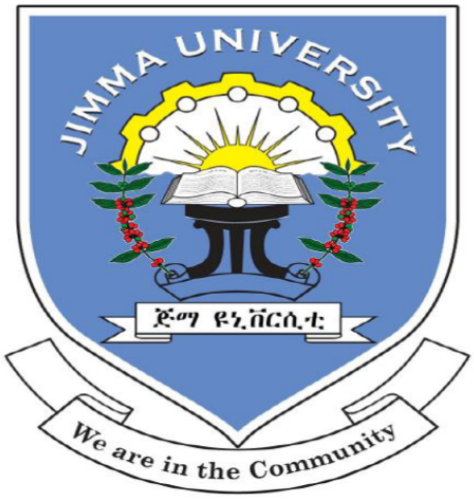


COLLEGE OF NATURAL SCIENCE

DEPARTMENT OF STATISTICS

BAYESIAN SURVIVAL ANALYSIS OF HEART FAILURE PATIENTS: A CASE STUDY IN JIMMA UNIVERSITY MEDICAL CENTER, JIMMA, ETHIOPIA

BY:

TAFESE ASHINE

A THESIS SUBMITTED TO JIMMA UNIVERSITY, COLLEGE OF NATURAL SCIENCE, DEPARTMENT OF STATISTICS IN THE PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE IN BIOSTATISTICS

FEBRUARY 28, 2020

JIMMA, ETHIOPIA

Bayesian Survival Analysis of Heart Failure Patients: A Case Study in Jimma University Medical Center, Jimma, Ethiopia

By:

Tafese Ashine

Advisor: Geremew Muleta (PhD Scholar)

Co-advisor: Kenenisa Tadesse (MSc)

February 28, 2020

Jimma, Ethiopia

JIMMA UNIVERSITY
COLLEGE OF NATURAL SCIENCE
DEPARTMENT OF STATISTICS

Bayesian Survival Analysis of Heart Failure Patients: A Case Study in Jimma University Medical Center, Jimma, Ethiopia

By
Tafese Ashine

As members of the Board of Examiners of M.Sc. thesis open defense examination of the above title, we have evaluated the thesis and examined the candidate.

Geremew Muleta (PhD Scholar)	.....	.....
Advisor	Signature	Date
Kenenisa Tadesse (MSc)	.....	.....
Co-advisor	Signature	Date
.....	.....	.....
Internal examiner	Signature	Date
.....	.....	.....
External examiner	Signature	Date
.....	.....	.....
Chairman, Department graduate committee	Signature	Date

DECLARATION

I, the undersigned, declare that the thesis is my original work, has not been presented for degrees in any other University and all sources of material used for the thesis have been duly acknowledged.

Tafese Ashine	. . . . .	. . . . .
Name	Signature	Date

This thesis has been submitted for examination with my approval as an advisors.

Geremew Muleta (PhD Scholar)	. . . . .	. . . . .
Name	Signature	Date
Kenenisa Tadesse (MSc)	. . . . .	. . . . .
Name	Signature	Date

ACKNOWLEDGMENT

Above all I would like to thank the almighty God who has been with me in all activities to be succeeded from the beginning to the end.

Next I would like to express sincere appreciation to my advisor Geremew Muleta (PhD Scholar) for his sharing substantial experience to do this research in the expect way and giving wonderful personality. Likewise, I am sincerely appreciate to my co-advisor Kenenisa Tadesse (MSc) for his valuable suggestions and comments.

I am also grateful to Dr. Birhanu Teshome (PhD) for his advice and guidance. I would like to thank Dr. Havard Rue (Professor of Statistics and developer of R-INLA) for his surprising continual email contact during R-INLA function and application of INLA. I appreciated all members of R-INLA group for their view on the questions I posted through the groups email.

Finally thanks to all staff members of Jimma University medical center for allowing me to collect the data set from the hospital. I am especially indebted to my beloved friends for their guidance and encouragement during the data collection. I am also very thankful to all my beloved parents and friends whose presence and support have always been important to me.

ABSTRACT

Heart failure is failure of the heart to pump blood with normal efficiency and globally growing public health issue with high death rate over the world including Ethiopia. The aim of this study was to identify factors affecting the survival time of heart failure patients in Jimma University Medical Center. To reach the aim, 409 heart failure patients were including in the study based on data taken from medical record of patients enrolled during January, 2016 to January, 2019. Kaplan Meier plots and log rank test were used for comparison of survival function; Bayesian survival models was used to identify factors affecting the survival time heart failure patients. Of the total patients in the study 164 (40.1%) were died. The estimated median survival time of patients was 31 months. Bayesian log-normal accelerated failure time model using Markov chain monte carlo and Integrated nested laplace approximation method fit heart failure data-set better than other Bayesian accelerated failure time models used in this study. The Bayesian log-normal accelerated failure time model using Integrated nested laplace approximation method was preferable than Markov chain monte carlo method due to smaller standard error and narrow credible interval. From the results of this model shows that the survival time of heart failure patients significantly affected by age, chronic kidney disease, diabetes mellitus, etiology of heart failure, hypertension, anemia, smoking cigarette and stages of heart failure. Bayesian log-normal accelerated failure time model using Integrated nested laplace approximation method describes the heart failure data-set well. Age group (49 to 65 years and greater than 65 years); etiology of heart failure (rheumatic valvular heart disease, hypertensive heart disease and Other diseases); presence of hypertension; presence of anemic; presence of chronic kidney disease; smokers; diabetes mellitus (type I and type II diabetic); and stages of heart failure (II, III and IV) were prolong the timing death of heart failure patients. The hospital, Jimma University medical center, need to improve public awareness for early detection of heart failure.

Keywords: Heart failure, Bayesian, Survival Analysis, MCMC, INLA

ACRONYMS

AFT:.....Accelerated Failure Time
AHA:.....America Health Association
AIC:.....Akaikes Information Criterion
BIC:.....Bayesian Information Criterion
Cox PH:.....Cox Proportional Hazard
CPO:.....Conditional Predictive Ordinate
DIC:.....Deviance Information Criteria
HF:.....Heart Failure
INLA:.....Integrated Nested Laplace Approximation
JUMC:.....Jimma University Medical Center
JUSH:.....Jimma University Specialized Hospital
KLD:.....Kullback-Leibler Divergence
MCMC:.....Markov Chain Monte Carlo
NYHA:.....New York Heart Association
OPD:.....Out Patient Department
PIT:..... Probability Integral Transform
SSA:.....Sub-Saharan Africa
TASH:.....Tikur Anbessa Specialized Hospital
USA:.....United State of America
WAIC:.....Watanabe Akaike Information Criterion

Table of Contents

List of Figures

List of Tables

1 INTRODUCTION

1.1 Background of the Study

Heart failure is defined as a clinical syndrome; specifically, failure of the heart to pump blood with normal efficiency, characterized by typical symptoms (shortness of breath, persistent coughing or wheezing, ankle swelling and fatigue) that may be accompanied by the following signs (jugular venous pressure, pulmonary crackles, increased heart rate and peripheral oedema) caused by a structural and functional cardiac abnormality, resulting in a reduced cardiac output and elevated intracardiac pressures at rest or during stress. In addition, HF is a syndrome and not a disease, its diagnosis relies on a clinical examination and can be challenging (Ponikowski *et al.*, 2016; Yancy *et al.*, 2013).

Heart failure is global major cause of death and is a rapidly grown public health issue affecting approximately 40 million individuals worldwide and an estimated 287,000 deaths occurred a year, making it the most quickly growing cardiovascular disorder. It is ever increasing prevalence across developed and developing countries resulted as a complications from an increasing aging population (Vos *et al.*, 2015). In the United State of America, the prevalence of HF is nearly 6.5 million, approximately 960,000 new cases of HF are diagnosed each year, the incidence of HF approaching 21 per 1,000 population and also an estimated 1 in 8 deaths in 2017 (Benjamin *et al.*, 2019). The prevalence of symptomatic HF is estimated to 5% of the population, and the mortality is estimated at about 13% in Europe (Huffman *et al.*, 2013).

In Africa, HF has emerged as a major public health problem, imposing enormous pressure on the health care systems; HF is not a disease by itself patients with HF have other causes of death. The sub-Saharan Africa Survey of HF, a prospective multi center study of HF across the continent, showed that HF is predominantly non-ischemic, most commonly hypertension; HF strikes individuals in sub-Saharan Africa at a much younger age than in the United States and Europe (Damasceno *et al.*, 2012). Similarly, HF is reported to have caused 2.5% of deaths among all age groups in a sampled hospital based mortality in Ethiopia (Misganaw *et al.*, 2014).

In this study, the researchers applied survival analysis since it addresses the limitation of classical regressions like logistic and linear regressions. In survival analysis, all the information of an investigated individual would be used until the last moment in the study, but in classical regressions, cannot be used all the available information. Most medical studies has been used cox regression model for assessing the survival distribution of heart failure patients, while alternative parametric models including exponential, weibull, log-normal, and, log-logistic model has been used to identify the prognostic factors (Giolo *et al.*, 2012; Hailay *et al.*, 2015).

The popular Cox-PH model is to evaluate simultaneously the effect of several factors on survival. The proportional hazards assumption holds with time fixed covariates, cannot specify

the general shape of the hazard curve (Collett, 2015). The parametric survival models could provide a more suitable description of the survival data if one is able to identify the distribution of the survival time (Khanal *et al.*, 2014). The parametric AFT models (i.e exponential, weibull, log-normal and log-logistic) has a more realistic interpretation and provides more informative results than Cox-PH model (Qi, 2009). Epidemiologists have documented several risk factors for the development of HF like as age, hypertension (Sheng *et al.*, 2018) and anemia (Ahmad *et al.*, 2017) were increased risk of mortality among HF patients. Factors such as age, sex, stages of HF, hypertension, anemia, and diabetes mellitus has statistically significant effect on the survival of HF patients (Zeru, 2018).

The parametric survival models play an important role in Bayesian survival analysis, since many Bayesian analysis in practice are carried out using parametric AFT models and provide computational advantages via the implementation of MCMC method with Gibbs sampling of estimation. It generates conclusions based on the synthesis of new information from an observed data and historical knowledge or expert opinion. Historical knowledge from past similar studies can be very helpful in interpreting the results of the current study by Ibrahim *et al.* (2001). The Bayesian approach assumes that the observed data is fixed and that model parameters are random. The prior probability distributions represents a powerful mechanism for incorporating information from previous studies and for controlling confounding (Ibrahim *et al.*, 2011). The Bayesian methods combine objective prior knowledge with the information acquired from the data by using Bayes theorem (Gelman *et al.*, 2014).

In this study, Bayesian Survival Models would be used to identify the factors that affecting the survival time of heart failure patients in JUMC, Jimma, Ethiopia so, the interesting application of MCMC and INLA method of estimation with Bayesian survival models are the most key for the motivation to apply it for the HF data-set under this study. The main aim of this study was to identify the factors that affecting survival time of HF patients. It quests to identify the prognostic factors of HF patients, to determine the best parametric survival models for heart failure data-set, to estimate the survival time of HF patients and to explore the bayesian AFT models using MCMC method with Gibbs sampling and INLA method.

The thesis was structured into five sections. The second section deals with the review of literature on HF patients in JUMC, Jimma, Ethiopia and the rest of the world, whereas third section data description and methodology and variables to be included in the study. Methods of data analysis are also described in this section. Section four reports results from the Bayesian survival analysis and provides discussions. Finally, the last section draws conclusions and makes recommendations for further studies.

1.2 Statement of the Problem

Heart failure is a serious condition during which the guts is unable to pump enough blood to satisfy the requirements of the body (Lloyd-Jones *et al.*, 2002). In approximately all regions of the world, HF is both common and increasing; it is predicted that the number of patients with HF to increase in countries with aging populations and the leading cause of HF death (Benjamin *et al.*, 2019). Studies show that HF is extremely increasing in African countries, including Ethiopia (Damasceno *et al.*, 2012; Misganaw *et al.*, 2014).

Several studies have been conducted in analyzing the survival data using parametric survival models (Hailay *et al.*, 2015), and Cox regression model (Ahmad *et al.*, 2017; Giolo *et al.*, 2012) to estimate the survival time of heart failure patients in hospital based. Many scholars used semi-parametric model, and parametric models (Sheng *et al.*, 2018; Hailay *et al.*, 2015), and Generalized additive model (Berarti and Goshu, 2015) to identify the prognostic factors of heart failure patients in hospital based. However, in JUSH the studies have been conducted on HF patients in hospital based using descriptive statistics (Habte *et al.*, 2010), and logistic regression analysis (Amare *et al.*, 2015). These statistical methodologies are not capable to consider the survival rate of the patients in the hospital and also multi-variable logistic regression does not account the censoring observations, that is, it does not hold for time-to-event data.

Few studies have been conducted based on survival analysis of heart failure patients hospital based with classical approach were fitted to identify factors affecting survival time of heart failure patients (Zeru, 2018; Hailay *et al.*, 2015). Bayesian approach is the possibility of improving the precision of the results by introducing external information in terms of the prior distribution. The advantages of Bayesian approach are to produce more accurate parameter estimates, and higher convergence (Ibrahim *et al.*, 2011). Thus, considering the advantages of Bayesian application is the most key for the motivation to apply it for the heart failure data-set under this study. So, chosen Bayesian Survival Analysis using MCMC and INLA method to analysis heart failure data-set.

Considering HF is a growing problem in the countries hospital based and gaps found with different studies, the researchers was explore the Bayesian Survival Analysis of HF patients in JUMC, Jimma, Ethiopia. Therefore, this study aims to answer the following scientific questions:-

- Which factors significantly affect the survival time of heart failure patients?
- What is the estimated survival time of heart failure patients?
- Which parametric survival models is the most appropriate for analyzing the heart failure data set?

1.3 Objectives of the Study

1.3.1 Generalized Objective

The aim of this study was to identify factors affecting the survival time of heart failure patients in Jimma University Medical Center, Jimma, Ethiopia using Bayesian Survival Models.

1.3.2 Specific Objectives

The specific objectives of this study were:-

- To identify the prognostic factors of heart failure patients.
- To estimate the survival time of heart failure patients.
- To determine the best parametric survival models for heart failure data-set.
- To explore the bayesian accelerating failure time models for the heart failure data-set using MCMC and INLA method.

