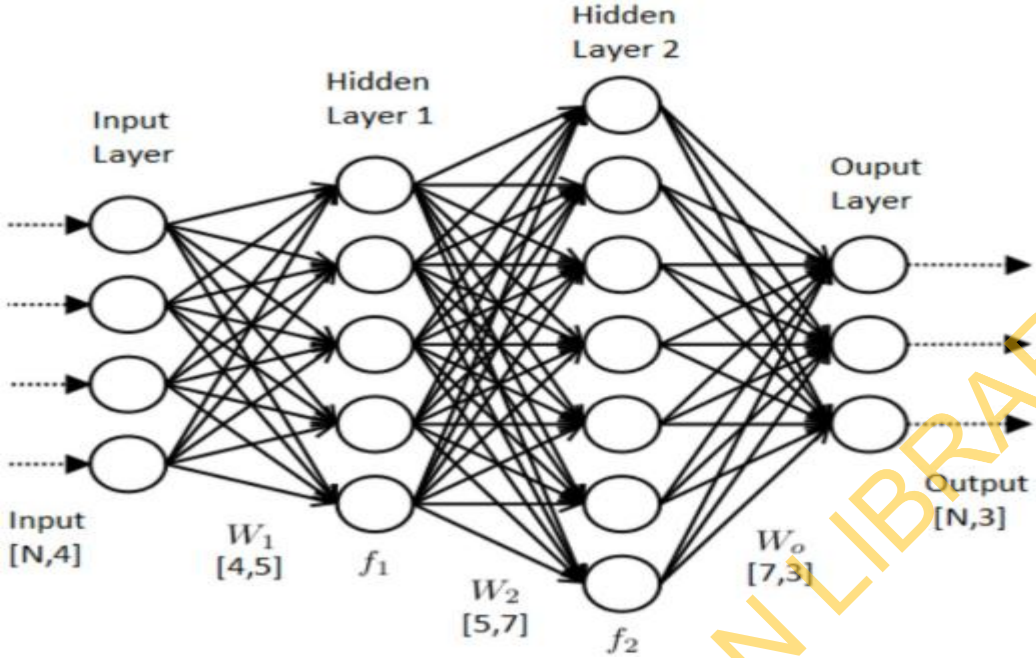

Figure 0.1.1: Neural Network with two hidden layers

CHAPTER FOUR

RESULTS

Risk factors of FGM from birth to the age or period at which women were exposed to cutting (in years) were model using Kaplan Meyer method (Non-parametric), Cox proportion hazard rate (parametric), and neural network (parametric). Thus, this study attempts to include socio-economic, demographic and environment related factors that are assumed as a potential determinants of FGM adopted from literature reviews and their theoretical justification. Possible explanatory variables for FGM includes education, religion, residence, ethnicity, daughter circumcision, region, religious belief, opinions about FGM from the respondents and wealth index.

4.1: Frequency Distribution of Background Characteristics of Respondents

Table 4.1 showed that the highest proportion of women falls within the age group 25-29 with 16.9% while the lowest proportion falls within age group 45-49 with 11.4%. Majority of the women are from North western region with 24% of the total respondents while the lowest proportions of women are from south southern region with 12.2% of the total respondents. The highest proportions of women are from other ethnic group (38.7%) while the lowest are from Yoruba ethnic group (15.7%). The highest proportions are from rural residence (54.1%).

The table also suggest that the highest proportion of the respondents are Christians (49.6%) and majority of them have secondary school qualification (38.3%) and minority have their daughter circumcised (15.8%) and majority religious belief do not support FGM (77.9%).

Among all the considered samples, most of the females support FGM practice should be stopped (69.20%) and are from the richest wealth index with 22.0%. The highest proportions of the women's household head are male with 82.2.

Table 4.1: Distribution of characteristics of the respondents

Variables	Levels	Frequency	Percentage (%)
Region	NC	1922	12.4
	NE	2953	19.0
	NW	3720	24.0
	SE	2607	16.8
	SS	1900	12.2
	SW	2410	15.5
Residence	Urban	7118	45.9
	Rural	8394	54.1
Respondent Circumcised	No	10724	69.1
	Yes	4788	30.9
Opinion	Continued	3425	22.1
	Stopped	10733	69.2
	Depends	1143	7.4
	Don't know	211	1.4
Religion Belief	No	12087	77.9
	Yes	2691	17.3
	Don't know	734	4.7
Daughter Circumcised	No	13065	84.2
	Yes	2447	15.8
Ethnicity	Yoruba	2438	15.7
	Igbo	3111	20.1
	Hausa	3961	25.5
	Others	5997	38.7
	Missing value	5	.0
Age group	15-19	2503	16.1
	20-24	2466	15.9
	25-29	2623	16.9
	30-34	2258	14.6
	35-39	2105	13.6
	40-44	1794	11.6
	45-49	1763	11.4
Wealth Index	Poorest	2889	18.6
	Poorer	2751	17.7
	Middle	3058	19.7
	Richer	3404	21.9
	Richest	3410	22.0
Religion	Christian	7695	49.6
	Islam	7725	49.8
	Traditional	62	.4
	Others	30	.2
Sex of the household head	Male	12745	82.2
	Female	2767	17.8
Highest Education Attained	No education	5143	33.2
	Primary	2387	15.4
	Secondary	5935	38.3
	Higher	2047	13.2
Total		15512	100.0

4.2: Prevalence of FGM by Women Background Characteristics

Table 4.2 shows that the overall prevalence rate of FGM was 31% for 4788 women. The table suggests that the prevalence rate of FGM among women in urban residence (35.61) was higher compared to rural residence. Southern region had the highest prevalence rate with south east and south west having 47.07 and 46.1 prevalence rate respectively.

The women's opinion on FGM, the number of women who support FGM practice should continue were the highest with (55.04) while the lowest prevalence were women who supported FGM should stopped with 22.42. The prevalence of FGM among women that had their women circumcised and whose household sex were female was higher. Majority of the women who had FGM do not circumcised their daughter (69.8) and had male as their household head.

The prevalence of FGM among women with primary and tertiary (higher) qualification was 38.42 and 26.62 respectively. Yoruba ethnic group had the highest prevalence with 51.31 followed by the Igbo with 43.88. The highest prevalence of FGM was among the richer women (34.46) who were Christians (34.04) and falls within the age group 45-49 years.

Table 4.2: Prevalence of FGM by women background characteristics

Variable	Levels	Total	Frequency	Percentage (%)
Age Distribution	15-19	2503	669	26.73
	20-24	2466	596	24.17
	25-29	2623	708	26.99
	30-34	2258	676	29.94
	35-39	2105	686	32.59
	40-44	1794	683	38.07
	45-49	1763	770	43.68
Residence	Urban	7118	2535	35.61
	Rural	8394	2253	26.84
Region	NC	1922	446	23.20
	NE	2953	243	8.23
	NW	3720	1221	32.82
	SE	2607	1227	47.07
	SS	1900	538	28.32
	SW	2410	1113	46.18
Opinion	Continued	3425	1885	55.04
	Stopped	10733	2406	22.42
	Depends	1143	409	35.78
	Don't know	211	88	41.71
Religious Belief	No	12087	3340	27.63
	Yes	2691	1172	43.55
	Don't know	734	276	37.60
Daughter Circumcised	No	13065	3342	25.58
	Yes	2447	1446	59.09
Household Head Sex	Male	12745	3770	29.58
	Female	2767	1018	36.79
Education	No education	5143	1384	26.91
	Primary	2387	917	38.42
	Secondary	5935	1942	32.72
	Higher	2047	545	26.62
Ethnicity	Yoruba	2438	1251	51.31
	Igbo	3111	1365	43.88
	Hausa	3961	1219	30.78
	Others	5997	952	15.87
Wealth Index	Poorest	2889	761	26.34
	Poorer	2751	814	29.59
	Middle	3058	1060	34.66
	Richer	3404	1173	34.46
	Richest	3410	980	28.74
Religion	Christian	7695	2619	34.04
	Islam	7725	2154	27.88
	Traditional	62	13	20.97
	Others	30	2	6.67
Age at Circumcision	Less than 5		4041	
	5-9		206	
	10-14		225	
	15-19		146	
	20-24		60	
	25-29		22	
	30-34		9	
	Greater than 34		5	
	Don't know		74	
Total		15512	4788	30.9

4.3 Survival Analysis

4.3.1 Kaplan Meier (KM) Curves

The KM survival curve is a plot of survival probability against time and can be used to estimate the median survival time. The horizontal axis represents time in years and the vertical axis shows the probability of surviving of women of ages 0-49 years.

Figure 4.1 suggested that there was a survival advantage for women from other ethnic groups to be circumcised compared to women from Yoruba, Igbo and Hausa ethnic groups.