1.4 Significance of the Study

Studying the survival time of HF patients is a mechanism of overcoming the problem of healthy in the society by identifying factors associated with death. On top of this, the result of this study might be used to improve awareness on the factors that trigger the death of HF patients. It also enables to provide scientific information about the finding to ministry of health in Ethiopia that helps policymakers to enhance the awareness of the society about factors that increase the probability of death due to HF which is protect-able and curable if it is screened and treated in its earlier stage with appropriate treatment.

2 LITERATURE REVIEW

2.1 Overview of Heart Failure

Heart failure is an increasingly common condition resulting in high rates of a major cause of premature morbidity, mortality, poor quality of life and a significant economic burden on national healthcare systems with a lifetime risk of 20% - 46% (Ambrosy *et al.*, 2014). The prevalence of HF is 1% to 2% in the general population of developed countries, and the lifetime risk of growing the disease is at least 1 in 5 among both men and women (Huffman *et al.*, 2013).

Globally, HF represents an increasing issue for health care systems and it is affecting millions of individuals worldwide. It is a common cause of death; is increasing in the prevalence and incidence worldwide (Vos *et al.*, 2015). According to the AHA, the prevalence of HF is nearly 6.2 million (46%) in USA had HF, nearly 800,000 new cases are diagnosed each year, and 78,356 deaths (35,424 males and 42,932 females) are occurred in 2016 since it is expected to rise 8.5 million by 2030 (Benjamin *et al.*, 2019). The study done by Dokainish *et al.* (2017) shows that 5823 HF patients within 1 year: the mortality were; highest in Africa (34%) and India (23%), intermediate in southeast Asia (15%), and lowest in China (7%), South America (9%), and the Middle East (9%).

In the study by Lam *et al.* (2016), shows that approximately 4.2 million people had HF, with 500,000 new cases diagnosed each year, accounting for an incidence of 0.9% in china; the predominant causes of HF were ischemic heart disease and rheumatic valvular heart diseases which causes 38.1% and 29.9% of the total population respectively.

The study done by Bloomfield *et al.* (2013) shows that HF was predominantly a major public health issue in the sub-Saharan Africa. In Ghana, the result shows that among those patients involved in the study, 398 of the patients had HF according to the modified framingham criteria for the diagnosis of HF; giving rise to a prevalence of 76% seen in the study supports the fact that HF is a major contributor to cardiovascular disease burden in SSA. Similar findings have been reported from Cameroon where HF is found to be the fifth to sixth cause of hospital admissions. In other parts of SSA, HF has been found to account 5% - 10% of hospital admissions (Owusu and Boakye, 2013).

Heart failure is an important cardiovascular disease due to its increasing prevalence and high mortality rate and it is found to be the third cause of death following hypertension and stroke in Ethiopia (Misganaw *et al.*, 2012). In the study done by Habte *et al.* (2010) in JUSH for the duration of 5 year period, they found that among 781 HF patients, rheumatic heart disease was 32.8%, hypertensive heart disease was 24.2%, cardiomyopathy was 20.2% and the remaining is other etiologies of HF. The patient registers from 2001 to 2012 among 3282 adult Ethiopian patients in TASH showed 9.1% of the patients had HF (Abdissa *et al.*, 2014).

The results of the Berarti and Goshu (2015) study that out of 263 patients considered in the analysis, 18.6% patients have died of HF while 81.4% were alive. A death proportion for female was 19.6% and that of male patients was 17.5% in Asella Referral Hospital. From the generalized additive model analysis showed that the predictors: age, anemia, diabetes mellitus, and hypertension significantly affect the death status of HF patient.

2.2 Risk Factors for Heart Failure

Heart failure is the non-communicable disease risk factors vary substantially across world regions, with hypertension being highly associated with HF in all regions but most commonly in Latin America, the Caribbean, Eastern Europe, and SSA (Khatibzadeh *et al.*, 2013)

According to a recent study by Benjamin *et al.* (2019), Epidemiologists there are many known HF causes including lifestyle factors, such as cigarette smoking and high blood pressure and other medical conditions factors, such as anemia and diabetes all dramatically increase the likelihood of HF. Similarly, HF has been estimated that a 5% reduction in the number of people with diabetes in the USA would prevent about 30,000 HF cases every year (Avery *et al.*, 2012). Likewise, HF factors includes higher level of anemia; high blood pressure (higher than normal range); aging were the key factors contributing towards increased risk of mortality among HF patients (Ahmad *et al.*, 2017).

Sex:- The study by Ahmad *et al.* (2017) has confirmed that among 299 HF patients, 62 (64%) death were males and 34 (36%) death were females. Thus, the death proportion for female HF patients was lower than that of male HF patients. In addition, the non-parametric kaplan-keier survival curve showed female heart failure patients had higher survival probability than male heart failure patients. However, their survival time of male patients seems lower.

Age:- Heart failure becomes more common with increasing age. A retrospective study conducted in Felege Hiwot referral hospital from 2013 to 2017 indicates that 384 HF patients were considered, the results of Cox-PH model shows HF patients was affect the middle age (49 to 65 years) group as well as the older age (greater than or equal to 65 years) group; since age has significant (p-value = 0.0001) factor of HF patients and the expected survival time of HF patients decreases as they gets older (Zeru, 2018; Adebayo *et al.*, 2017)

Residence:- In the study of Hailay *et al.* (2015) shows that among 147 heart failure patients, the mortality were 33 (77.2%), 13 (22.8%) heart failure patients live in rural area and urban area respectively. However, the results shows the death proportion for HF patients live in urban area was seems lower than that of HF patients live in rural area.

Chronic Kidney Disease:- Chronic kidney disease is an important risk factor for HF pa-

tients. In addition, the results of Cox-PH model shows that the HF patients with chronic kidney disease were less survival time than those without chronic kidney disease and chronic kidney disease has a significant (p-value = 0.0071) factor for heart failure patients (Zeru, 2018).

Anemia:- Anemia is associated with more symptoms, worse functional status, greater risk of HF patients, and reduced survival time (Ponikowski *et al.*, 2016). As study shown by Ahmad *et al.* (2017), among 299 HF patients, 54 (56%) death were HF patients with anemic and 42 (44%) death were HF patients without anemic. In addition, the results of semi-parametric model shows that HF patients with anemic were less survival time than those without anemic and anemia has significant (p-value = 0.0096) factor for HF patients.

History of Heart Failure:- Study by Hailay *et al.* (2015) indicates that among 147 heart failure patients, mortality were 15 (39.5%) was new patients, 31 (28.4%) was HF patient before and 22 (25.6%) was medical OPD. Thus, the death proportion for new HF patients were higher than that of HF patient before and HF patients in medical OPD.

Etiology of Heart Failure:- The main etiology of HF were ischemic heart disease 15.8%, rheumatic valvular heart disease 40.1%, cardiomyopathy 12.5%, hypertensive heart disease 16.0% and the remaining were from other causes constituted the majority of all admissions due to HF (Abebe *et al.*, 2016). In addition, the etiologies of HF in Africa's remain largely hypertensive heart disease, rheumatic valvular heart disease and cardiomyopathy heart disease were the main contributors to the etiology of heart failure in SSA accounting for over 90% of cases (Bloomfield *et al.*, 2013).

Smoking Cigarette:- Smoking is an important risk factor in the development of HF. The study by Ahmad *et al.* (2017) indicates that among 299 HF patients, about 66 (69%) were the death proportion for HF patients without smoking cigarette and 30 (31%) were the death proportion for HF patients with smoking cigarette

Treatments Taken:- Study by Zeru (2018) shows that among those heart failure patients, 261 (68%) used spironolactone, 228 (59.4%) used atorvastatin, 130 (33.9%) used digoxin and 24 (6.2%) used other treatments. Although, survival rates for all patients with HF have improved during the past several decades, the greatest gains have been made in the treatment of patients with heart failure with reduced heart failure. The patients went to medication they had diagnosed in different mechanisms, since method of diagnosis had an effect on the type of treatment they had taken (McManus *et al.*, 2013).

Stages of Heart Failure:- The New York Heart Association functional classification system stratified HF patients in to 4 groups, which was originally developed in 1928 and later by criteria committee in 1964 that described the functional classification system. The NYHA functional classification system was designed for clinical assessment of patients by physicians has been categorized in to 4 classes (I, II, III and IV) (Bennett *et al.*, 2002). As study shown by Zeru

(2018), HF patients with NYHA class, among 384 HF patients 82.6% of patients had NYHA class IV, 11.5% had NYHA class III and 6% of the cases had NYHA class II, this reflects NYHA class IV takes the highest portion for the cause of HF as compared to the other NYHA class. In addition, the results of Cox-PH model shows that stage of HF has statistically significant (p-value = 0.008) factor of HF patients, and the HF Patients with late stage disease had much lower survival rates than those with early stage disease.

Alcohol Consumption:- Study done by Urrutia *et al.* (2016) has confirmed that among 327 heart failure patients considered, 50% of the heart failure patients used alcoholic and 50% of the heart failure patients were non-alcoholic users.

Diabete Mellitus:- Diabetic patients have an increased risk of developing HF. Study done by Fadini *et al.* (2015) shows that among the patients were considered, 58% HF patients without diabetic and 42% with diabetic have an increased mortality. Thus, the results of cox regression model shows that diabetes mellitus has significant (p-value = 0.016) factor of HF patients and expected that heart failure patients with diabetic was less survival time than those with out diabetic Miyagawa *et al.* (2019).

Hypertension:- Hypertension is associated with an increased risk of developing HF patients and it is the dominant cause of HF in Africa, responsible for up to 46% of cases of HF patients (Lip *et al.*, 2015; Callender *et al.*, 2014). The results of semi-parametric model shows hypertension has significant (p-value = 0.0195) factor for HF patients and expected that HF patients with hypertension was less survival time than those without hypertension (Ahmad *et al.*, 2017).

2.3 Overview of Model Used

Survival analysis studies the time duration until the occurrence of an event and time-to-event endpoints are widely used in many medical fields. According to the study conducted by Hailay *et al.* (2015) shows that to estimate the survival time of HF patients using parametric AFT models (i.e exponential, log-normal, weibull, log-logistic) and Cox-PH model in Gondar university hospital, Gonder, Ethiopia.

Ahmad *et al.* (2017) conducted survival analysis of heart failure patients using cox regression model in Allied hospital Faisalabad-Pakistan; the results shows anemia, high blood pressure (higher than normal range), age were significant factor of heart failure patients and also those key factors were contributing towards increased risk of mortality among heart failure patients. Sheng *et al.* (2018) conducted survival analysis of HF patients using Cox proportional hazards model, a total of 1789 patients with heart failure were collected from Shanghai Shuguang Hospital, China.

The study has been done in survival data analysis using Semi-parametric survival model for analyzing the prognostic factors on survival of heart failure patients in Sao Paulo, Brazil (Heart Institute of the Sao Paulo University Medical School) (Giolo *et al.*, 2012). Cox-PH model was the most popular model for analyzing the prognostic factors on survival of gastric cancer patients; when the proportional hazard assumption does not hold, the parametric AFT models estimate the parameter more efficiently than the Cox-PH model. In addition, the parametric AFT models (i.e Weibull, Exponential, Log-normal, and Log-logistic) can provide more accurate estimates than Cox-PH model (Pourhoseingholi *et al.*, 2011).

According to Abrha *et al.* (2018), study the Comparison of Parametric and Bayesian Survival Regression Models in Simulated and HIV Patient Antiretroviral Therapy Data using MCMC method in win-bugs software: Case Study of Alamata Hospital, North Ethiopia; the results Bayesian survival analysis was better performance than classical parametric survival analysis. The efficiency and relevance of the Bayesian survival model with application to clinical research and medicine or public health were supported by the work of different researchers which in fact applied for the different data-set (Khanal *et al.*, 2014; Khan and Khan, 2013; Kumar *et al.*, 2019). The analyzing of heart failure data set based on the theory and algorithms for learning Bayesian networks, the death of a patient can be determined by hypertension, diabetes mellitus, anemia, chronic kidney disease were significant factors of HF patients (Berarti and Goshu, 2015).

The study has been done using data on hormone receptor status for breast cancer cases hospital based by Avi (2017) which indicates that the Bayesian log-normal model was chosen over different Bayesian AFT models using MCMC method with Gibbs sampling in win-bugs software. Akerkar *et al.* (2010) conducted Implementing Approximate Bayesian Inference for Survival Analysis on Kidney Dialysis Patients data-set using INLA method in R-software and MCMC method in win-bugs software uses Gibbs sampling. Bayesian Inference for Survival models using INLA method was supported by the work of the researcher applied for the different data-set (Martino *et al.*, 2011).

3 DATA AND METHODOLOGY

3.1 Data Description

3.1.1 Study Area

The study has been conducted on the data taken from Jimma University Medical Center which is located in Oromia National Regional State, Jimma town 350Km Southwest of Addis Ababa, Ethiopia. JUMC is the only medical center in Jimma zone serving the majority of peoples living in Jimma city and its surrounding.

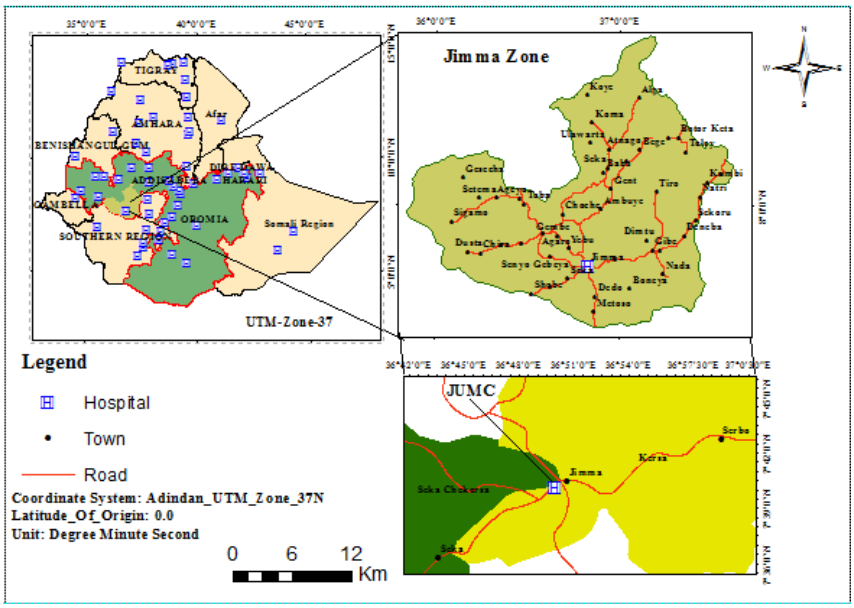


Figure 3.1: Location map of Jimma University Medical Center

Source: The data has been taken from Geographical Information System Masters Student.