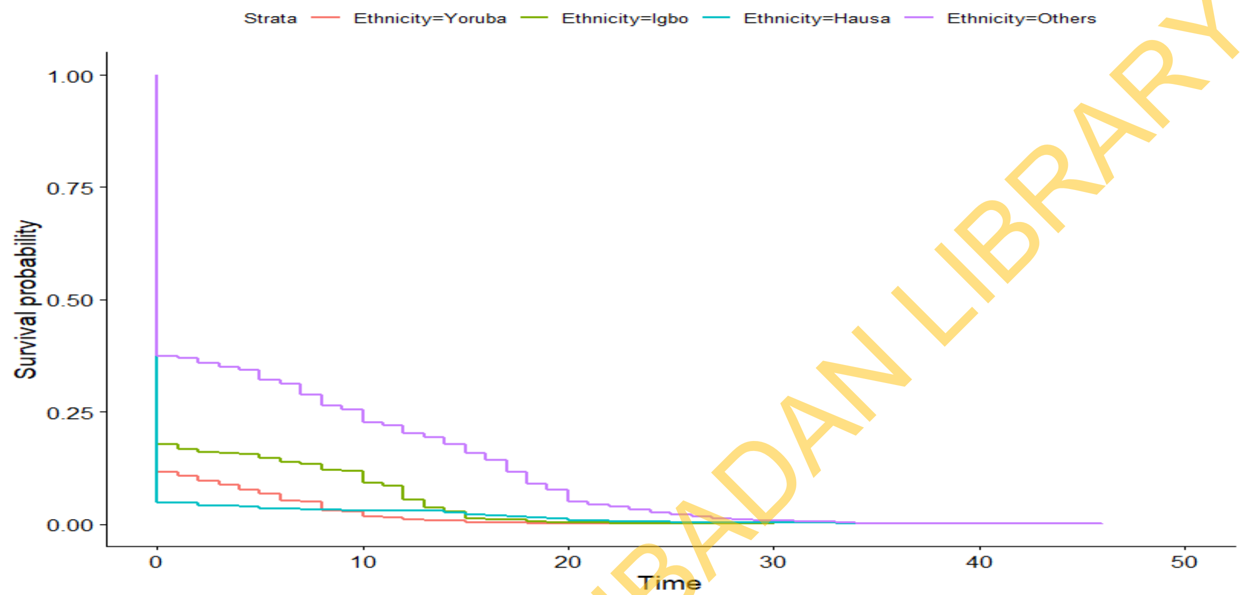

Figure 4.1: Survival curve for comparing groups for Ethnicity

Figure 4.2 suggested that there was a survival advantage for women with higher education to be circumcised compare to other women with other educational qualification.

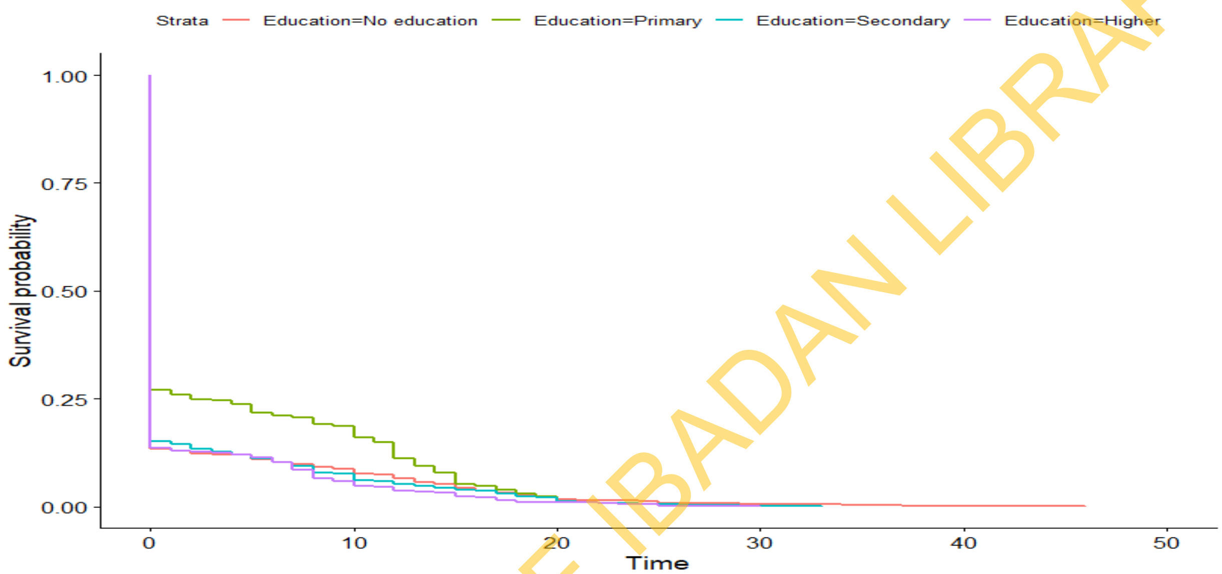

Figure 4.2: Survival curve for comparing groups for Education

Figure 4.3 suggested that there was a survival advantage for women whose daughters were not circumcised to be circumcised compared to other women whose daughters were circumcised.

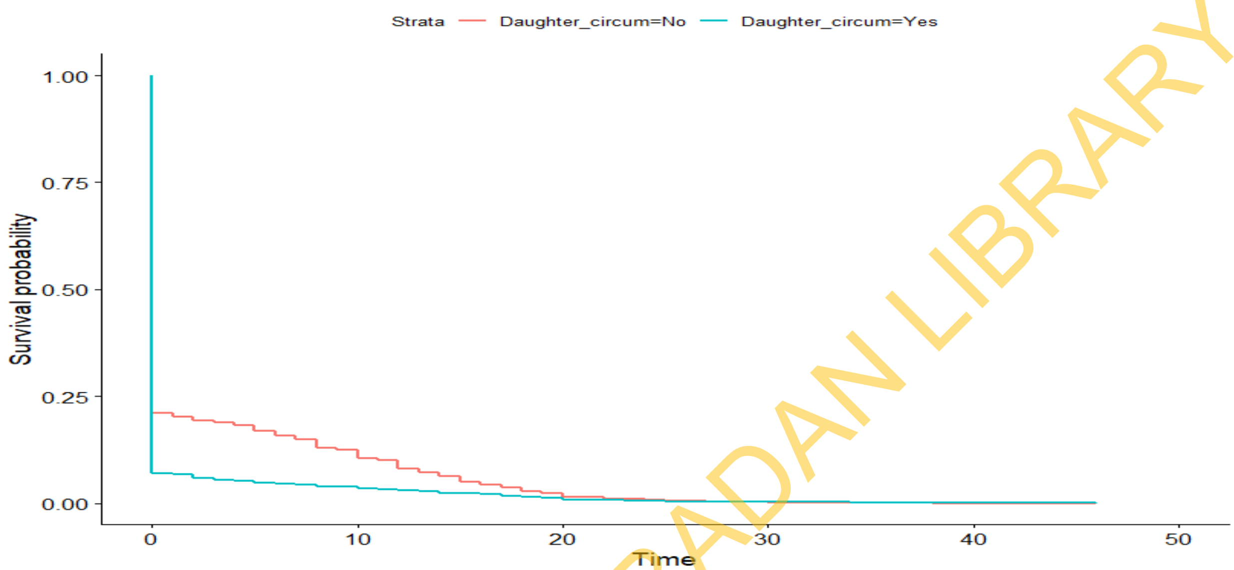

Figure 4.3: Survival curve for comparing groups for Daughter Circumcision

Figure 4.4 suggested that there was survival advantage for women whose household heads were male to be circumcised compared to other women whose household heads were female.

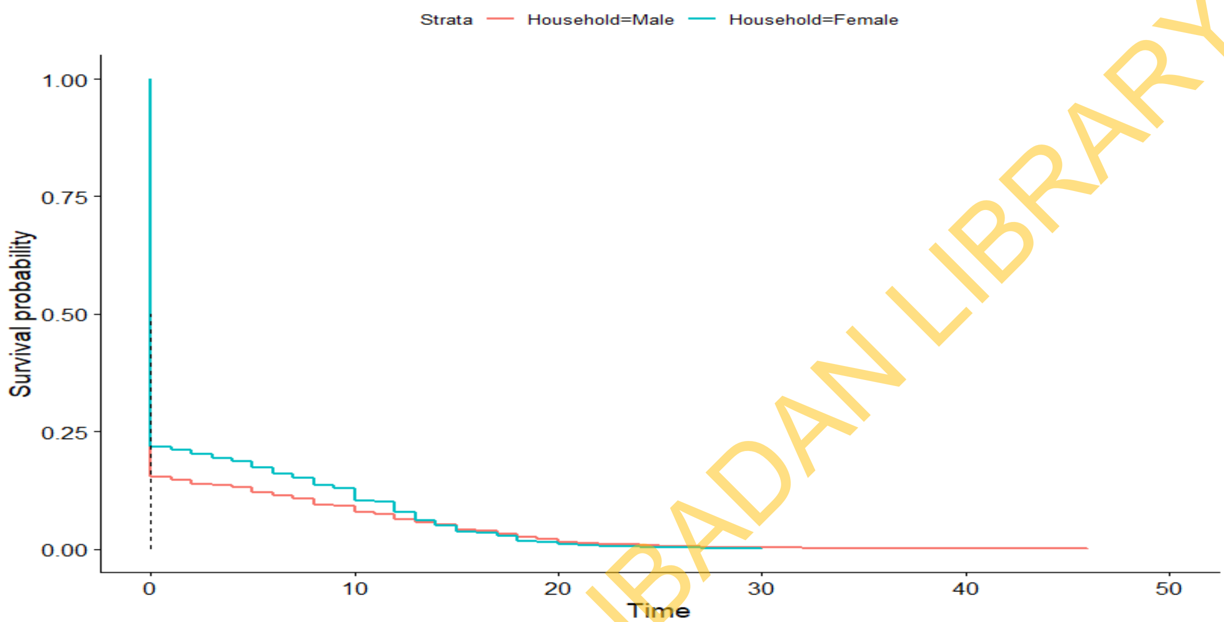

Figure 4.4: Survival curve for comparing groups for Sex of the Household Head

Figure 4.5 suggested that there was survival advantage for women opinion was that FGM should be stopped to be circumcised compared to other women's opinion on FGM.