3.1.2 Study Design and Population

A retrospective study has been applied to obtain data on HF patients that recorded in JUMC, Jimma, Ethiopia. The population of this study was all HF patients who had been registered at JUMC for 3 years starting from first January, 2016 up to first January, 2019.

The data has been carefully reviewed from the registration log book and patients registration card; any inadequate information encountered was checked from the file and excluded from analysis if proven to be inadequate. Thus, the data has been collected from patient follow up records based on the variables in the study.

3.1.3 Inclusion and Exclusion Criteria

Inclusion criteria:- All person registered with full information including study variables of interest in the registration book or in the chart were considered to be eligible for the study. The patients was to be included in the study they must take treatment at least for one time from the hospital.
Exclusion criteria:- The patient with insufficient information regarding study variables on the registration book or in the card were not eligible. Thus, the HF patients lost from the study without starting any treatment was not included.

3.1.4 Data Collection Methods

Ethical permission has been obtained from the JUMC, Jimma, Ethiopia. Then secondary data were taken based on data existing in the hospital by trained enumerator and the principal investigator using check list (data extraction form).

3.2 Variables in the Study

The response variable was survival time of heart failure patients (in months), which defined as the difference between time of diagnosis and time to one of the events ”death”, ”lost to follow up”, ”dropped out”, ”stopped”, ”transferred out to other health centers or hospitals” occurred. Death was considered to be an event of interest. The status variable was coded as 0 for censored and 1 for death.

Starting Time:- the start time of the interval (in months). Time origin or the beginning of the study, the entry of the survival data would be considered from the day that the heart failure patients starts diagnosis; when the patient first received the treatment.

Ending Time:- the time (in months) at which the event was occurred, when the heart failure patients was died or was lost to follow-up at first January, 2019 (at the end of study). This means that the type of the survival data is right censored.
The factors considered for the purpose of Bayesian survival analysis of HF patients were as follows:-

1. Sex (0=Female, 1=Male)
2. Age (1=$\leq$ 49 years, 2=49-65 years, 3=$\geq$ 65 years)
3. Residence (0=Urban, 1=Rural)

4. History of Heart Failure (1=New, 2=HF patient before, 3=Medical OPD)
5. Etiology of Heart Failure (1=Ischemic heart disease, 2=Rheumatic valvular heart diseases, 3=Cardiomyopathy heart disease, 4=Hypertensive heart disease and 5=Others)
6. Chronic kidney disease (0=No, 1=Yes)
7. Smoking Cigarette (0=No, 1=Yes)
8. Hypertension (0=No, 1=Yes)
9. Diabetes Mellitus (1=Not, 2=Type I and 3=Type II)
10. Treatments (1=Digoxin, 2=Spironolactone, 3=Atorvastatin, 4=Others and 5=Combination of two or more)
11. Anemia (0=No, 1=Yes)
12. Alcohol Consumption (0=No, 1=Yes)
13. Stages of Heart failure (1=I, 2=II, 3=III, 4=IV)

3.3 Method of Data Analysis

3.3.1 Descriptive Statistics

The description of survival data utilizes non-parametric methods to compare the survival functions of two or more groups and kaplan-meier plot(s) would be employed for this purpose (Kaplan and Meier, 1958). The frequency distribution table was used to summarize the data obtained from registration book of patients based on the study variables in JUMC, Jimma, Ethiopia.

3.3.2 Statistical Models

3.3.2.1 Survival Data Analysis

Survival analysis is the statistical analytic method used for modeling and analyzing the data that have a principal end point the time until an event occurs. It is used in analyzing the time-to-event data arises in several applied fields like medicine, public health, epidemiology and etc. Survival data are censored in the sense that they did not provide complete information since subjects of the study may not have experienced the event of interest. Survival analyses consider

a key analytic problem of censoring that occur when some information about individual survival time is known, but not the exact survival time (Aalen *et al.*, 2008).
Censoring is common in survival analysis and it is considered as an important feature of survival data. Survival analysis is well suited for heart failure data-set which are very common in medical research since studies in medical areas have a special feature that follow-up studies could start at a certain observation time and could end before all experimental units had experienced an event.

Right censoring:- occurs to the right of the last known survival time and the observation of patient is terminated before the event occurs. Survival time is said to be right censored when it is recorded from its beginning to a defined time before its end time. This type of censoring is commonly recognized survival analysis and also considered in this study (Klein and Moeschberger, 2006).

Survival Function:- The distribution of survival time is characterized by survival function, probability density function and hazard function. Let T be a random variable associated with the survival times and t be the specified value of the random variable T and f(t) be the underlying probability density function of the survival time T. The survival function, S(t), is given by:-

$$S(t) = P(T \geq t) = 1 - F(t), t \geq 0$$

Where, F(t) is cumulative distribution function, which represents the probability that a subject selected at random will have a survival time less than some stated value t, given by:-

$$F(t) = P(T < t) = \int_0^t f(u)du, t \geq 0$$

The probability density function, f(t), is given by:-

$$f(t) = \frac{d}{dt}F(t) = \frac{-d}{dt}S(t)$$

The hazard function is the instantaneous probability of having an event at time t given that one has survived up to time t (Kleinbaum and Klein, 2012). Hazard function is given by:-

$$\lambda(t) = \frac{f(t)}{S(t)}F(t) = \frac{-d}{dt}lnS(t)$$

The cumulative hazard function is defined as:-

$$\Lambda(t) = \int_0^t \lambda(u)du = -lnS(t), S(t) = e^{-\Lambda(t)}$$

3.3.2.1.1 Estimation of Survival Function

The Kaplan-Meier estimator, non-parametric estimator used to estimate the survival function

with censoring, which is not based on the actual observed event and censoring times, but rather on the order in which events and censored observations occur. It incorporates information from all of the observations available, both censored and uncensored, by considering any point in time as a series of steps defined by the observed survival and censored times (Kaplan and Meier, 1958). Therefore, the Kaplan Meier estimate of the survival function at time t, (S(t)) is given by:-

$$\hat{S}(t) = \prod_{j:\tau_j \leq t} [1 - \frac{d_j}{r_j}]$$

where:- τ_j denote the set of k distinct death time in the observed in the sample, r_j is the number of subjects alive (at risk) just before time t_j (the j^{th} ordered survival time) and d_j denotes the number who died at time t_j.

3.3.2.1.2 Comparison of Survival Function

The Kaplan-Meier plots are used to see whether there is difference in survival times or not between groups of covariate under investigation. But, the plot cannot be used to decide whether the survival time of heart failure patients in each covariate is different or not and log-rank test was used for this purpose (Mantel and Haenszel, 1959).

The test statistic for log rank test is given by:

$$\chi^2_{logrank} = \frac{[\sum(d_{0j} - r_{0j}\frac{d_j}{r_j})]^2}{\sum \frac{r_{0j}r_{1j}d_j(r_j - d_j)}{r_j^2(1-r_j)}} \sim \chi^2_{(1)}$$

where: d_{0j} is the number of failure in j^{th} time of 1^{st} group, d_{1j} is the number of failure in j^{th} time of 2^{nd} group, d_j is the number of failure in j^{th} time ($d_{0j} + d_{1j}$), r_{0j} is the number at risk at j^{th} time of 1^{st} group, r_{1j} is the number at risk at j^{th} time of 2^{nd} group and r_j is the number at risk at j^{th} time ($r_{0j} + r_{1j}$).

The hypotheses to be tested are:-
$H_0 : -$ There is no difference between the survival curves.
$H_1 : -$ There is difference between the survival curves.

3.3.2.1.3 Cox Proportional Hazards Model

The purpose of Cox PH model is to evaluate simultaneously the effect of several factors on survival. In other words, it allows us to examine how specified factors influence the rate of a particular event happening at a particular point in time. This rate is commonly referred as the hazard rate. Thus, the relationship of predictors and the time-to-event in survival analysis is given through hazard function as follows:-

$$\lambda(t|Z) = \lambda_0(t)e^{\beta' Z} = \lambda_0(t)e^{\beta_1 Z_1 + ,\ldots, + \beta_p Z_p}$$

Where
$\lambda(t|Z)$ is the hazard at time t for a subject with a set of predictors $Z_1, ..., Z_p$, $\lambda_0(t)$ is the baseline hazard function, and $\beta_1, ..., \beta_p$ are the model parameters describing the effect of the predictors on the overall hazard.
The interpretation of the Cox PH model can be done using hazards ratios, which defined as the ratio of two individuals with different covariate. Since hazard ratio is time independent we call Cox PH model as Proportional Hazards model. The corresponding survival function for Cox PH model is given by :-

$$S(t|Z) = [S_0(t)]^{e^{\beta_1 Z_1 + ,...,+\beta_p Z_p}} \quad (1)$$

Where $S_0(t)$ is the baseline survival function
In this model, as model in equation (1), no distributional assumption is made for the survival time and the only assumption is that the hazard ratio does not change over time (i.e. proportional hazard) that is why this model is also known as semi-parametric model.

3.3.2.1.4 Accelerated Failure Time Models

Cox-PH model has been extensively used for modeling survival data. The AFT model is parametric model and an alternative model to fit the data when the proportional hazards assumption fails. The AFT models are useful for comparison of survival times whereas the Cox-PH is applicable for comparison of hazards. The key differences between the two models are baseline hazard function and ways of estimating coefficients (Kleinbaum and Klein, 2012).

AFT model is obtained by regressing of the logarithm of the survival time over the covariates and the effect of the explanatory variables on the survival time is directly measured. Some of the standard parametric accelerated failure time models are Exponential, Weibull, Log-normal, and Log-logistic (Dätwyler and Stucki, 2011).

The survival function of an individual with covariate X at time t, in the accelerated failure time model, is the same as the baseline survival function at time t * $exp(\beta_1 X_{1i} +, ..., +\beta_p X_{pi})$, where $\beta_1, ..., \beta_p$ are coefficients of the regression. Thus, $S(t|X) = S_0[t * exp(\beta_1 X_{1i} +, ..., +\beta_p X_{pi})]$ for all t. The effect of the covariates on the survival function is that the time scale is changed by a factor $exp(\beta' X)$, which called accelerated factor (γ). A γ greater than one will increase the survival time, while a factor less than one is harmful on survival time.

The AFT model treats the logarithm of survival time as the response variable and includes an error term that is assumed to follow a particular distribution. The AFT model can be written as follows:-

$$\log(T_i) = \mu + \beta_1 X_{1i} +, ..., +\beta_p X_{pi} + \sigma\epsilon_i$$

This model shows the log linear representation of the AFT model for the i^{th} individual, where: μ is intercept, $log(T_i)$ is the log-transformed survival time, T_i is survival time, $X_1, ..., X_p$ are

explanatory variables with coefficients $\beta_1, ..., \beta_p$, ϵ_i represents residual or unexplained variation in the log-transformed survival times and σ denote scale factor.

Table 3.1: Commonly used distributions and parameters in AFT models

Distribution	f(t)	S(t)	λ(t)
Exponential	$\lambda e^{-\lambda t}$	$e^{-\lambda t}$	λ
Weibull	$\lambda\rho t^{\rho-1}e^{-\rho t}$	$e^{-\rho t}$	$\lambda\rho t^{\rho-1}$
Log-logistic	$\frac{\lambda\rho t^{\rho-1}}{[1+\lambda\rho t^{\rho}]^2}$	$\frac{1}{1+\lambda\rho t^{\rho}}$	$\frac{\lambda\rho t^{\rho-1}}{1+\lambda\rho t^{\rho}}$
Log-normal	$\frac{1}{\sqrt{2\pi}\sigma}exp[\frac{-[logt-\mu]^2}{2\sigma^2}]$	$1-\Phi[\frac{logt-\mu}{\sigma}]$	$\frac{\frac{1}{\sqrt{2\pi}\sigma}exp[\frac{-[logt-\mu]^2}{2\sigma^2}]}{1-\Phi[\frac{logt-\mu}{\sigma}]}$

where:

- λ (lambda) and ρ (rho) denotes scale parameter and shape parameter respectively for Exponential, Weibull, & Log-logistic distribution.
- σ and μ denote scale parameter and shape parameter respectively for Log-normal distribution.
- $\Phi(.)$ denotes the standard normal distribution function.

3.3.2.1.5 Estimation of Parameters in AFT Models

The parameters of Cox-PH model were estimated by partial likelihood function. Partial likelihood is a technique developed to make inference about the regression parameters, β, in the presence of nuisance parameters $\lambda(t|Z)$ (Cox, 1972).

The parameters of AFT models were estimated by maximum likelihood method and Newton-Raphson procedure was used to obtain maximum likelihood parameters estimates (Qi, 2009).

3.3.2.1.6 Model Selection Criterion

Akaikes Information Criterion was used to choose the best AFT model from models like Exponential, Weibull, Log-normal, and Log-logistic model, that fit the data results. It is a method proposed to compare different models, models that are not nested, and/or models with different numbers of parameters (Akaike, 1974). AIC is obtained by:-

$$AIC = 2log(L) + kp$$

where: p is the number of parameters in the model, L is the likelihood, k is a constant and can be seen as a penalty for additional parameters between 2 and 6 (often 2). The recommendation is to use a larger k with small sample. The model which as smallest AIC value is considered as best fitted model.

Bayesian Information Criteria was used to select the best AFT model. The BIC is given by Schwarz (Schwarz *et al.*, 1978). Bayesian Information Criteria is obtained by:-

$$BIC = 2log(L) + kp * log(n)$$

where: p is the number of parameters in the distribution, L is the likelihood, $log(n)$ is the number of observations, k is a constant. The distribution which has the lowest BIC value is considered as best fitted model.

3.3.2.2 Bayesian Survival Analysis

Bayesian approach is preferred over the frequentist approach in survival analysis is that the power of information obtained from the approach is much better as it is the combination of likelihood data and prior information about the distribution of the parameter. In addition, Bayesian approach has several advantages over classical methods, it is well known that survival models are generally quite hard to fit, especially in the presence of complex censoring schemes. With the use of the Gibbs sampler and other MCMC techniques, fitting complex survival models is fairly straightforward, and the availability of software like BUGS eases the implementation greatly (Ibrahim *et al.*, 2001). MCMC methods, has some limitation like the burden of time in approximating the posterior and convergence problem (Brooks and Gelman, 1998; Berger, 2013). As of 2009, the other news was welcomed with very flexible and fast approximation techniques called Integrated Nested Laplace Approximation. Bayesian approach with INLA method is focused on providing a good approximation to the posterior marginal distributions of the parameters in the model (Rue *et al.*, 2009). Bayesian approach is the best method to obtain the appropriate estimates of the model (Gelfand and Mallick, 1995).

The main reasons why one might choose to use Bayesian statistics to produce more accurate parameter estimates. In addition, Bayesian statistics one can incorporate uncertainty about a parameter and update this knowledge through the prior distribution (Depaoli, 2014). Bayesian approach is more useful in clinical data analysis over frequentist and suitable data analysis technique for clinical researchers (Bhattacharjee, 2014). Bayesian approach considers the parameters of the model as random variables and requires that prior distributions be specified for them and data are considered as fixed. In addition, Statistical inferences done by using Bayesian approach is based on the posterior distribution of the model generated (Dezfuli *et al.*, 2009).

Components of Bayesian inference:-

Prior Distribution:- $\pi(\theta)$, It probability distribution used to expresses uncertainty about unknown quantities parameter θ, before the data are taken into account. It is prior distribution, which is a probability distribution that represents the prior information associated with the parameter of interest. A conjugate prior distribution is intended for an unknown parameter which leads to a posterior distribution. Since there is no information available about the parameter, uniform distribution is most commonly used non-informative prior (Ibrahim *et al.*, 2001).

Likelihood:- Ł$(\theta|Data)$, it is a likelihood functions, which is a function that gives the probability of observing of the sample data given the current parameters. For set of unknown parameters in the presence of right censoring it can be written as:

$$L(\theta|Data) = \prod_{i=1}^{n}[f(t_i|X_i;\theta)^{\sigma_i} * S(t_i|X_i;\theta)^{1-\sigma_i}]$$

where

- σ_i is censoring indicator (0=censored and 1=death)
- $f(t_i|X_i;\theta$ and $S(t_i|X_i;\theta)$ are the probability density and survival distributions respectively (Ganjali and Baghfalaki, 2012).

Posterior Distribution:- Posterior Distribution is a combination of prior distribution and likelihood using the Bayes rule, likelihood which includes information about model parameters based on the observed data, and a prior, which includes prior information (before observing the data) about model parameters. It is obtained by multiplying the prior distribution over all parameters, θ, by the full likelihood function $L(\theta|X)$ (Christensen *et al.*, 2011). Given by

$$Posterior = \frac{Likelihood * prior}{\int Likelihood * prior d\theta}$$

Assuming that θ is a random variable and has a prior distribution denoted by $\pi(\theta)$, then posterior distribution, $\pi(\theta|X)$, of θ is given by:

$$\pi(\theta|X) = \frac{L(X|\theta)\pi(\theta)}{\int L(X|\theta)\pi(\theta)d\theta}$$

It is clear that $\pi(\theta|X)$ is proportional to the likelihood multiply by the prior, $\pi(\theta|X) \propto L(X|\theta)\pi(\theta)$, and thus it involves a contribution from the observed data through $L(X|\theta)$ and contribution from prior information quantified through $\pi(\theta)$. The quantity $m(x) = \int L(X|\theta)\pi(\theta)d\theta$ is the normalizing constant of $\pi(\theta|X)$, and is often called the marginal distribution of the data or the prior predictive distribution (Ibrahim *et al.*, 2001).

3.3.2.2.1 Bayesian Accelerated Failure Time Models

Parametric models play an important role in Bayesian survival analysis, since many Bayesian analyses in practice are carried out using parametric models. Parametric modeling offers straightforward modeling and analysis techniques (Ibrahim *et al.*, 2001).

Exponential Model

The Exponential model is the most fundamental parametric model in survival analysis. Suppose we have independent identically distributed survival times $t = (t_1, t_2, ..., t_n)'$ each having an exponential distribution with parameter β, where the censoring indicators $\sigma_i = 0$ if t_i is right censored and $\sigma_i = 1$ if t_i is failure time. Let $f(t_i|\beta)$ =$\beta exp(-\beta t_i)$ denotes the density for t_i, $S(t_i|\beta)$ =$exp(-\beta t_i)$ denotes the survival function for t_i and $X = (n, t, \sigma)$ denotes the observed data. We can write the likelihood function of β as

$$L(\beta|X) = \prod_{i=1}^{n}[f(t_i|\beta)^{\sigma_i} * S(t_i|\beta)^{1-\sigma_i}] = \beta^X exp(-\beta \sum_{i=1}^{n} t_i)$$

where $X = \sum_{i=1}^{n} \sigma_i$. If we assume a p dimensional normal prior for β with mean as μ_0 and covariance matrix as Σ_0 (Ibrahim *et al.*, 2001). Then the posterior distribution of β is given by

$$\pi(\beta|X) \propto L(\beta|X)\pi(\beta|\mu_0, \Sigma_0) \propto \beta^{\sum_{i=1}^{n}\sigma_i} exp(-\beta \sum_{i=1}^{n} t_i)(\beta^{\mu_0 - 1} exp(-\Sigma_0 \beta))$$

$$\pi(\beta|X) = \beta^{\mu_0 + x - 1} exp(-\beta(\Sigma_0 + \sum_{i=1}^{n} t_i))$$

Weibull Model

The Weibull model is perhaps the most widely used parametric survival model. Suppose we have independent identically distributed survival times $t = (t_1, t_2, ..., t_n)'$ each having an Weibull distribution with parameter α, where $\alpha > 0$ and λ. The density for t_i is given by $f(t|\alpha, \lambda) = \alpha t^{\alpha - 1} exp(\lambda - exp(\lambda) t^{\alpha})$. The survival function is given by $S(t|\alpha, \lambda) = exp(-exp(\lambda) t^{\alpha})$.

We can write the likelihood function of (α, λ) as

$$L(\alpha, \lambda|X) = \prod_{i=1}^{n}[f(t_i|\alpha, \lambda)^{\sigma_i} * S(t_i|\alpha, \lambda)^{1-\sigma_i}]$$

$$= \alpha^x exp[\lambda x + \sum_{i=1}^{n}(\sigma_i(\alpha - 1) log(t_i) - exp(\lambda) t_i^{\alpha})]$$

where σ_i is the indicator variable taking value 1 if t_i is failure time and 0 if t_i is right censored. Then the posterior distribution of (α, λ) is given by

$$\pi(\alpha, \lambda|X) \propto L(\alpha, \lambda|X)\pi(\alpha|\alpha_0, k_0)\pi(\lambda|\mu_0, \sigma_0^2)$$

Let $\lambda_i = x_i^{'}\beta$, where x_i is a covariates, β is a parameters. Assuming Normal prior with parameters (μ_0, σ_0^2) for β and gamma prior for α (Ibrahim *et al.*, 2001). The posterior distribution of $(\beta, \alpha|X)$ is given by

$$\pi(\beta, \alpha|X) \propto \alpha^{\alpha_0+X-1} exp[\sum_{i=1}^{n} (\sigma_i + x_i^{'}\beta + \sigma_i(\alpha-1)log(t_i) - t_i^{\alpha} exp(x_i{'}\beta)) -$$

$$k_0\alpha - 1/2(\beta-\mu_0)\Sigma_0^{-1}(\beta-\mu_0)]$$

where $X = (n, t, \sigma)$ denote the observed data for survival model.

Log-Normal Model

Another commonly used parametric survival model is the log-normal model. For this model, we assume that the logarithms of the survival times are normally distributed. If t_i has a log-normal distribution with parameters (μ, σ^2), denoted by $LN(\mu, \sigma^2)$, then the density function is given by

$$f(t_i|\mu, \sigma) = (2\pi)^{-1/2}(t_i\sigma)^{-1} exp(\frac{-1}{2\sigma^2}(log(t_i) - \mu)^2).$$

The survival function is given by

$$S(t_i|\mu, \sigma) = 1 - \phi(\frac{log(t_i) - \mu}{\sigma})$$

Then the likelihood function of (μ, σ) as

$$L(\mu, \sigma|X) = \prod_{i=1}^{n} [f(t_i|\mu, \sigma)^{\sigma_i} * S(t_i|\mu, \sigma)^{1-\sigma_i}]$$

$$= (2\pi\sigma^2)^{-x/2} exp(\frac{-1}{2\sigma^2}) \sum_{i=1}^{n} [\sigma_i(log(t_i)) - \mu)^2]$$

$$* \prod_{i=1}^{n} [t_i^{-\sigma_i}(1 - \phi(\frac{log(t_i) - \mu}{\sigma}))^{1-\sigma}$$

Let $\tau = 1/\sigma 2$ and $\mu_i = x_i^{'}\beta$, where x_i is a covariates, β is a parameters. Assuming Normal prior for β and gamma prior for τ (Ibrahim *et al.*, 2001). The posterior distribution of (β, τ) is given by

$$\pi(\beta, \tau|X) \propto \tau^{\frac{\alpha_0+X}{2}-1} exp[\frac{-\tau^2}{2}(\sum_{i=1}^{n} \sigma_i(log(t_i) - x_i^{'}\beta)^2 + (\beta-\mu_0)^{'}\Sigma_i^{-1}(\beta-\mu_0) + \lambda_0)]$$

$$* \prod_{i=1}^{n} [t_i^{-\sigma_i}(1 - \phi(\tau^{-1/2}(\log(t_i) - x_i^{'}\beta)))^{1-\sigma_i}$$

Log-Logistic Model

Log-logistic model is used parametric survival model. If t_i has a log-logistic distribution with parameters (μ, σ), denoted by $LL(\mu, \sigma)$, then the density function is given by

$$f(t|\mu,\sigma) = \frac{\exp(\frac{\log(t)-\mu}{\sigma})}{\sigma(1-\frac{\exp(\log(t)-\mu)}{\sigma})^2}$$

The survival function is given by

$$S(t|\mu,\sigma) = (1+\exp(\frac{\log(t)-\mu}{\sigma}))^{-1}$$

Then the likelihood function of (μ, σ) as

$$L(\mu,\sigma|X) = \prod_{i=1}^{n}[f(t_i|\mu,\sigma)^{\sigma_i} * S(t_i|\mu,\sigma)^{1-\sigma_i}]$$

Let $\tau = 1/\sigma 2$ and $\mu_i = x_i^{'}\beta$, where x_i is a covariates, β is a parameters. Assuming Normal prior for β and gamma prior for τ (Ibrahim *et al.*, 2001). The posterior distribution of (β, τ) is given by

$$\pi(\beta,\tau|X) \propto \prod_{i=1}^{n}[(\frac{(\tau^{1/2}\exp(\log(t)-x^{'}\beta\tau^{1/2}))}{(1+\exp(\log(t)-x^{'}\beta\tau^{1/2}))^2})^{\sigma_i} * (\frac{1}{(1+\exp(\log(t)-x^{'}\beta\tau^{1/2}))})^{1-\sigma_i}$$

$$*[(2\pi\sigma^2)^{-1/2}\exp(\frac{(\beta-\mu_0)}{2\sigma_0})]$$

3.3.2.2.2 Estimations of Parameter in Bayesian Survival Models

Markov Chain Monte Carlo Method

A Markov Chain Monte Carlo method was used to carry out simulations in estimation of Bayesian parameters. The simulation is used to do the integration numerically rather than analytically by sampling from the posterior distribution of interest even when the form of that posterior has unknown algebraic form (Spiegelhalter *et al.*, 2004).

The basic MCMC methods are Gibbs sampler and metropolis hasting algorithm. Gibbs sampler is the first choice for conditionally conjugate models, the posterior densities are easy to obtain and its also easy to draw samples from each conditional posterior distribution, but metropolis algorithm can be used for models that are not conditionally conjugate. Therefore, In this study the models are conditionally conjugate, because of that Gibbs sampler was used rather than metropolis algorithms (Ibrahim *et al.*, 2001).

Gibbs sampler is an algorithm that sequentially generates samples from a joint distribution of two or more random variables. The means of the posterior samples provide point estimates

for the model parameters, while the standard deviations provide measures of precision. The 95% credible intervals, calculated using the 2.5^{th} and 97.5^{th} percentiles of the posterior samples, provide an alternative indication of the covariates effects along with estimation precision.

Integrated Nested Laplace Approximation Method

The Integrated Nested Laplace Approximation Method was used to estimate the parameters in Bayesian survival model. Survival analysis consists of a great body of work using latent gaussian models. According to Rue *et al.* (2009), INLA computes posterior marginals for each component in the model and it is from these that the posterior expectations and standard deviations can be found. The survival models can be expressed as a latent gaussian model on which the integrated nested Laplace approximations can be applied (Akerkar *et al.*, 2010).

The main aim is to approximate the posterior marginals of the latent field, $\pi(X_i|Y)$ and the posterior marginals of the hyper parameters $\pi(\theta|Y)$ and $\pi(\theta_j|Y)$. The posterior marginals are given by:

$$\pi(X_i|Y) = \int \pi(X_i|\theta)\pi(\theta/Y)d\theta$$

$\pi(\theta_j|Y) = \int \pi(\theta/Y)d\theta\_j$ where $\pi(X_i|Y)$, $\pi(\theta|Y)$ and $\pi(\theta_j|Y)$ are the posterior marginals to be approximated by the latent gaussian models. INLA has a great improvement in speed compared to the other MCMC and also a higher level of accuracy (Akerkar *et al.*, 2010). In addition, INLA provide both extremely fast and very accurate approximations to the posterior marginal through a clever use of laplace approximations and advanced numerical methods and it can be adapted to fit survival models. An R package called R-INLA works as an interface for INLA and it is used just as the other R functions. The INLA programme and the R package for INLA are freely available from http://www.r-inla.org.

3.3.2.2.3 Bayesian Model Selection Criterion

For Bayesian models, we might prefer the Deviance Information Criteria was used for Bayesian survival model comparison. The preferable model is the one with the lowest value of the DIC (Spiegelhalter *et al.*, 2004). We define the deviance of the model as:

$$D(\theta) = -2log(P(y|\theta))$$

In a Bayesian model, this is a random variable so we use the expected deviance $E(D(\theta))$ under the posterior distribution as a measure of fit. For counting the parameters, we introduce the idea of the effective number of parameters:

$$pD = E(D(\theta)) - D(E(\theta)) = \bar{D} - \bar{D}\ (\theta),$$

and the DIC is then:

$$DIC = \bar{D} + pD$$

An alternative is the Watanabe Akaike information criterion, (Watanabe, 2010) which follows a more fully Bayesian approach to construct a criterion. (Gelman *et al.*, 2014) claims the WAIC is preferable to the DIC.