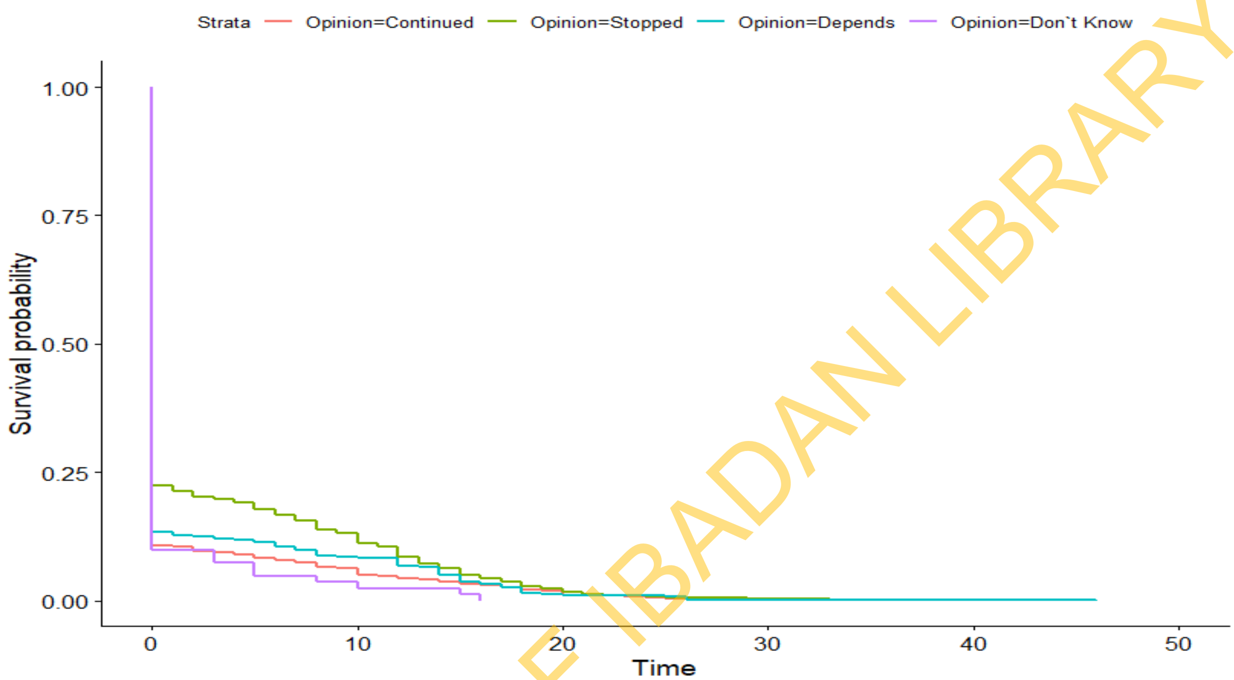

Figure 4.5: Survival curve for comparing groups for Opinion about FGM

Figure 4.6 suggested that there was survival advantage for women from rural residence to be circumcised compared to women from urban residence.

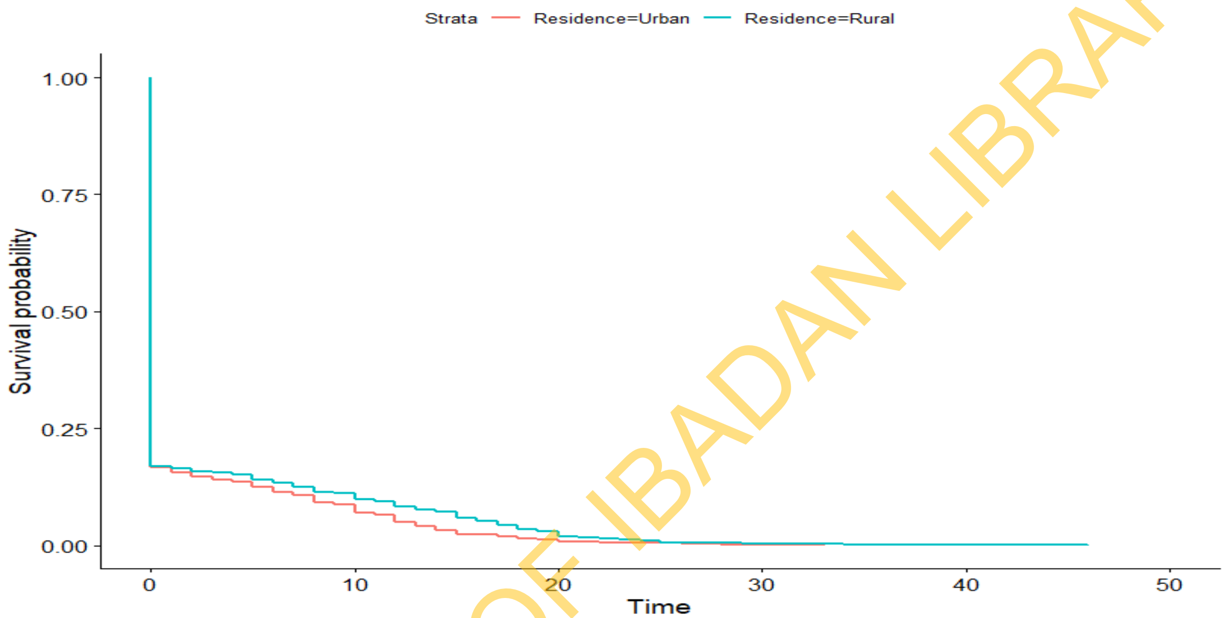

Figure 4.6: Survival curve for comparing groups for Residence

Figure 4.7 suggested that there was survival advantage for women from traditional religious group to be circumcised compare to other women from other religious groups.

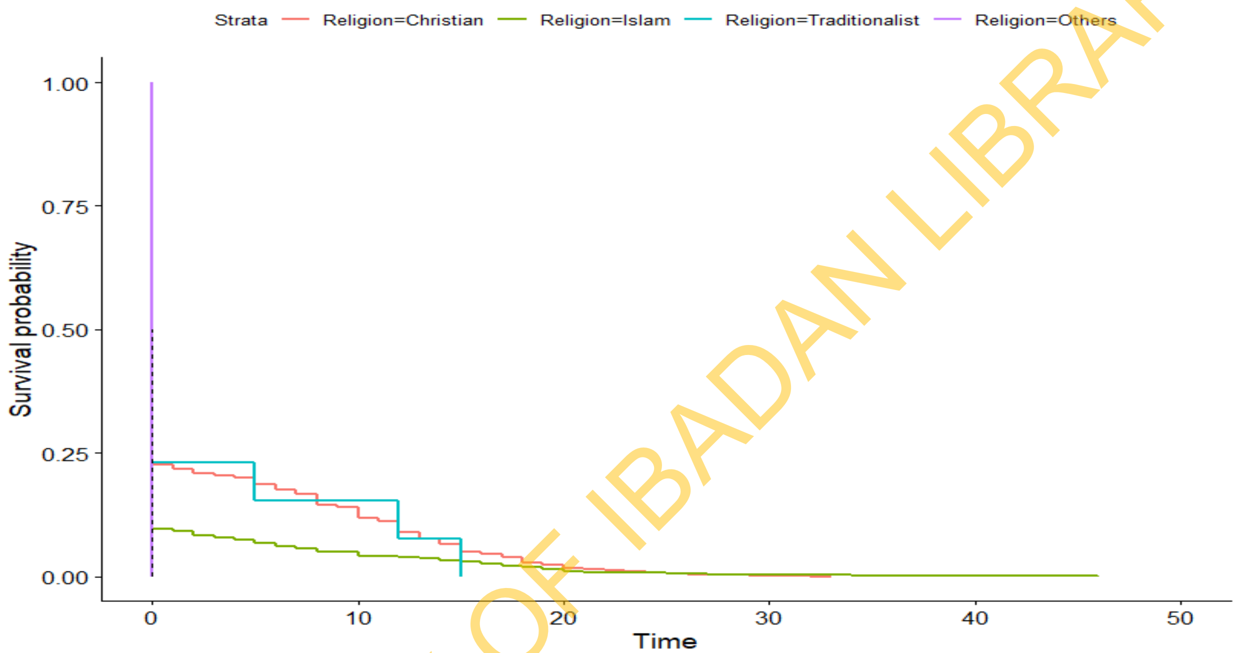

Figure 4.7: Survival curve for comparing groups for Religion

Figure 4.8 suggested that there was survival advantage for women whose religion belief do not support FGM to be circumcised compared to other women religious belief.