3.3.2.2.4 Bayesian Model Diagnostics

Bayesian Cox-Snell residual plot

Model checking and adequacy play an important role in models for survival data. In Bayesian analysis, Chaloner (1991) defined the Bayesian version of the residuals:-

$$r_{ci} = H_i(t_i, \theta|x_i) = -\log(S(t_i, \theta|x_i)), i = 1, 2, ..., n$$

Each r_{ci} is just a function of unknown parameters, and posterior distribution is therefore straightforward to calculate. The posterior mean or median of the r_{ci}'s can be calculated and evaluated. More simply, Wakefield (2013) suggested that one could substitute the posterior mean or median of θ directly to obtain the approximate Bayesian residuals.

If the model fits well and the posterior mean is close to the true value, then the posterior mean or median of r_{ci}s should look like a censored sample from a unit exponential distribution. In order to check whether the r_{ci}s behave as a sample from a unit exponential distribution, we could compute the Nelson Aalen estimator of the cumulative hazard rate of r_{ci}s, which is defined as $H(t) = \sum_{r_{ci}} \frac{d_i}{m_i}$ with d_i the number of events at r_{ci} and m_i the total individuals at risk just prior to time r_{ci} .

If the exponential distribution fits the residuals, the estimate should be very close to the true cumulative hazard rate of the unit exponential model, that is, $H(t) = t$. Hence, one could check the so-called Cox-Snell residual plot, a plot of the residual r_{ci} versus its Nelson Aalen estimate (r_{ci}). If a model fits well, this plot should follow a straight line through the origin with a slope of 1.

Predictive Distribution

The idea on how to classify the sample data and techniques undergone for the application of criticism for predictive distribution (Piironen and Vehtari, 2017). The predictive distribution for observation x is

$$p(x|y) = \int_\theta p(x|\theta)p(\theta|y)d\theta$$

To evaluate the goodness of the model in this perspective and whether there are any outliers, the conditional predictive ordinates and probability integral transform values can be examined. The conditional predictive ordinate introduced by Pettit (1990), is defined as:

$$CPO_i = P(y_i|y\_i)$$

where: y_i means all the data except for the i^{th} observation. If a choice of model leads to many small CPOs, the model may be flawed in some way. This is similar to investigating the residuals

of a model, whereas a residuals only measure the distance between the data and the model, the CPO measures the probability of that distance, i.e. this takes the distribution of the model into account. In addition, the sum of the CPO values is a measure of fit (Held *et al.*, 2010).

The probability integral transform is similar to the CPO statistic. It was introduced by (Dawid, 1984) and is defined as:-

$$PIT_i = P(Y_i < y_i|y\_i)$$

The observations were drawn from the predictive distribution, If the model represents the observation well, the distribution of the different values should be close to a uniform distribution between 0 and 1 (Martino *et al.*, 2011).

3.4 Ethical Consideration

The Research Ethics Review Board of Jimma University would provide an ethical clearance for the study. The data has been collected after written permeation was obtained from Jimma University Medical Center and department of statistics write an official co-operation letter to the Hospital for the permeation.

The data has been carefully reviewed from the registration log book and patients registration card. Confidentiality of any information related to the patients and their clinical history has been maintained by keeping both the hardcopy and softcopy of every collected data in a locked cabinet and password secured computer. Only the researcher would access to the de-identified data that has been kept in a secure place. All data has been coded with numbers and without personal identifiers. All analysis has been on de-identified and coded data. During the study, there is no contact between the patients and the researcher. The study is non-invasive and without any harm to the patients. Then, the data obtained from the hospital has been secured.

3.5 Statistical Software Used

The statistical software used were:-

- SPSS version 21 used for data coding and entry.
- WinBUGS version 14 used for data analysis in MCMC method.
- R version 3.6.1 used for others data analysis.

4 RESULTS AND DISCUSSIONS

4.1 Results

4.1.1 Descriptive Summaries

The data for this study has been taken from 409 patients receive treatments for HF, at least one time, at Jimma University Medical Center, Jimma, Ethiopia in the year first January, 2016 to first January, 2019. The minimum and maximum event time observed from HF patients follow up where 6 and 36 months respectively. Among those HF patients, about 59.90% were censored (right censored) and remaining 40.10% were died. Fifty percent of HF patients who receive treatments, survived 31 months or above it. Almost half, 52.81%, of the HF patients were female and the remaining were male during the follow up study. However, their survival time of male patients seems lower.

Most of HF patients, about 64.79% live in rural area and the remaining were in urban area. The survival time of HF patients seems less as they gets older. About 20.05%, 22.25%, 23.72%, 25.43% and 8.55% of HF patients were ischemic heart disease, rheumatic valvular heart diseases, cardiomyopathy heart disease, hypertensive heart disease and other disease respectively. By observing the smoking status of HF patients, most HF patients were, 74.82%, non-smokers and the death proportion seems highest for those HF patients who smoker which was 54.88% compared to non-smokers which was 45.12%. About, 64.55% of HF patients were not alcohol users and 35.45% were alcohol users.

Moreover, about 19.08% HF patients treated in the hospital with a combination of two or more treatments and 19.32% HF patients take digoxin. In addition, the remaining 24.2%, 25.18% and 11.49% of HF patients treated with spironolactone, atorvastatin and other treatments respectively. About 58.19%, 13.69% and 28.12% HF patients were non-diabetic, type I diabetes mellitus and type II diabetes mellitus respectively.

By observing the chronic kidney disease of HF patients, about 30.32% and 69.68% were HF patients with chronic kidney disease and without chronic kidney disease respectively, in which HF patients with chronic kidney disease seems lower survival time. Most of HF patients has no hypertension, 60.64%, and the remaining has hypertension.

Looking the stage at which the HF patients goes to the hospital for treatment, about 36.92%, 28.61%, 19.07% and 15.4% were in stage IV, in stage III, in stage II and in stage I respectively. Most of, about 54.87% death, HF patients go for treatment into the hospital at later stage and their survival time seems low at this stage.

Table 4.1: Descriptive summaries of patient's for HF data set

		Patients Status		
Covariates	Categories	No of Censored(%)	No of Death(%)	Total
Sex	Female	148(60.41)	68(41.46)	216(52.81)
	Male	97(39.59)	96(58.54)	193(47.19)
Age	≤ 49	98(40.0)	17(10.36)	115(28.12)
	49-65	78(31.84)	53(32.32)	131(32.03)
	≥ 65	69(28.16)	94(57.32)	163(39.85)
Alcohol	No	172(70.20)	92(56.10)	264(64.55)
	Yes	73(29.80)	72(43.90)	145(35.45)
Residence	Urban	97(39.59)	47(28.66)	144(35.21)
	Rural	148(60.41)	117(71.34)	265(64.79)
History of HF	New	74(30.20)	46(28.05)	120(29.34)
	HF patient before	84(34.29)	79(48.17)	163(39.85)
	Medical OPD	87(35.51)	39(23.78)	126 (30.81)
CKD	No	224(91.43)	61(37.20)	285(69.68)
	Yes	21(8.57)	103(62.80)	124(30.32)
Hypertension	No	203(82.86)	45(27.44)	248(60.64)
	Yes	42(17.14)	119(72.56)	161(39.36)
Anemia	No	205(83.67)	62(37.80)	267(65.28)
	Yes	40(16.33)	102(62.20)	142(34.72)
DM	Not	204(83.26)	34(20.73)	238(58.19)
	Type I	18(7.35)	38(23.17)	56(13.69)
	Type II	23(9.39)	92(57.0)	115(28.12)
Etiology of HF	IHD	62(25.31)	20(12.19)	82(20.05)
	RVHD	47(19.18)	44(26.83)	91(22.25)
	Cardiomyopathy	55(22.45)	42(25.61)	97(23.72)
	HHD	55(22.45)	49(29.88)	104(25.43)
	Others	26(10.61)	9(5.49)	35(8.55)
Cigarette	No	232(94.69)	74(45.12)	306(74.82)
	Yes	13(5.31)	90(54.88)	103(25.18)
Treatments	Digoxin	48(19.59)	31(18.90)	79(19.32)
	Spironolactone	65(26.53)	34(20.73)	99(24.20)
	Atorvastatin	62(25.31)	41(25.0)	103(25.18)
	Others	27(11.02)	20(12.20)	47(11.49)
	Combination ≥ 2	43(17.55)	38(23.17)	81(19.80)
Stages	I	58(23.67)	5(3.05)	63(15.40)
	II	61(24.90)	17(10.37)	78(19.07)
	III	65(26.53)	52(31.71)	117(28.61)
	IV	61(24.90)	90(54.87)	151(36.92)

No: Number; %: percent; DM: Diabetes mellitus; CKD: Chronic kidney disease; IHD: Ischemic heart disease; RVHD: Rheumatic valvular heart disease; and HHD: Hypertensive heart disease.

Source: JUMC, Jimma, Ethiopia; from first January, 2016 to first January, 2019.

The Kaplan Meier Estimate for some Covariate:-

Figure 4.1 (a) below, the overall survival rate at the end of the first year was almost 93.1%, and the overall survival rate at the end of 34 months in this study was 31%, 95% confidence interval was (23.9%, 40.2%).

Figure 4.1 (b) below, indicated that HF patients whose age was below 49 years were at a higher probability of surviving than patients whose age was 49 to 65 years and also patients whose age was greater than or equal to 65 years. The probability of surviving becomes less for the patients whose age was greater than or equal to 65 years.

Figure 4.2 (c) below, shows that HF patients with stage I were higher chance of surviving than other stages. The survival curve for patients with stage II was above the survival curve of those patients with stage III and stage IV. The probability of surviving becomes less for the HF patients with stage IV.

Figure 4.2 (d) below, one sees that the HF patients without hypertension was at a higher chance of survival than the HF patients with hypertension.

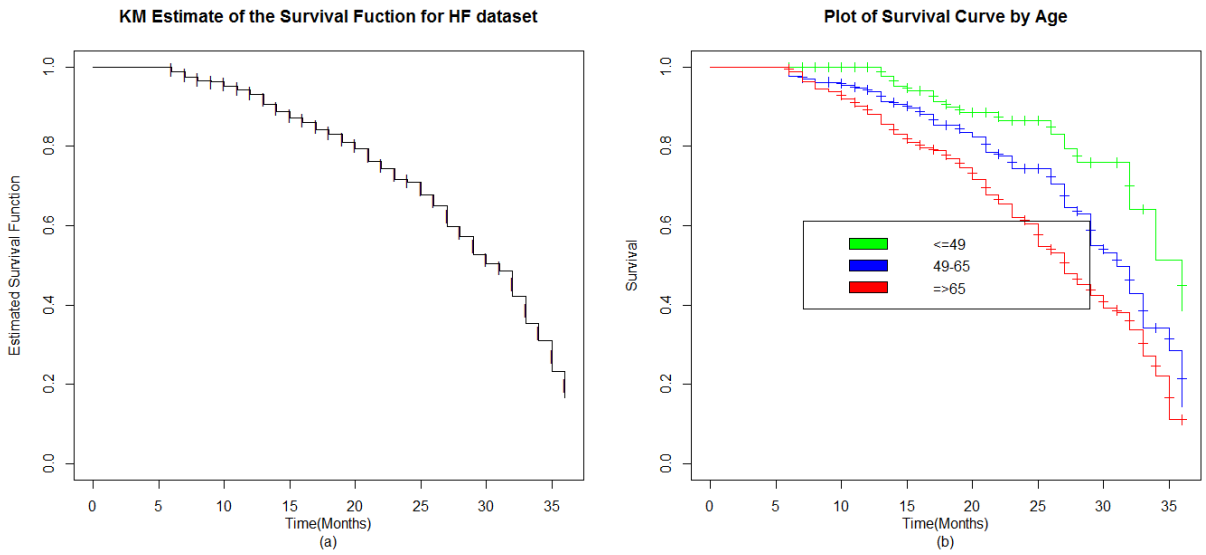


Figure 4.1: Kaplan-Meier estimates of the survival curves of HF data set for (a): Overall survivor function for heart failure patients in the study, (b): Age group

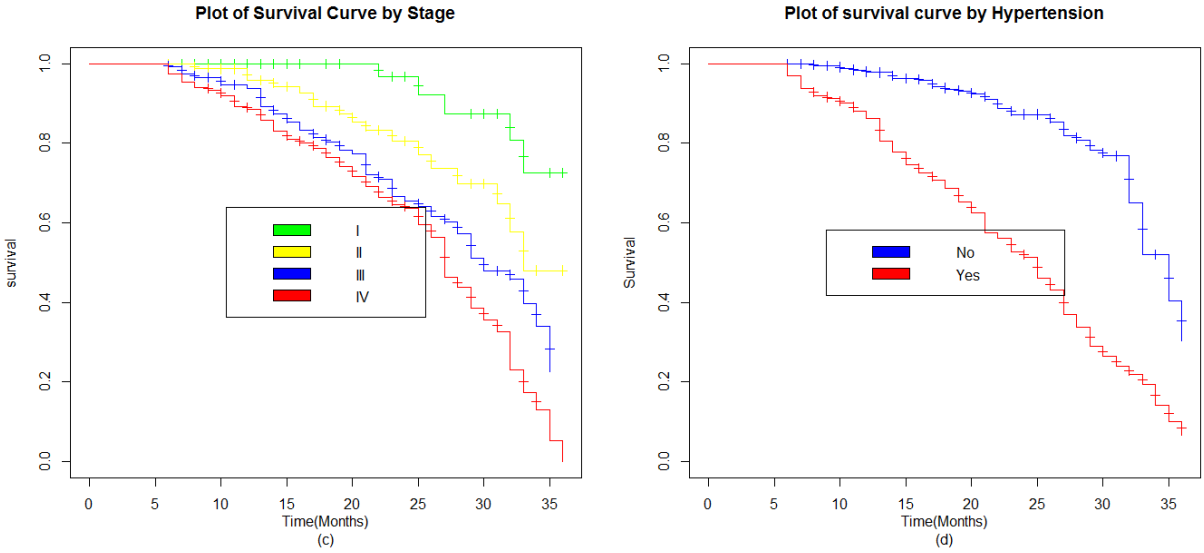


Figure 4.2: Kaplan-Meier estimates of the survival curves of HF data set for (c): Stages, (d): Hypertension

4.1.2 Comparison of Survival time of HF patients

From Table 5.1 in appendix, indicating log-rank for the covariates, there is difference in the survival time of HF patients with respect to sex, age, chronic kidney disease, diabetes mellitus, etiology of heart failure, hypertension, anemia, smoking cigerette and stages of heart failure since there corresponding p-value was smaller than common alpha level of significance (5%). On the other hand, the remaining covariates like residence, treatment taken, alcohol consumption and history of heart failure were not statistically significant.