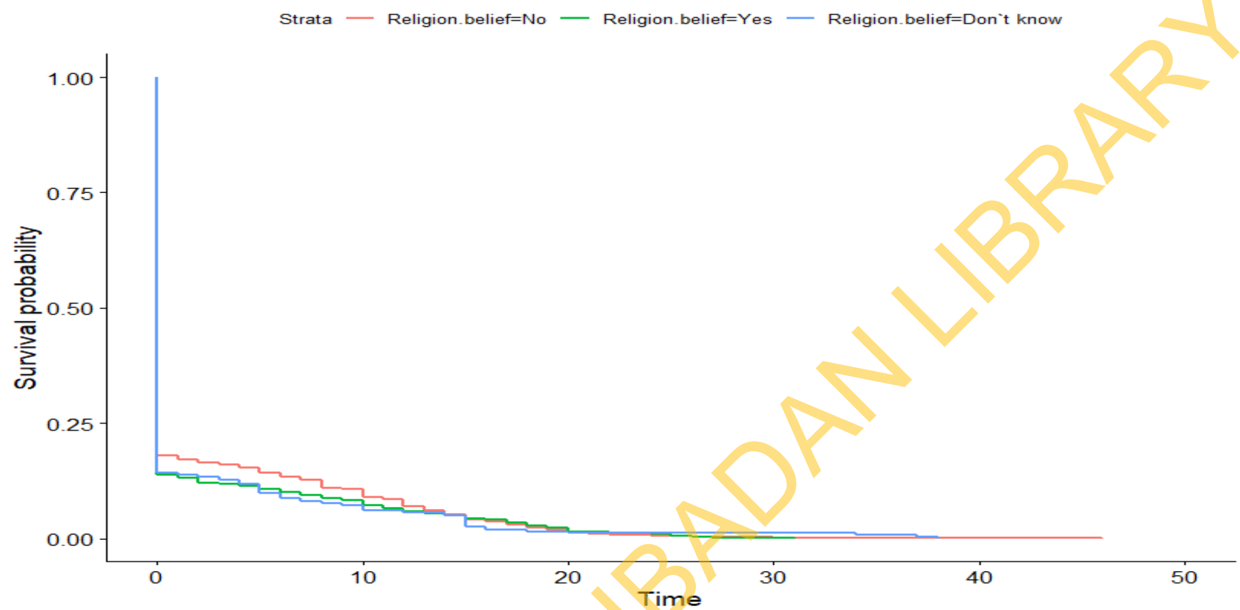

Figure 4.8: Survival curve for comparing groups for Religion Belief

Figure 4.9 suggested that there was survival advantage for women from north central region to be circumcised compare to women from other regions.

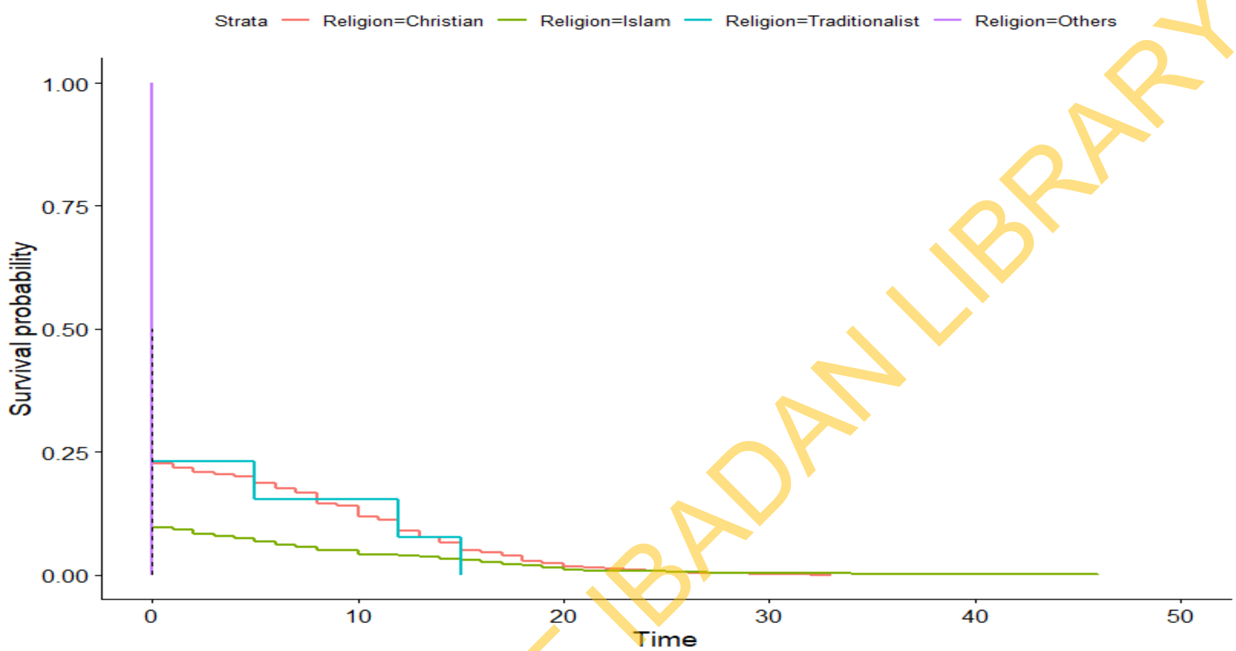

Figure 4.9: Survival curve for comparing groups for Region

Figure 4.10 suggested that there was survival advantage for women with the poorest wealth index to be circumcised compare to other women with other wealth index.

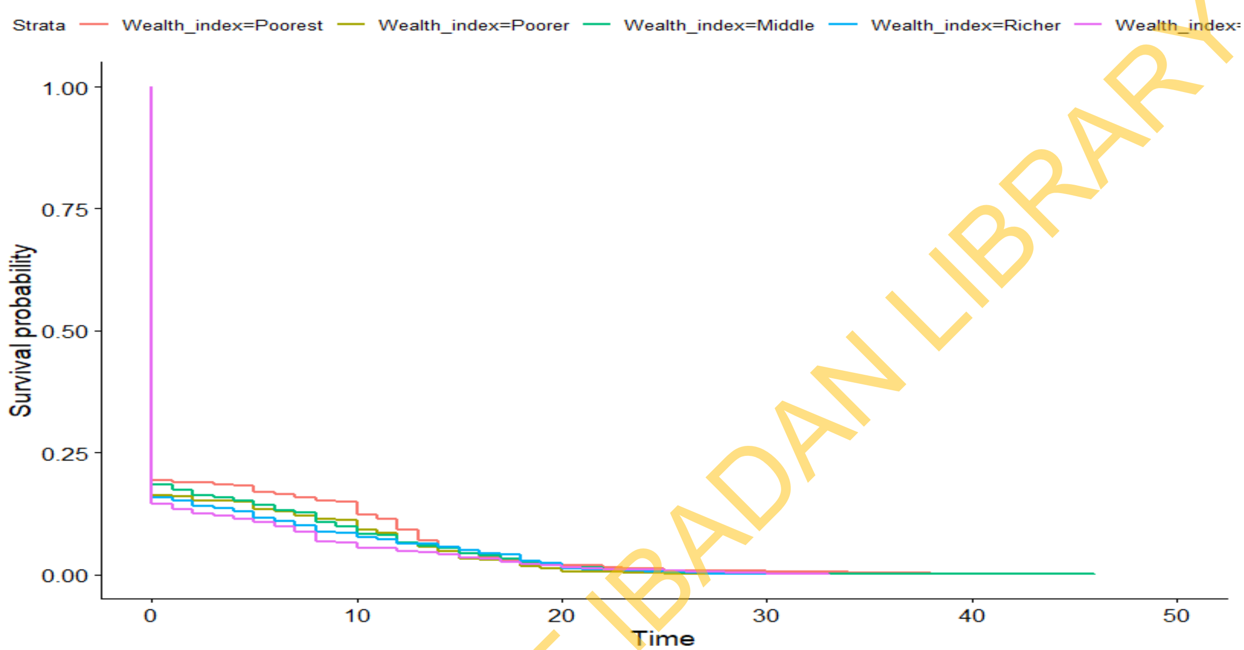

Figure 4.10: Survival curve for comparing groups for Wealth Index

4.3.2: Cox Proportional Hazard Model

The analysis was carried out with R software. Data was loaded and missing values removed. To avoid the problem of over fitting, the data was split into a training set and testing set. 70 percent of the data set was randomly selected without replacement into the training set while the rest was used as the testing set. The training set was used to fit the cox proportional hazard model and the testing set was used to test the predictive ability (accuracy) of the model.

The survival function and the cumulative hazard model are calculated relative to the baseline (lowest value of covariates or reference category) at each time point. The reference category for the covariates is denoted by RC in table 4.3.

Time to Female Genital mutilation (FGM) using the covariates like region, type of residence, religion, ethnicity, education, daughter circumcision, religious belief, opinion about FGM and wealth index. The table suggest that there is a significance difference in the hazard function between North East & North Central, South East & North Central and South South& North Central since their p-value is less than 0.01 level of significance. There is a significance difference in the hazard function between south west and north central since its p-value is less than 0.05 level of significance but there is no significance difference between hazard function of North West and north central since the p-value is greater than 0.01 level of significance. Females from North east and north west had 33.47% and 99.38% respectively lower hazard of being circumcised compared to females from north central region while females from the south east, south south and south west region had 62.44%, 67.63% and 22.69% respectively higher hazard of being circumcised compare to female from north central region. There is a significance difference in the hazard function between type of residence at 0.01 level of significance and females from urban residence had 13.46% higher hazard of being circumcised compared to females from rural residence.

Also, there is a significance difference in the hazard function between Islam and Christianity at 0.01 level of significance. Females who practice Islamic and traditional religion had 54.29% and 3.68% respectively higher hazard of being circumcised compared to females practicing Christianity religion while females practicing other religion had 34.88% lower hazard of being circumcised compare to females practicing Christianity religion. There is a significance difference in the hazard function between daughter circumcision and females whose daughter was circumcised had 26.62% higher hazard of being circumcised compared to females whose daughter was not circumcised.

There is also a significance difference in the hazard function between Yoruba & Hausa and Igbo & Hausa at 5% and 1% level of significance respectively. The table also suggest that females from Yoruba ethnic group had 26.28% higher hazard of being circumcised compared to females from Hausa ethnic group while females from Igbo and other ethnic group had 98.30% and 62.14% respectively lower hazard of being circumcised compare to females from Hausa ethnic group. There is a significance difference in the hazard function across religion belief. Females whose religious belief support FGM and females that do not know their religion stands on FGM had 16.89% and 23.92% higher hazard of being circumcised compared to females whose religion belief do not support FGM.

There is a significance difference in the hazard function of opinion (continued & stopped, continued & depends) and wealth index (poorest & poorer, poorest & middle, poorest & richer and poorest & richest) at 1% level of significance. Females whose opinion support stopping FGM, females whose opinion depends on circumstances surrounding their decisions and females whose don't know either to continue or stop FGM had 38.16%, 72.96% and 87.75% lower hazard of being circumcised compared to females whose opinion support continuing FGM. Also, females whose wealth index are poorer, middle, richer and richest had 94.72%, 78.27%, 65.35% and 54.30% respectively lower hazard of being circumcised compare to females whose wealth index is poorest.

Finally, there is a significance difference in the hazard function across the education level at 1% level of significance. Females with secondary education had 18.66% higher hazard of being circumcised compare to females with tertiary education while females with primary and no education had 73.23% and 50.72% lower hazard of being circumcised compare to females with tertiary education.

Table 4.3: Cox Proportional Hazard Model estimate

Variables	Levels	Coef	HR	S.E (Coef)	z	P-Value	95% c.I (HR)
Region	North Central	RC	1				
	North East	-1.0947	0.335	0.118	-9.318	<0.001	(0.266,0.421)
	North West	-0.006	0.994	0.098	-0.063	0.949	(0.820,1.204)
	South East	0.485	1.624	0.1275	3.804	0.0001	(1.265,2.086)
	South South	0.517	1.676	0.09	5.722	<0.001	(1.405, 2.001)
	South West	0.204	1.227	0.086	2.38	0.017	(1.037, 1.452)
Residence	Rural	RC	1				
	Urban	0.126	1.135	0.044	2.884	0.004	(1.041,1.236)
Religion	Christianity	RC	1				
	Islam	0.434	1.543	0.062	6.946	<0.001	(1.365, 1.744)
	Traditionalist	0.036	1.037	0.293	0.123	0.902	(0.584, 1.842)
	Others	-1.053	0.349	1.002	-1.051	0.293	(0.049, 2.485)
Ethnicity	Hausa	RC	1				
	Yoruba	0.233	1.263	0.108	2.156	0.031	(1.022,1.561)
	Igbo	-0.017	0.983	0.139	-0.123	0.902	(0.749, 1.290)
	Others	-0.476	0.621	0.09	-5.302	<0.001	(0.521, 0.741)
Education	Tertiary	RC	1				
	Secondary	0.171	1.187	0.063	2.711	0.007	(1.049,1.343)
	Primary	-0.312	0.732	0.075	-4.144	<0.001	(0.632, 0.849)
	No education	-0.679	0.507	0.086	-7.894	<0.001	(0.429, 0.60)
Daughter Circumcised	No	RC	1				
	Yes	0.236	1.266	0.054	4.402	<0.001	(1.140, 1.407)
Religion Belief	No	RC	1				
	Yes	0.156	1.169	0.052	3.021	0.003	(1.056, 1.294)
	Don't know	0.214	1.239	0.083	2.574	0.0101	(1.053,1.459)
Opinion	Continued	RC	1				
	Stopped	-0.963	0.382	0.051	-18.891	< 0.001	(0.345, 0.422)
	Depends	-0.315	0.73	0.075	-4.205	<0.001	(0.629, 0.845)
	Don't know	-0.131	0.878	0.14	-0.936	0.349	(0.668, 1.154)
Wealth Index	Poorest	RC	1				
	Poorer	-0.054	0.947	0.072	-0.755	0.45	(0.823, 1.090)
	Middle	-0.245	0.783	0.073	-3.348	0.0008	(0.678, 0.903)
	Richer	-0.425	0.654	0.077	-5.516	<0.001	(0.562, 0.760)
	Richest	-0.611	0.543	0.084	-7.274	<0.001	(0.461, 0.640)
Model Summary	Statistic		Value		df		P-value
	Likelihood ratio test		1704		25		<2E-16
	Wald test		1594		25		<2E-16
	Score (logrank) test		1769		25		<2E-16

N.B: RC stands for Reference Category. Variables with color red stand for there is statistical differences at 1% level of significance while variables with blue color stands for statistical differences at 5% level of significance

4.3.3: Artificial Neural Network (ANN) Model

Figure 4.11 is known as a feed forward network where each layer of nodes receives inputs from the previous layers. Figure 4.11 has a total of 30 weights, B1 and B2 represent the bias, O1 and O2 represent the output layers, H1 represents the hidden layer and I1 to I25 represent the input layers. The weights are:

b->h1 (308.01) i1->h1 (48.90), i2->h1 (57.68), i3->h1 (65.10), i4->h1 (47.10), i5->h1 (40.89), i6->h1 (151.45), i7->h1 (122.08), i8->h1 (2.04), i9->h1 (-0.17), i10->h1 (39.68), i11->h1 (72.26), i12->h1 (139.54), i13->h1 (125.16), i14->h1(53.07), i15->h1 (82.70), i16->h1 (29.42), i17->h1 (38.82), i18->h1 (15.28), i19->h1 (228.02), i20->h1 (19.33), i21->h1(5.35), i22->h1(42.06), i23->h1(73.58), i24->h1(88.85), i25->h1 (58.79), b->o1(5948.16), h1->o1(1420.06), b->o2(-31.97), h1->o2(-7.57).