4.1.3 Cox Proportional Hazard Model

To determine the covariates which are associated with the observed time to death of HF patients, the variable with p-value less than or equal to 25% in uni-variable analysis were considered for multi-variable analysis (Hosmer Jr *et al.*, 2008). Then, the multi-variable Cox proportional hazard model was fitted including all the potential covariates that were significant in the uni-variable at 25% level of significance. The fitted Cox-PH model, as shown from Table 5.3 in the appendices, shows that the survival time of HF patients significantly affected by sex, age, chronic kidney disease, diabetes mellitus, etiology of heart failure, hypertension, anemia, smoking cigarette and stages of heart failure since there corresponding p-value was smaller than common alpha level of significance (5%).

4.1.3.1 Checking the Assumption of Cox-PH

As it can be shown in the Table 4.2 below, the p-values for alcohol consumption, chronic kidney disease and anemia are less than common (5%) level of significance using correlation test (rho) and these shows as the assumption of Cox-PH model was not valid for HF data set. In addition, by looking for global test the assumption of Cox-PH fails since the test result was significant.

Table 4.2: Shows test of assumption in Cox-PH model

Covariates	rho	Chi-square	P-value
Age	0.0125	0.0269	0.86981
Sex	-0.0349	0.2344	0.62829
History of HF	0.0367	0.2201	0.63896
Alcohol consumption	0.1639	4.5498	0.03292
Hypertension	-0.0639	0.7364	0.39081
Chronic kidney disease	-0.1970	7.9647	0.00477
Etiology of HF	-0.1205	2.4705	0.11600
Smoking cigarette	0.0907	1.6229	0.20269
Diabetes mellitus	-0.0179	0.0573	0.81077
Anemia	0.2030	8.2169	0.00415
GLOBAL TEST	NA	23.4782	0.00911

Source: JUMC, Jimma, Ethiopia; from first January, 2016 to first January, 2019.

4.1.4 Accelerated Failure Time Models

Proportionality assumption of cox proportional hazard model was not fulfilled for HF data-set. In this case parametric AFT models was used for the HF data set. To determine the covariates which are associated with the observed time to death of HF patients, the variable with p-value less than or equal to 25% in uni-variable analysis were considered for multi-variable analysis (Hosmer Jr *et al.*, 2008). For the survival time of HF data-set, AFT model such as exponential, weibull, log-normal and log-logistic distribution were fitted by including all the covariates that were significant in the uni-variable at 25% level of significance.

To compare the efficiency of these different models AIC and BIC were used and the one with the smallest value and seems best fit. All AFT models and the corresponding AIC and BIC values were displayed in Table 5.5 of the appendices. As it can be observed from this table, the AIC and BIC value for log-normal model is less than all proposed AFT models. Thus,

log-normal AFT model (AIC = 1298.620 and BIC = 1365.253) found to be the best fit for the survival time of HF patients data set from the given alternative, the step-wise procedure was applied to select the significant covariates. The results for the log-normal AFT model as shown from Table 5.4 in the appendices shows the survival time of HF patients statistically significantly affected by age, chronic kidney disease, diabetes mellitus, etiology of HF, hypertension, anemia, smoking cigarette and stages of HF since there corresponding p-value was smaller than common alpha level of significance (5%).

4.1.5 Bayesian Accelerated Failure Time Models

As it can be shown in the Table 4.2, the assumption of Cox-PH model was not valid for heart failure data set; in this case parametric AFT models was used for HF data set. For the HF data set, the time t_i where i = 1, 2, ..., 409 of heart failure patients. Given that $\beta = (\beta_0, \beta_1, ..., \beta_p)^{'}$ is the vector of coefficients of the covariates considered for analysis, β_0 is the intercept and p the number of covariates (p = 13), we assume that all these coefficients have a normal prior with mean 0 and variance 1000. We assume that scale parameter have a gamma prior with shape parameter 1 and inverse scale parameter 0.001, this prior was used for Weibull, Log-normal and Log-logistic distribution for INLA and MCMC method of estimation (Akerkar *et al.*, 2010; Ibrahim *et al.*, 2001). In this simulation study of Bayesian inference using MCMC, the Gibbs sampler algorithm was implemented with 40,000 iterations in three different chains, 15,000 burn-in terms discarded, as to obtain 60,000 samples for full posterior distribution.

From Table 4.3 below, shows that analysis of HF data set for model comparison using MCMC and INLA method. To compare the efficiency of these different models DIC and WAIC were used and the one with smallest value and seems best fit. Accordingly, Bayesian log-normal AFT model using INLA (DIC = 1297.84; WAIC = 1297.47) and Bayesian log-normal AFT model using MCMC Method (DIC = 1321.73) found to be the best for survival time of HF patients data-set from the give alternative. The Bayesian log-normal AFT model using INLA method has the smallest DIC than using MCMC method.

Table 4.3: The comparisons of Bayesian AFT model using MCMC and INLA methods

	MCMC		**INLA**		
Distributions	**pD**	**DIC**	**pD**	**DIC**	**WAIC**
Exponential	15.764	1497.050	12.24	1400.62	1522.88
Log-Normal	16.796	**1321.730**	17.06	**1297.84**	**1297.47**
Weibull	16.360	1411.070	11.63	1389.20	1383.43
Log-logistic	16.424	1386.650	7.59	1326.88	1326.39

Source: JUMC, Jimma, Ethiopia; from first January, 2016 to first January, 2019.

The Results of Bayesian Log-normal AFT model using MCMC and INLA method

As shown in Table 4.4 below, the Bayesian log-normal AFT model using INLA method was better than Bayesian log-normal AFT model using MCMC method due to smaller standard error and narrow credible interval for all significant parameters in HF data set. The standard error of MCMC results were larger than INLA method with wider credible interval. Thus, INLA method provides a faster and more accurate alternative to simulation based MCMC schemes with higher precision.

Table 4.4: The results of Bayesian log-normal AFT model using MCMC and INLA method

Method of Estimations							
		MCMC			INLA		
Cova	Categ	PM	Sd	CrI	PM	Sd	CrI
	Intercept	5.18	0.2417	[4.397, 5.645]	4.95	0.221	[4.541, 5.409]
Age	≤ 49	Ref			Ref		
	49-65	-0.27	0.123	[-0.503, -0.032]	-0.25	0.115	[-0.488, -0.036]
	≥ 65	-0.35	0.121	[-0.561, -0.113]	-0.33	0.110	[-0.557, -0.125]
Hyper	No	Ref			Ref		
	Yes	-0.31	0.107	[-0.465, -0.129]	-0.30	0.076	[-0.452, -0.153]
CKD	No	Ref			Ref		
	Yes	-0.40	0.101	[-0.541, -0.206]	-0.38	0.075	[-0.537, -0.244]
Etio	IHD	Ref			Ref		
	RVHD	-0.31	0.126	[-0.545, -0.057]	-0.30	0.116	[-0.533, -0.076]
	Cardio	-0.16	0.122	[-0.390, 0.064]	-0.15	0.113	[-0.382, 0.063]
	HHD	-0.27	0.123	[-0.504, -0.031]	-0.25	0.115	[-0.486, -0.035]
	Others	-0.39	0.183	[-0.713, -0.059]	-0.38	0.160	[-0.693, -0.066]
Stages	I	Ref			Ref		
	II	-0.43	0.207	[-0.817, -0.321]	-0.40	0.190	[-0.782, -0.038]
	III	-0.42	0.180	[-0.793, -0.080]	-0.42	0.176	[-0.781, -0.090]
	IV	-0.51	0.175	[-0.879, -0.178]	-0.50	0.173	[-0.857, -0.180]
Ciga	No	Ref			Ref		
	Yes	-0.16	0.088	[-0.362, -0.080]	-0.15	0.073	[-0.300, -0.014]
DM	Not	Ref			Ref		
	TypeI	-0.24	0.120	[-0.468, -0.029]	-0.23	0.100	[-0.431, -0.036]
	TypeII	-0.43	0.11	[-0.605, -0.256]	-0.42	0.086	[-0.593, -0.255]
Anem	No	Ref			Ref		
	Yes	-0.16	0.085	[-0.297, -0.011]	-0.15	0.072	[-0.298, -0.013]

Cov: Covariates; Categ: Categories; PM: Posterior mean; CrI: Credible Interval; Sd: standard deviation; Ref: Reference; Hyper: Hypertension; CKD: Chronic kidney disease; Etio: Etiology of HF; Anem: Anemia; DM: Diabete mellitus and Ciga: Smoking cigarette.
Source: JUMC, Jimma, Ethiopia; from first January, 2016 to first January, 2019.

4.1.5.1 Bayesian Log-Normal AFT Model using INLA method

For the HF data set, the time t_i where i = 1, 2, ..., 409 of heart failure patients follows a log-normal distribution. The final model is defined as:
$\log(T_i) = \beta_0 + \beta_{12}$ (49-65) years age group $+\beta_{13}(\geq 65)$ years age group $+\beta_{22}$ had hypertension $+\beta_{32}$ had CKD $+\beta_{42}$ RVHD $+\beta_{43}$ HHD $+\beta_{44}$ Others $+\beta_{52}$ Stage II $+\beta_{53}$ Stage III $+\beta_{54}$ Stage IV $+\beta_{62}$ Smokers $+\beta_{72}$ TypeI DM $+\beta_{73}$ TypeII DM $+\beta_{82}$ Anemia $+\sigma\epsilon_i, i = 1, ..., 409$.
where:- CKD: Chronic kidney disease, DM: Diabetes mellitus, HHD: Hypertensive heart disease, RVHD: Rheumatic valvular heart disease.

The final results for the Bayesian log-normal AFT model using INLA method was shown as in Table 4.5 and as this result shows the survival time of HF patients statistically significantly affected by age, chronic kidney disease, diabetes mellitus, etiology of heart failure, hypertension, anemia, smoking cigarette and stages of heart failure.

Table 4.5: Indicating the results for Bayesian log-normal AFT model using INLA method

Covariates	Categories	Pmean	sd	Median	CrI	Mode	Kld
	Intercept	4.953	0.221	4.945	[4.541, 5.409]*	4.929	0
Age	≤ 49	Ref					
	49-65	-0.258	0.115	-0.256	[-0.488, -0.036]*	-0.253	0
	≥ 65	-0.336	0.110	-0.335	[-0.557, -0.125]*	-0.331	0
Hyper	No	Ref					
	Yes	-0.301	0.076	-0.300	[-0.452, -0.153]*	-0.299	0
CKD	No	Ref					
	Yes	-0.389	0.075	-0.388	[-0.537, -0.244]*	-0.387	0
EthiHF	IHD	Ref					
	RVHD	-0.302	0.116	-0.302	[-0.533, -0.076]*	-0.300	0
	Cardiomy	-0.158	0.113	-0.158	[-0.382, 0.063]	-0.157	0
	HDD	-0.258	0.115	-0.257	[-0.486, -0.035]*	-0.255	0
	Others	-0.381	0.160	-0.381	[-0.693, -0.066]*	-0.382	0
Scigarette	No	Ref					
	Yes	-0.156	0.073	-0.155	[-0.300, -0.014]*	-0.154	0
Stages	I	Ref					
	II	-0.400	0.190	-0.397	[-0.782, -0.038]*	-0.389	0
	III	-0.423	0.176	-0.419	[-0.781, -0.090]*	-0.410	0
	IV	-0.506	0.173	-0.501	[-0.857, -0.180]*	-0.492	0

$\star$ is indicated statistically significant. Pmean: Posterior mean; CrI: Credible Interval; Sd: standard deviation; Kld: Kullback-leibler divergence; Hyper: Hypertension; CKD: Chronic kidney disease; EthiHF: Etiology of heart failure; IHD: Ischemic heart disease; RVHD: Rheumatic valvular heart disease, and Scigarette: Smoking cigarette.

Extension of Table 4.5, Indicating the results for Bayesian log-normal AFT model using INLA method

Covariates	Categories	Pmean	sd	Median	CrI	Mode	Kld
DM	Not	Ref					
	Type I	-0.232	0.10	-0.231	[-0.431, -0.036]*	-0.230	0
	Type II	-0.422	0.086	-0.421	[-0.593, -0.255]*	-0.419	0
Anemia	No	Ref					
	Yes	-0.154	0.072	-0.153	[-0.298, -0.013]*	-0.152	0
Tau parameter	for log-normal	4.30	0.497	4.28	[3.38, 5.33]*	4.24	-

$\star$ is indicated statistically significant. Pmean: Posterior mean; CrI: Credible Interval; Sd: standard deviation; Kld: Kullback-leibler divergence; and DM: Diabetes mellitus.
Source: JUMC, Jimma, Ethiopia; from first January, 2016 to first January, 2019.

Interpretation of Bayesian log-normal AFT model using INLA method

From Table 4.5, the final model were interpreted using acceleration factor, 95% credible interval of Bayesian accelerated failure time estimated values. The estimated acceleration factor is defined as $\gamma = [\exp(\hat{\beta})] = [\exp(\text{posterior mean})]$.

Under the Bayesian log-normal AFT model, keeping the effect of other factors constant, the estimated acceleration factor for age group of HF patients were 49 to 65 and greater than or equal to 65 years old are estimated to be 0.7726 with [95% CrI: 0.6138, 0.9646] and 0.7146 with [95% CrI: 0.5729, 0.9875] respectively. Thus, the expected survival time of HF patients decrease by 22.74% and 28.54% for HF patients aged group 49 to 65 and 65 or above 65 years older respectively as compared to HF patients of aged group 49 or below 49 years (Reference). The 95% Credible Interval for acceleration factor of both age group did not include one which implies that both age group has significant effect on the survival time of HF patients.