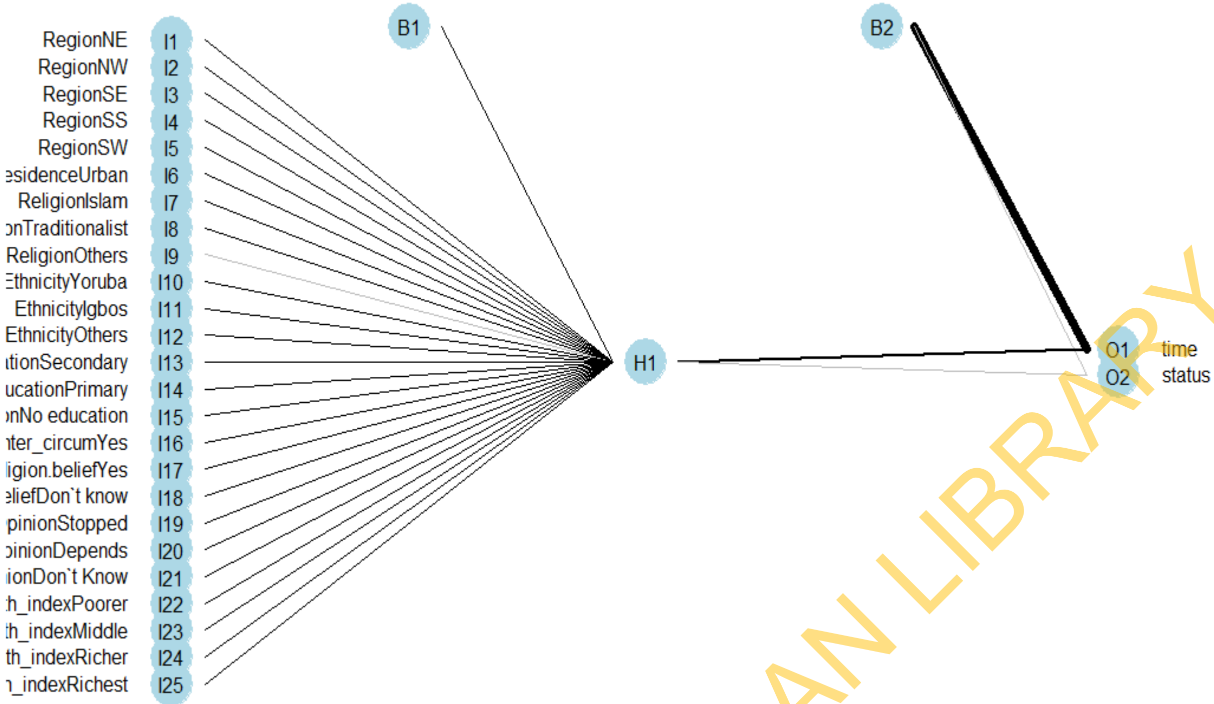

Figure 4.11: A neural network with twenty five inputs layers, one hidden layer and two outputs layers

The olden method was used to determine the relative importance of the predictors. The combined effects of the weight on the model predictions represent the relative importance of the predictors with the outcome variable.

The relative importance of each predictor is shown in table 4.5 and figure 4.12 below. It suggest that the type of residence is the most important predictor associated with risk of FGM followed by level of education. The variable importance values are the weighted values that showed how important each levels of the variables are.

Table 1.4: Relative Importance of the explanatory variables using neural network model

Variables	Levels	Variable Importance	Mean
Region	North Central (RC)		
	North East	533065.62	
	North West	555912.66	
	South East	552700.16	
	South South	380547.87	
	South West	374252.18	479295.7
Residence	Rural (RC)		
	Urban	1183809.7	1183810
Religion	Christianity (RC)		
	Islam	1214381.1	
	Traditionalist	14798.346	
	Others	4319.649	411166.4
Ethnicity	Hausa (RC)		
	Yoruba	368197.59	
	Igbo	660928.67	
	Others	1120784	716636.8
Education	Tertiary (RC)		
	Secondary	896068.69	
	Primary	468006.44	
	No education	963834.57	775969.9
Daughter Circumcised	No (RC)		
	Yes	233251.12	233251.1
Religion Belief	No (RC)		
	Yes	438550.18	
	Don't know	102870.13	270710.2
Opinion	Continued (RC)		
	Stopped	2113816.1	
	Depends	160531.62	
	Don't know	23808.58	766052.1
Wealth Index	Poorest (RC)		
	Poorer	517854.81	
	Middle	536741.15	
	Richer	519872.63	
	Richest	676924.82	562848.4

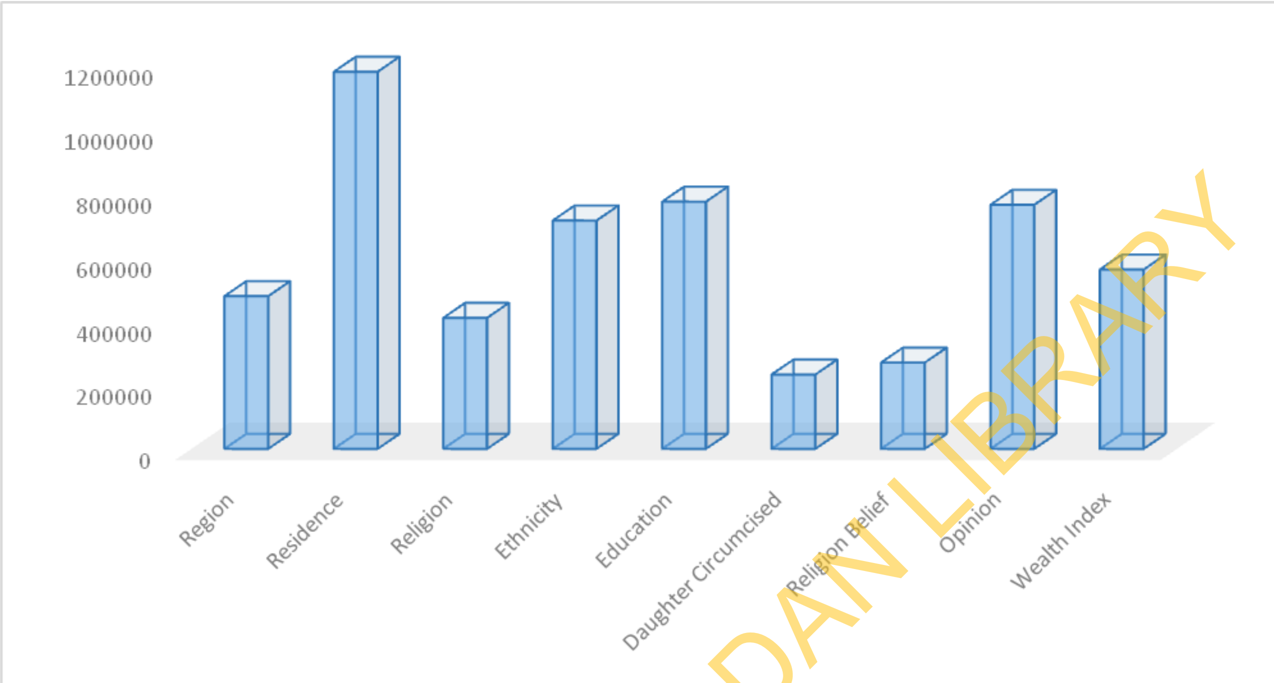

Figure 4.12: Plot of the relative importance of the variables using neural network

4.3.4: Comparison between Cox Proportional Hazard model and Artificial Neural Network (ANN) model

One common measure used to compare two or more classification models is to use the area under the ROC curve (AUC) as a way to indirectly assess their performance. In this case the model with a larger AUC is usually interpreted as performing better than those with a smaller AUC. Cox proportional hazard model is better compare to ANN because it has the higher AUC value of 0.1908 compare to 0.0989.

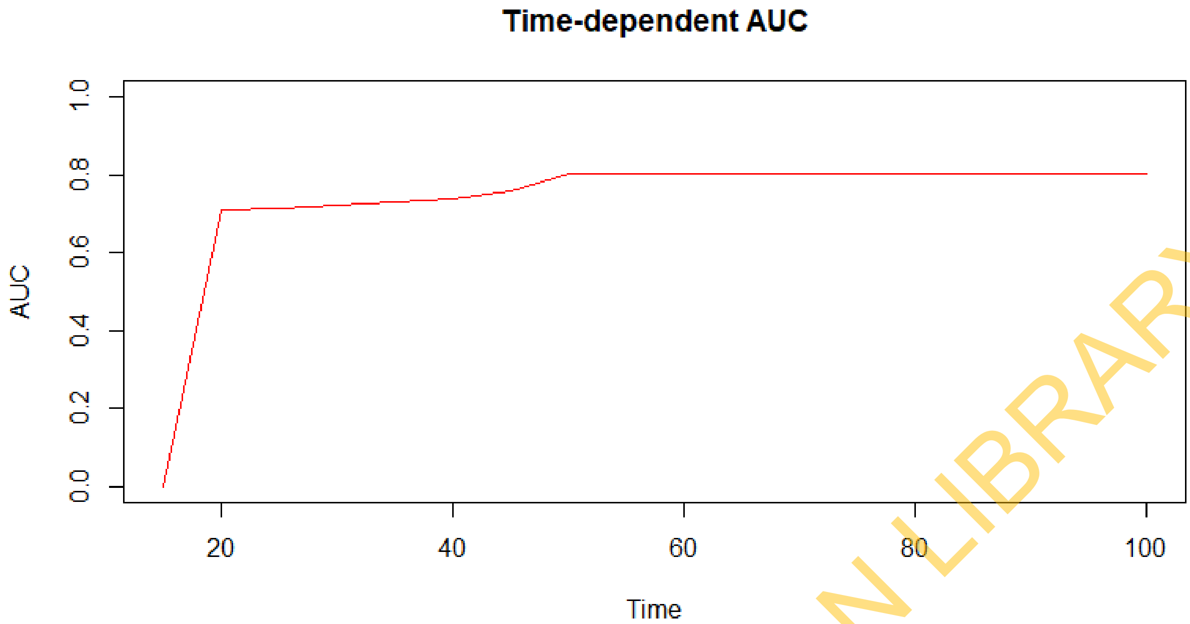

Figure 4.13: ROC for Cox Proportional Hazard Model

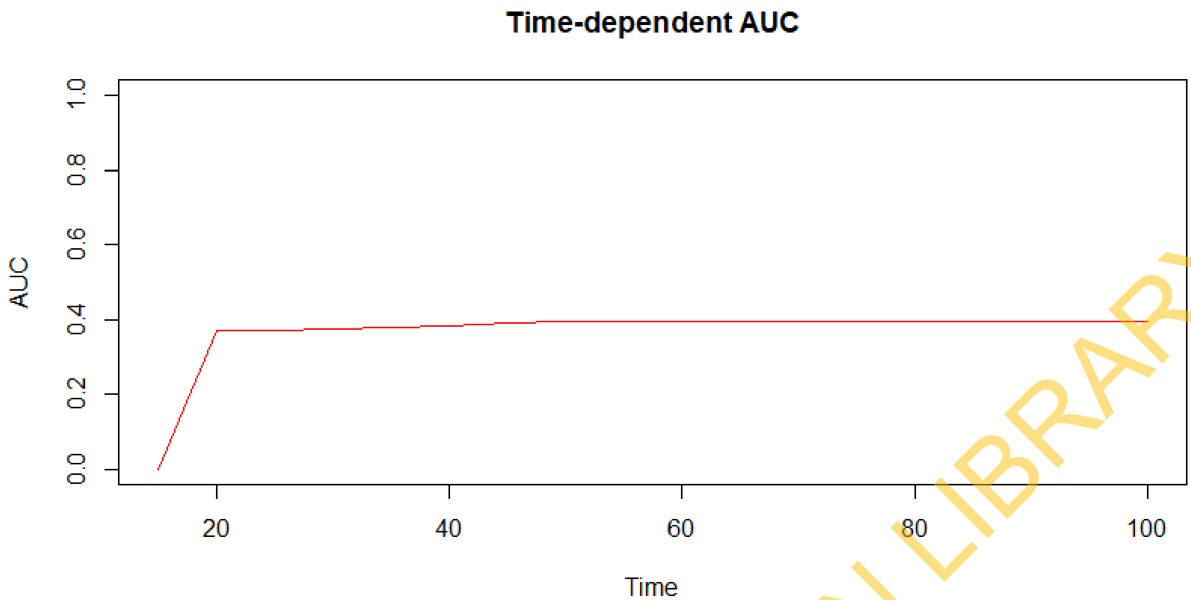

Figure 4.14: ROC for Artificial Neural Network Model

CHAPTER FIVE

DISCUSSION, CONCLUSION AND RECOMMENDATION

The findings on Female Genital Mutilation (FGM), the distribution and prevalence of FGM on the background characteristics of the respondent were summarized in this chapter. Also, the findings using Kaplan Meier method (Non-parametric), Cox proportion hazard rate (parametric), and neural network (parametric) to model the socio-economic, demographic and environment related factors as a potential determinants of FGM were also summarized.

5.1 Discussion

This study evaluated the performances of Artificial Neural Network (ANN) and Cox Proportional Hazard (CPH) in modeling time to Female Genital Mutilation (FGM) in Nigeria. The data suggest that the prevalence rate of FGM in Nigeria using the 2018 NDHS data was about 31%. The prevalence of FGM was dominant in the urban area and mainly concentrated in the southern part of the country.

Majority of the women that undergo FGM were from South Eastern region and were Igbos. Among the women who had FGM, most of the women support FGM practice should be stopped and majority of them were Christians and their religious belief do not support FGM. Majority of the women who had FGM do not circumcised their daughter and had male as their household head.FGM prevalence was higher in richer households and majority of them have secondary school qualification and they falls within the age group (45-49) and had their circumcision at age less than 5 years.

The findings from Kaplan Meier (KM) Curves showed that there was a significant difference between the survival curves among the levels of region, religious belief, opinion, wealth index, daughter circumcision, education, ethnicity, residence, religion and sex of the household head at 1% level of significance. The findings from the CPH model suggested that women from urban residence had higher hazard of being circumcised compared to females from rural residence. This was against the findings of (Setegn *et al*., 2016) and(Daniel, 2019)which revealed that girls whose mothers lived in urban areas had longer survival with respect to age at circumcision than girls whose mothers resided in rural areas.

Women from North east and North west had lower hazard of being circumcised compared to women from north central region while women from the south east, south south and south west region had higher hazard of being circumcised compare to female from north central region. Also, women Yoruba ethnic group had higher hazard of being circumcised compared to women from Hausa ethnic group while women from Igbo and other ethnic group had lower hazard of being circumcised compare to women from Hausa ethnic group. This may be due to higher prevalence of FGM in the southern part of the country and the concentration of FGM abandonment programs in the south (Uchenna Mberu, 2017).

Women who practice Islamic and traditional religion had higher hazard of being circumcised compared to women practicing Christianity religion while women practicing other religion had lower hazard of being circumcised compare to women practicing Christianity religion. Also, women whose religious belief support FGM and women that do not know their religion stands on FGM had higher hazard of being circumcised compared to women whose religion belief do not support FGM. This was supported by finding by (Abdisa *et al*., 2017)that found girls whom their mothers are Muslims are more vulnerable towards circumcision.

A woman's opinion on the continuation of FGM is a good indicator of the preventing social norms in a community and the risk for the practice being perpetuated across generations(Shell-

Duncan *et al*., 2018). Women whose opinion support stopping FGM, females whose opinion depends on circumstances surrounding their decisions and females whose don't know either to continue or stop FGM had lower hazard of being circumcised compared to women whose opinion support continuing FGM.

The study also showed that there was survival advantage for women with no educational qualification to be circumcised compare to other women with other educational qualification. This is consistent with the study conducted in Burkina Faso by (Karmaker *et al*., 2011) they revealed that the prevalence of girl's circumcision was high for those daughters whom their mothers are illiterate.

Also, women whose wealth index were poorer, middle, richer and richest had lower hazard of being circumcised compared to women whose wealth index was poorest and women whose daughter was circumcised had higher hazard of being circumcised compared to women whose daughter was not circumcised.

The result of the Cox proportional suggested that women from North east and north west had lower hazard of being circumcised compared to women from north central region while women from the south east, south south and south west region had higher hazard of being circumcised compared to female from north central region. Women from urban residence had higher hazard of being circumcised compared to women from rural residence.

Women who practice Islamic and traditional religion had higher hazard of being circumcised compared to women practicing Christianity religion while women practicing other religion had lower hazard of being circumcised compared to females practicing Christianity religion. Women whose daughter was circumcised had higher hazard of being circumcised compared to women whose daughter was not circumcised.

Also, women from Yoruba ethnic group had higher hazard of being circumcised compared to women from Hausa ethnic group while women from Igbo and other ethnic group had lower hazard of being circumcised compare to females from Hausa ethnic group. Women whose religious belief supported FGM and women that do not know their religion stands on FGM had higher hazard of being circumcised compared to women whose religion belief do not support FGM.

Women whose opinion support stopping FGM, women whose opinion depends on circumstances surrounding their decisions and women who don't know either to continue or stop FGM had lower hazard of being circumcised compared to females whose opinion support continuing FGM. Also, women whose wealth index are poorer, middle, richer and richest had lower hazard of being circumcised compared to women whose wealth index was poorest.

Finally, women with secondary education had higher hazard of being circumcised compared to women with tertiary education while women with primary and no education had lower hazard of being circumcised compared to women with tertiary education.

The findings from the ANN model suggested that the type of residence, level of education, opinion of the women on FGM and ethnicity are the most important predictor associated with risk of FGM. The comparison between the CPH and ANN model suggested that CPH model was better in terms of classification ability than ANN because it has the higher AUC value.

5.2 Conclusion

Female genital mutilation is a crime against womanhood, posing a great health and financial burden to individuals, families and the society. Although its prevalence is on the increase in many parts of Nigeria, more sustained and coordinated efforts of stakeholders at all levels are needed to fast-track the elimination of this practice in Nigeria. The type of residence, level of education, opinion of the women on FGM and ethnicity are the most important predictor associated with risk of FGM. Artificial Neural Network and Cox Proportional Hazard models are appropriate for predicting time to FGM.

5.3 Strengths and Limitation

This report has several shortcomings. Reliance on cross-sectional NDHS data means that causation between individual and community factors and FGM cannot be implied. Secondly, there was an unavoidable incompleteness of data since it was a secondary data analysis. Therefore, researchers should interpret the findings from this study with caution and precise rates of prevalence may well be under or over reported. Furthermore, there is a strong possibility that female uncut at the time of the survey were still at risk of being cut in the future. The survival analysis techniques used here addressed this issue by allowing for data censoring.