Looking for chronic kidney disease, keeping the effect of other factors constant, the estimated acceleration factor of HF patients with chronic kidney disease is estimated to be 0.6777 with [95% CrI: 0.5844, 0.7835] which implies the expected survival time decreases by 32.23% than HF patients without chronic kidney disease. The 95% CrI for acceleration factor of HF patients with chronic kidney disease did not include one which implies that HF patients with chronic kidney disease has significant effect on the survival time of HF patients.

By observing hypertension, keeping the effect of other factors constant, the estimated acceleration factor for HF patients with hypertension is estimated to be 0.74 with [95% CrI: 0.5844, 0.7834] in which the expected survival time is 26% decrease as compared to HF patients with out hypertension (Reference). The 95% credible interval for acceleration factor of HF patients with hypertension did not include one which implies that HF patients with hypertension has significant (in the Bayesian sense) effect on the survival time of HF patients.

On other hand, keeping the effect of other factors constant, the estimated acceleration factor for HF patients who were smoking cigarette is estimated to be 0.8555 with [95% CrI: 0.7408, 0.986]. The 95% credible interval for the acceleration factor of HF patients who were smoking cigarette did not include one. Thus, HF patients who were smoking cigarette has significant effect on the survival time of patients and the expected survival time of HF patients who were smoking cigarette decreases by 14.45% than not smoking cigarette.

Regarding to etiologies of HF, keeping the effect of other factors constant, the estimated acceleration factor for etiologies of HF were rheumatic valvular heart disease, hypertensive heart disease and other heart disease are estimated to be 0.7393 with [95% CrI: 0.5868, 0.9268], 0.772 with [95% CrI: 0.615, 0.965] and 0.683 with [95% CrI: 0.5, 0.936] respectively. Thus, the expected survival time of HF patients decreases by 27.07% were rheumatic valvular heart disease, 22.8% were hypertensive heart disease and 31.7% were other heart disease as compared to ischemic heart disease of HF patients. The 95% CrI for acceleration factor of HF patients for etiology of HF were rheumatic valvular heart disease, hypertensive heart disease and other heart disease did not include one which implies that etiology of HF were rheumatic valvular heart disease, hypertensive heart disease and other heart disease has significant effect on the survival time of HF patients, while the etiology of HF were cardiomyopathy heart disease has not significant effect on the survival time of HF patients.

Moreover, for diabetes mellitus, keeping the effect of other factors constant, the estimated acceleration factor for HF patients with type I diabetic and type II diabetic are estimated to be 0.793 with [95% CrI: 0.649, 0.964] and 0.655 with [95% CrI: 0.552, 0.774] respectively. Thus, the expected survival time of HF patients decreases by 20.7% for type I diabetic and 34.5% for type II diabetic as compared to HF patients non-diabetic (Reference). The 95% credible interval for acceleration factor of HF patients with both type of diabetes did not include one which implies that HF patients with both type of diabetes has significant effect on the survival time of HF patients.

Looking for anemia, keeping the effect of other factors constant, the estimated acceleration factor of HF patients with anemia is estimated to be 0.857 with [95% CrI: 0.742, 0.987] which implies the expected survival time decreases by 14.3% than HF patients without anemia. The 95% CrI for acceleration factor of HF patients with anemia did not include one which implies that HF patients with anemia has significant effect on the survival time of HF patients.

Finally, observing stages of HF, keeping the effect of other factors constant, the estimated acceleration factor for stage II, III and IV of HF patients are estimated to be 0.67 with [95% CrI: 0.457, 0.962], 0.655 with [95% CrI: 0.457, 0.913] and 0.602 with [95% CrI: 0.424, 0.835] respectively. Thus, the expected survival time of HF patients decreases by 33%, 34.5% and 39.8% for stage II, III and IV of HF patients respectively as compared to stage I. The 95%

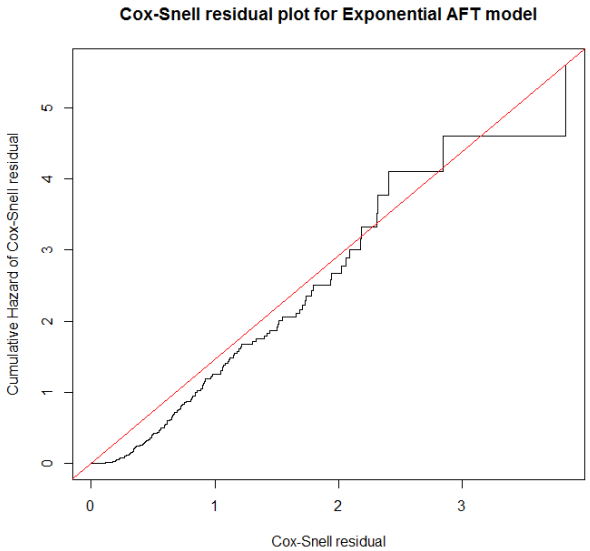
credible interval for acceleration factor of HF patients having stage II, III and IV did not include one this indicates that stage II, III and IV has significant effect on the survival time of heart failure patients.

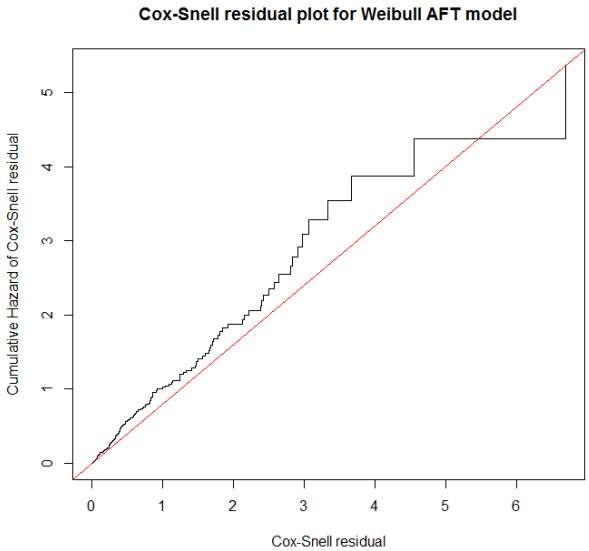
From Table 4.5, the Kullback-leibler divergence values for all significant parameters in Bayesian log-normal AFT model were 0, and thus, small values indicate that the posterior distribution was well approximated by a normal distribution. A simplified laplace approximation was the most efficient algorithm with improved efficiency and results to higher computation speed. In addition, there is no need to perform the more computationally intense full Laplace approximation. Therefore, simplified laplace approximation was appropriate.

4.1.5.2 Bayesian Model Diagnostics

Bayesian Cox Snell Residual Plots

By observing Bayesian cox-snell residual plots figure below, the Bayesian log-normal AFT model best fit HF data-set among the five models, since the plot of Cox-Snell residuals against cumulative hazard function of residuals was approximately a straight line with slope one and Bayesian cox-snell residual plot for Bayesian log-normal AFT model were nearest to the line through the origin. In addition, the plot also indicated that Bayesian log-normal model describes the HF data-set well.





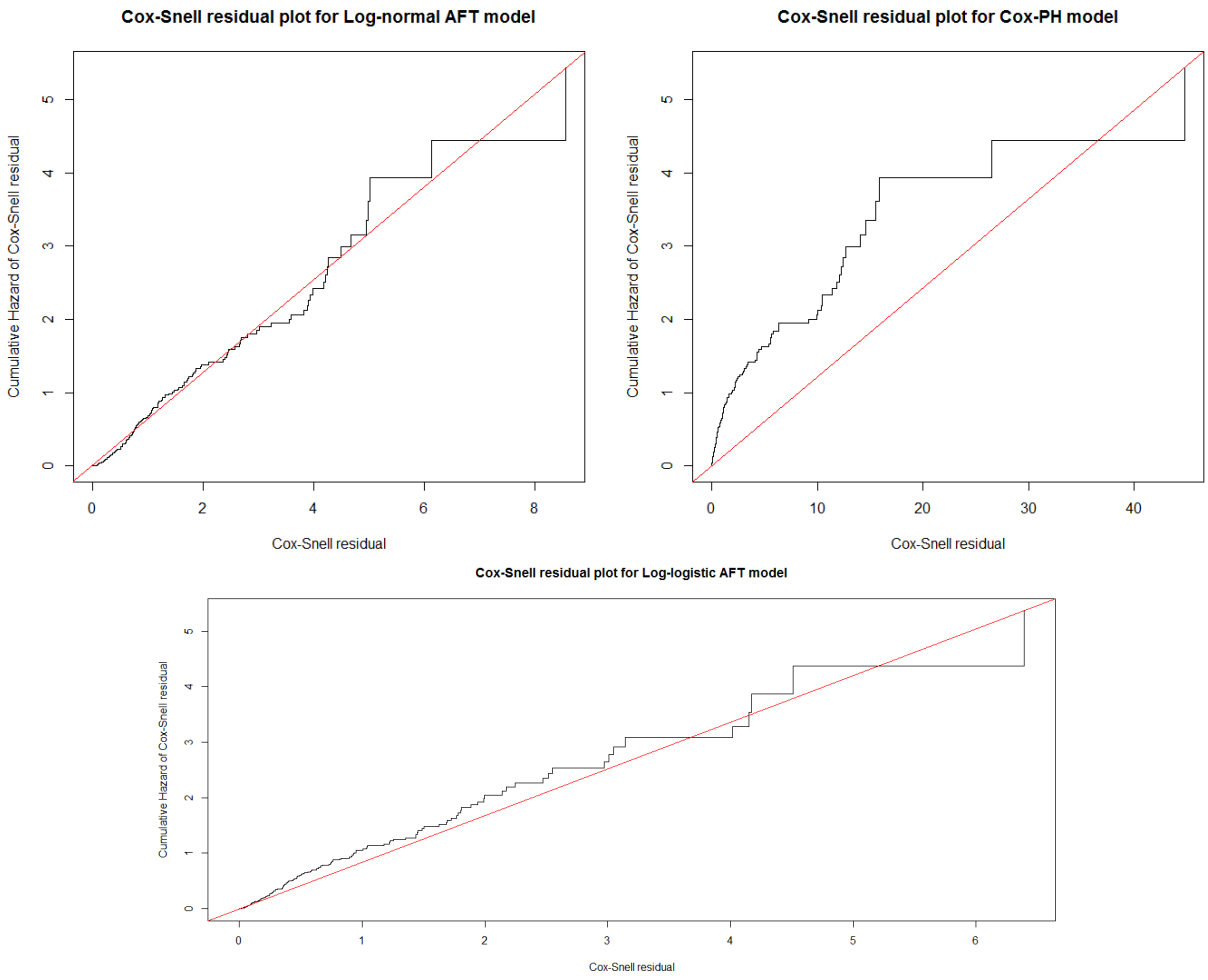


Figure 4.3: Bayesian Cox-Snell residual plots for baseline distribution and Cox-PH that were used to fit the Heart failure data set

The histograms of the cross-validated probability integral transform values in appendix Figure 5.4, shows that the posterior predictive p-values are some extent closer to a uniformly distributed with some observations outliers in HF data-set. The Conditional predictive ordinate values that are considerably smaller (order of magnitude smaller) than the others, so with respect to the Bayesian log-normal model, the observed values would be considered 'surprising', because the sum of the observations associated with failure flags are equal to zero in HF data-set.

From Figure 5.5, in appendix shows that the plots including 95% credibility interval by observing the posterior density for the parameters was normally distributed in HF data-set. The kullback-leibler divergence is a diagnostic that measures the accuracy of the INLA approximation, since from Table 4.5 shows that the kld for all significant parameters in the Bayesian log-normal AFT model were 0.

4.2 Discussions

The main aim of this study was to identify factors affecting the survival time of HF data set, which was obtained from JUMC, Jimma, Ethiopia. Heart failure is a growth problem in the world and the overall prevalence of HF in the adult population in developing countries is 7% - 10% with exponential rise with age (Adebayo *et al.*, 2017). The descriptive results of the study indicated that a total of 409 HF patients were included in this study, the minimum and maximum event time observed from HF patients follow up where 6 and 36 months respectively. In addition, fifty percent of HF patients who receive treatments, survived 31 months or above it.

In this study, among those HF patients, about 59.90% were censored (right censored) and remaining 40.10% were died. This finding was similar to a study conducted by Hailay *et al.* (2015) shows among those HF patients, 31.3% of them were dead while the rest 68.7% were censored. The Cox-PH model was applied for this data set. But, the assumption of Cox-PH model was violated and the AFT models with different baseline distribution (exponential, weibull, log-normal, log-logistic) were fitted for the HF data set. Among those different AFT models the log-normal AFT model was chosen over the other since AIC and BIC were small (Akaike, 1974; Schwarz *et al.*, 1978). Bayesian approach was applied on parametric AFT models and to compare the efficiency of different AFT models DIC and WAIC were used (Spiegelhalter *et al.*, 2004; Watanabe, 2010). Bayesian log-normal AFT model was the best model to describe HF data set from the given alternative. This result was similar with study done by Avi (2017).

Bayesian survival analysis using INLA and MCMC method helped to increase the accuracy of the results by observing narrow credible interval and minimum the standard error, this results, also confirmed by other studies done by El-Hakim and Uthman (1999), Akerkar *et al.* (2010), and Martino *et al.* (2011). The efficiency and relevance of the Bayesian survival models using MCMC method of estimation with application to clinical research were supported by the work of different researchers which in fact applied for the different data set (Khanal *et al.*, 2014; Khan and Khan, 2013; Kumar *et al.*, 2019).

In this study, Bayesian log-normal AFT model using INLA method was showed that smaller standard error and narrow credible interval for all significant parameters better than that of Bayesian log-normal AFT model using MCMC method and log-normal AFT (classical or frequentist) model in HF data-set. This results was similar to a study done by Abrha *et al.* (2018) shows that Bayesian survival analysis was better performance than classical parametric survival analysis and also the result was consistent with study done by Akerkar *et al.* (2010) shows that Bayesian survival models using INLA method was better than Bayesian survival models using MCMC method. Therefore, INLA method was provides a faster and more accurate alternative to simulation based MCMC schemes in HF data set. This result was consistent with studies done by Akerkar *et al.* (2010), and Martino *et al.* (2011).