5.4 Recommendation

Artificial Neural Network and Cox Proportional Hazard Model are recommended for predicting and determining the influence of risk factors on the time to female genital mutilation. Government, private and Non-Governmental organizations should raise alarm and create more awareness of the risk affecting FGM.

REFERENCES

Abdisa, B., Desalegn, M., & Tesew, A. (2017). Assessment of the Prevalence of FGM and Associated Factors among Women's of Reproductive Age Group in Kebirbeyah Town, Somali Region Eastern Ethiopia, 2017. *Health Science Journal*, *11*(4), 517. https://doi.org/10.21767/1791-809x.1000517

Acar, D. (2019). *Evaluating the effect of region, religion and socioeconomic parameters on female genital mutilation in Egypt from 1987 to 2014*.

Achia, T. N. O. (2014). Spatial modelling and mapping of female genital mutilation in Kenya. *BMC Public Health*, *14*(1). https://doi.org/10.1186/1471-2458-14-276

Adegoke, P. (2005). Female genital mutilation: An African humanist view. Ibadan University Humanist Society.

Amir Hossein Hashemian, Behrouz Beiranvand, Mansour Rezaei, Abdolrasoul Bardideh and Eghbal Zand-Karimi (2013). Comparison of Artificial Neural Networks and Cox Regression Models in Prediction of Kidney Transplant Survival. International Journal of Advanced Biological and Biomedical Research.Volume 1, Issue 10. Page 1204-1212.

A. Xiang, P. Lapuerta, A. Ryutov, J. Buckley, and S. Azen, (2020) ''Comparison of the performance of neural network methods and Cox regression for censored survival data,'' Computational Statistics & amp; Data Analysis, 34. 2, pp. 243-257, 2000.

Babińska, M., Chudek, J., Chełmecka, E., Janik, M., Klimek, K., & Owczarek, A. (2015). Limitations of Cox proportional hazards analysis in mortality prediction of patients with acute coronary syndrome. *Studies in Logic, Grammar and Rhetoric*, *43*(56), 33–48.

Faraggi, D., & Simon, R. (1995). A neural network model for survival data. *Statistics in Medicine*, *114*(1), 73–82.

Gayawan, E., & Lateef, R. S. (2019). Estimating geographic variations in the determinants of attitude towards the practice of female genital mutilation in Nigeria. *Journal of Biosocial Science*, *51*(5), 645–657.

Hosseini Teshnizi S., Tzazhibi M., T. F. . (2013). Comparison of Cox Regression and Artificial Neural Network Model in Predicting of Survival in Acute Leukemia Patients. *The Scientific Journal of Iranian Blood Transfusion Organization (KHOON).*, *10*(39), 154 To 162.

Ishtiaq, M. (2019). Book Review Creswell, J. W. (2014). Research Design: Qualitative, Quantitative and Mixed Methods Approaches (4th ed.). Thousand Oaks, CA: Sage. *English Language Teaching*, *12*(5), 40. https://doi.org/10.5539/elt.v12n5p40

Kandala, N.-B., Nwakeze, N., & Kandala, S. N. I. (2009). Spatial distribution of female genital mutilation in Nigeria. The American Journal of Tropical Medicine and Hygiene, 81(5), 784–792.

Karmaker, B., Kandala, N. B., Chung, D., & Clarke, A. (2011). Factors associated with female genital mutilation in Burkina Faso and its policy implications. *International Journal for Equity in Health*, *10*(1), 20. https://doi.org/10.1186/1475-9276-10-20

Khosravi, B., Pourahmad, S., Bahreini, A., Nikeghbalian, S., & Mehrdad, G. (2015). Five Years Survival of Patients After Liver Transplantation and Its Effective Factors by Neural Network and Cox Poroportional Hazard Regression Models. *Hepatitis Monthly*, *15*(9).

Lundin, M., Lundin, J., Burke, H. B., Toikkanen, S., Pylkkänen, L., & Joensuu, H. (1999). Artificial neural networks applied to survival prediction in breast cancer. Oncology, 57(4), 281–286.

Lee, S., & Lim, H. (2019). Review of statistical methods for survival analysis using genomic data. In *Genomics and Informatics* (Vol. 17, Issue 4). Korea Genome Organization. https://doi.org/10.5808/GI.2019.17.4.e41

Morhason-Bello, I. O., Fagbamigbe, A. F., Kareem, Y. O., & Ojengbede, O. A. (2020). Economic status, a salient motivator for medicalisation of FGM in sub-Saharan Africa: Myth or reality from 13 national demographic health surveys. *SSM - Population Health*, *11*,

Morris, B. J., Waskett, J. H., Banerjee, J., Wamai, R. G., Tobian, A. A., Gray, R. H., Bailis, S. A., Bailey, R. C., Klausner, J. D., Willcourt, R. J., Halperin, D. T., Wiswell, T. E., & Mindel, A. (2012). A "snip" in time: what is the best age to circumcise? *BMC Pediatrics*, *12*(1), 20.

Masho, S. W., & Matthews, L. (2009). Factors determining whether Ethiopian women support continuation of female genital mutilation. International Journal of Gynecology & Obstetrics, 107(3), 232–235

National Population Commission(NPC)[Nigeria], & ICF International. (2019). *Nigeria Demographic and Health Survey 2018*.

Ngianga-Bakwin Kandala, Chibuzor Christopher, G. A. (2019). Statistical Modelling of Complex Correlated and Clustered Data Household Surveys in Africa. *Enviromental Research and Public Health*. *Nigeria28 Too Many*. (n.d.).

Odukogbe, A. T. A., Afolabi, B. B., Bello, O. O., & Adeyanju, A. S. (2017). Female genital mutilation/cutting in Africa. In *Translational Andrology and Urology* (Vol. 6, Issue 2, pp. 138–148). AME Publishing Company. https://doi.org/10.21037/tau.2016.12.01

Okeke, T., Anyaehie, U., & Ezenyeaku, C. (2012). An overview of female genital mutilation in Nigeria. *Annals of Medical and Health Sciences Research*, *2*(1), 70.

Osuorah, D. I. C. (2015). Sociodemographic Predictors of Genital Mutilation (Circumcision) of the Girl Child in Nigeria: A Population-Based Study. *International Journal of Women's Health and Reproduction Sciences*, *3*(3), 142–150. https://doi.org/10.15296/ijwhr.2015.30

Puddu, P. E., & Menotti, A. (2012). Artificial neural networks versus proportional hazards Cox models to predict 45-year all-cause mortality in the Italian Rural Areas of the Seven

Countries Study. *BMC Medical Research Methodology*, *12*(1), 100. https://doi.org/10.1186/1471-2288-12-100

Sabouri, S., Esmaily, H., Shahidsales, S., & Emadi, M. (2020). Survival Prediction in Patients with Colorectal Cancer Using Artificial Neural Network and Cox Regression. *International Journal of Cancer Management*, *13*(1). https://doi.org/10.5812/ijcm.81161

Sadiq, R., Oyejola, B. ., & Abiodun, A. . (2016). Comparison of Artificial Neural Networks and Cox Regression Models in Population Based Survival. *International Journal of Advanced Research in Science,Engineering and Technology*, *3*(3).

Setegn, T., Lakew, Y., & Deribe, K. (2016). Geographic Variation and Factors Associated with Female Genital Mutilation among Reproductive Age Women in Ethiopia: A National Population Based Survey. *PLOS ONE*, *11*(1), e0145329. https://doi.org/10.1371/journal.pone.0145329

Shell-Duncan, B., Moreau, A., Wander, K., & Smith, S. (2018). The role of older women in contesting norms associated with female genital mutilation/ cutting in Senegambia: A factorial focus group analysis. *PLoS ONE*, *13*(7). https://doi.org/10.1371/journal.pone.0199217

Uchenna Mberu, B. (2017). *FEMALE GENITAL MUTILATION/CUTTING IN NIGERIA: A SCOPING REVIEW*.

Unicef. (2001). Children's and women's rights in Nigeria: A wake-up call. Situation Assessment and Analysis. Harmful Traditional Practice (FGM) Abuja NPC and UNICEF Nigeria, 195–200.

WHO, U. A. U. (1997). FEMALE GENITAL MUTILATION. *WORLD HEALTH ORGANIZATION*.

Yang, X. (2011). *Cox's proportional hazards (Cox's PH model) is one of the most widely applied models in survival modeling with a hazard ratio of a group assumed to be constant to a baseline group through the observed time (Yang, 2011 - Google Search*.

ZHU, L., LUO, W., SU, M., WEI, H., WEI, J., ZHANG, X., & ZOU, C. (2013). Comparison between artificial neural network and Cox regression model in predicting the survival rate of gastric cancer patients. *Biomedical Reports*, *1*(5), 757–760. https://doi.org/10.3892/br.2013.140