However in this study, the results of Bayesian log-normal AFT model using INLA method shows that the survival time of HF patients significantly affected by age, chronic kidney disease, diabetes mellitus, etiology of HF, hypertension, anemia, smoking cigarette and stage of HF. From the result of this study the age group has a significant effect on the survival time of HF patients. In addition, the survival time of HF patients seems less as they gets older (greater than or equal to 65 years) and different studies were also persisted with this results Adebayo *et al.* (2017), Zeru (2018), and Sheng *et al.* (2018). The survival time of HF patients has no hypertension was higher than that of with hypertension and thus, hypertension had a significant effect on HF patients, the studies done by Ahmad *et al.* (2017), and Sheng *et al.* (2018) shows the same results.

On other hand, the survival time of smoker HF patients were decreases as compared to non-smoker which is similar to study done by Ahmad *et al.* (2017). The survival time of HF patients significantly affected by both type of diabetes mellitus and the expected survival time of HF with both type of diabetes mellitus was less as compared to HF patients without diabetic, this results consistent with studies done by Ahmad *et al.* (2017), and Zeru (2018). In addition, chronic kidney disease was significantly affected the survival time of HF patients and the survival time was high for HF patients do not have chronic kidney disease as compared to HF patients having chronic kidney disease, this results, also confirmed with study by Zeru (2018).

The studies done by Ahmad *et al.* (2017), and Zeru (2018) shows that the survival time of HF patients significantly affected by anemia and the expected survival time of HF with anemia was less as compared to HF patients without anemic. This studies was consistent with the current study. The stages of HF patients has significant effect on the survival time of HF patients. The study done by Zeru (2018) shows that the stages of HF patients has been significantly affected the survival time of HF patients. From the results of these study the survival time of HF patients was smaller as the stage increases as in result of this study.

For checking adequacy of the model, the cumulative hazard plots for the Bayesian Cox Snell residuals of the Cox-PH, Exponential, Weibull, Log-normal and the Log-logistic models were plotted as in Figure 4.3. The plots were more approached to the line in case of the Bayesian log-normal model that indicates the Bayesian log-normal was best in HF data-set. This result was consistent with other study done by Avi (2017). The conditional predictive ordinate and probability integral transform were also used for model checking. Before adequacy checking using graphical methods, it can be important to check whether the usual numerical problem occurred during the computation of conditional predictive ordinate. Thus, since the sum of the number of failure in conditional predictive ordinate was zero, no failure was detected and meaning that no numerical problem has occurred in HF data-set. The histogram and scatter plot of probability integral transform were plotted as in Figure 5.4, indicated that the plots of predictive residual based values were to some extent uniformly distributed with some deviated

outlier and there is reasonable predictive distribution matches the actual data. This result was persisted with other studies done by Akerkar *et al.* (2010), and Martino *et al.* (2011).

The Bayesian log-normal AFT model diagnostic plots including 95% credibility interval were plotted as in Figure 5.5 shows that the plot of posterior density for the parameters was normally distributed. Similarly, the kullback-leibler divergence is a diagnostic that measures the accuracy of the INLA approximation. In this study the values of kld for all significant parameters in the Bayesian log-normal AFT model were 0. This indicate that Bayesian log-normal AFT model using INLA method was fast and higher accuracy. This results, also confirmed by other studies done by Martino *et al.* (2011), and Akerkar *et al.* (2010).

However, the thesis was not done without limitation. The study was conducted based on secondary data gathered from registration log book and patients registration card, which might have incomplete and biased information; Lack of published literature's on the countries hospital based related to the survival time of heart failure patients using Bayesian survival models; As different literature pointed out, there are different prognostic factors (Body mass index and Weight) that are assumed to have impacts on the survival time of HF. However, data on those variables could not be available in hospital records.

5 CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

This study used survival time of heart failure patients data set, for those patients who were receiving treatments for at-least one time in Jimma University Medical Center. Bayesian log-normal AFT model performing better than various parametric models with baseline distribution (Exponential, Weibull, Log-logistic and Log-normal) for this study. The small standard error and narrow credible interval for all significant parameters for Bayesian log-normal AFT model using INLA method was better than Bayesian log-normal AFT model using MCMC method for the HF data set. Fifty percent of heart failure patients who receive treatments, survived 31 months or above it.

The survival time of HF patients significantly affected by age, chronic kidney disease, diabetes mellitus, etiology of heart failure, hypertension, anemia, smoking cigarette and stages of heart failure. Of all this statistically significant covariates; age (49 to 65 years and greater than 65 years); etiology of heart failure (rheumatic valvular heart disease, hypertensive heart disease and Other diseases); presence of hypertension; presence of anemic; presence of chronic kidney disease; smokers; diabetes mellitus (type I and type II diabetic); and stages of heart failure (II, III and IV) were prolong the timing death of heart failure patients.

5.2 Recommendations

Based on the finding of the study it is recommended as follows:-

- The ministry of health and policy makers should work on awareness by letting to know the risk factors for heart failure.
- The hospital, JUMC, need to improve public awareness for early detection of HF.
- Awareness has to be given for the society regarding smoking cigarette. The mass media can play an effective role in this regard.
- The researchers who are interested to investigate on the same area are recommended to introduce frailty modeling to account the correlation which comes from the cluster and to accounts un-observable random effect using Bayesian survival analysis using MCMC and INLA method.

References

P. Ponikowski, A. A. Voors, S. D. Anker, H. Bueno, J. G. Cleland, A. J. Coats, V. Falk, J. R. González-Juanatey, V.-P. Harjola, E. A. Jankowska *et al.*, *European journal of heart failure*, 2016, **18**, 891–975.

C. W. Yancy, M. Jessup, B. Bozkurt, J. Butler, D. E. Casey, M. H. Drazner, G. C. Fonarow, S. A. Geraci, T. Horwich, J. L. Januzzi *et al.*, *Journal of the American College of Cardiology*, 2013, **62**, e147–e239.

T. Vos, R. M. Barber, B. Bell, A. Bertozzi-Villa, S. Biryukov, I. Bolliger, F. Charlson, A. Davis, L. Degenhardt, D. Dicker *et al.*, *The Lancet*, 2015, **386**, 743–800.

E. J. Benjamin, P. Muntner and M. S. Bittencourt, *Circulation*, 2019, **139**, e56–e528.

M. D. Huffman, J. D. Berry, H. Ning, A. R. Dyer, D. B. Garside, X. Cai, M. L. Daviglus and D. M. Lloyd-Jones, *Journal of the American College of Cardiology*, 2013, **61**, year.

A. Damasceno, B. M. Mayosi, M. Sani, O. S. Ogah, C. Mondo, D. Ojji, A. Dzudie, C. K. Kouam, A. Suliman, N. Schrueder *et al.*, *Archives of internal medicine*, 2012, **172**, 1386–1394.

A. Misganaw, D. H. Mariam, A. Ali and T. Araya, *Journal of health, population, and nutrition*, 2014, **32**, 1.

S. R. Giolo, J. E. Krieger, A. J. Mansur and A. C. Pereira, *PloS one*, 2012, **7**, e37392.

A. Hailay, E. Kebede and K. Mohammed, *American Journal of Health Research*, 2015, **3**, 257–269.

D. Collett, *Modelling survival data in medical research*, Chapman and Hall/CRC, 2015.

S. P. Khanal, V. Sreenivas and S. K. Acharya, *Int. J. Sci. Res*, 2014, **3**, 161–166.

J. Qi, *Ph.D. thesis*, 2009.

J. Sheng, X. Qian and T. Ruan, *Open Journal of Statistics*, 2018, **8**, 651.

T. Ahmad, A. Munir, S. H. Bhatti, M. Aftab and M. A. Raza, *PloS one*, 2017, **12**, e0181001.

M. A. Zeru, *Journal of Public Health and Epidemiology*, 2018, **10**, 326–331.

J. G. Ibrahim, M.-H. Chen and D. Sinha, *Bayesian Survival Analysis*, Springer Science & Business Media, 2001.

J. G. Ibrahim, H. Zhu and N. Tang, *Lifetime data analysis*, 2011, **17**, 43–70.

A. Gelman, J. Hwang and A. Vehtari, *Statistics and computing*, 2014, **24**, 997–1016.

D. M. Lloyd-Jones, M. G. Larson, E. P. Leip, A. Beiser, R. B. DAgostino, W. B. Kannel, J. M. Murabito, R. S. Vasan, E. J. Benjamin and D. Levy, *Circulation*, 2002, **106**, 3068–3072.

M. A. Berarti and A. T. Goshu, *Cardiology and Angiology: An International Journal*, 2015, 65–79.

B. Habte, F. Alemseged and D. Tesfaye, *Ethiopian journal of health sciences*, 2010, **20**, year.

H. Amare, L. Hamza and H. Asefa, *BMC cardiovascular disorders*, 2015, **15**, 128.

A. P. Ambrosy, G. C. Fonarow, J. Butler, O. Chioncel, S. J. Greene, M. Vaduganathan, S. Nodari, C. S. Lam, N. Sato, A. N. Shah *et al.*, *Journal of the American College of Cardiology*, 2014, **63**, 1123–1133.

H. Dokainish, K. Teo, J. Zhu, A. Roy, K. F. AlHabib, A. ElSayed, L. Palileo-Villaneuva, P. Lopez-Jaramillo, K. Karaye, K. Yusoff *et al.*, *The Lancet Global Health*, 2017, **5**, e665–e672.

C. S. Lam, T.-H. K. Teng, W. T. Tay, I. Anand, S. Zhang, W. Shimizu, C. Narasimhan, S. W. Park, C.-M. Yu, T. Ngarmukos *et al.*, *European heart journal*, 2016, **37**, 3141–3153.

G. S. Bloomfield, F. A. Barasa, J. A. Doll and E. J. Velazquez, *Current cardiology reviews*, 2013, **9**, 157–173.

I. K. Owusu and Y. Boakye, *J Cardiovasc Dis Diagn*, 2013, **1**, 2.

A. Misganaw, D. H. Mariam and T. Araya, *Preventing chronic disease*, 2012, **9**, year.

S. G. Abdissa, K. Oli, Y. Feleke, D. Y. Goshu, D. M. Begna and A. Tafese, *Ethiop Med J*, 2014, **52**, 9–17.

S. Khatibzadeh, F. Farzadfar, J. Oliver, M. Ezzati and A. Moran, *International journal of cardiology*, 2013, **168**, 1186–1194.

C. L. Avery, L. R. Loehr, C. Baggett, P. P. Chang, A. M. Kucharska-Newton, K. Matsushita, W. D. Rosamond and G. Heiss, *Journal of the American College of Cardiology*, 2012, **60**, 1640–1646.

S. O. Adebayo, T. O. Olunuga, A. Durodola, O. S. Ogah *et al.*, *Nigerian Journal of Cardiology*, 2017, **14**, 9.

T. B. Abebe, E. A. Gebreyohannes, Y. G. Tefera and T. M. Abegaz, *BMC cardiovascular disorders*, 2016, **16**, 232.

D. D. McManus, G. Hsu, S. H. Sung, J. S. Saczynski, D. H. Smith, D. J. Magid, J. H. Gurwitz, R. J. Goldberg, A. S. Go and C. R. N. P. Study, *Journal of the American Heart Association*, 2013, **2**, e005694.

J. A. Bennett, B. Riegel, V. Bittner and J. Nichols, *Heart & Lung*, 2002, **31**, 262–270.

J. Urrutia, R. Tampis, J. Mercado and A. Baygan, *International Journal of Pharmacy and Pharmaceutical Sciences*, 2016, **8**, 6.

G. P. Fadini, A. Avogaro, L. Degli Esposti, P. Russo, S. Saragoni, S. Buda, G. Rosano, S. Pecorelli, L. Pani, O. H.-D. Network *et al.*, *European heart journal*, 2015, **36**, 2454–2462.

S. Miyagawa, K. Pak, S. Hikoso, T. Ohtani, E. Amiya, Y. Sakata, S. Ueda, M. Takeuchi, I. Komuro and Y. Sawa, *Circulation Reports*, 2019, **1**, 29–34.

G. Y. Lip, F. Skjøth, K. Overvad, L. H. Rasmussen and T. B. Larsen, *Clinical Research in Cardiology*, 2015, **104**, 1088–1096.

T. Callender, M. Woodward, G. Roth, F. Farzadfar, J.-C. Lemarie, S. Gicquel, J. Atherton, S. Rahimzadeh, M. Ghaziani, M. Shaikh *et al.*, *PLoS medicine*, 2014, **11**, e1001699.

M. Pourhoseingholi, A. Pourhoseingholi, M. Vahedi, B. M. Dehkordi, A. Safaee, S. Ashtari and M. Zali, *Iranian Journal of Cancer Prevention*, 2011, **4**, year.

S. K. Abrha, Z. G. Asfaw and D. G. Degefu, *Ł* , 2018.

Y. Khan and A. A. Khan, *International Journal of Innovative Research in Science, Engineering and Technology*, 2013, **2**, 7199–7204.

M. Kumar, P. K. Sonker, A. Saroj, A. Jain, A. Bhattacharjee and R. K. Saroj, *Cancer Reports*, 2019, e1210.

M. A. Berarti and A. T. Goshu, *Cardiology and Angiology: An International Journal*, 2015, 43–50.

E. Avi, *International Journal of Data Analysis Techniques and Strategies*, 2017, **9**, 63–74.

R. Akerkar, S. Martino and H. Rue, *Preprint Statistics, Norwegian University of Science and Technology*, 2010, **1**, 1–38.

S. Martino, R. Akerkar and H. Rue, *Scandinavian Journal of Statistics*, 2011, **38**, 514–528.

E. L. Kaplan and P. Meier, *Journal of the American statistical association*, 1958, **53**, 457–481.

O. Aalen, O. Borgan and H. Gjessing, *Survival and event history analysis: a process point of view*, Springer Science & Business Media, 2008.

J. P. Klein and M. L. Moeschberger, *Survival analysis: techniques for censored and truncated data*, Springer Science & Business Media, 2006.

D. Kleinbaum and M. Klein, *New York*, 2012.

N. Mantel and W. Haenszel, *Journal of the national cancer institute*, 1959, **22**, 719–748.

C. Dätwyler and T. Stucki, *Parametric survival models*, 2011.

D. R. Cox, *Journal of the Royal Statistical Society: Series B (Methodological)*, 1972, **34**, 187–202.

H. Akaike, in *Selected Papers of Hirotugu Akaike*, Springer, 1974, pp. 215–222.

G. Schwarz *et al.*, *The annals of statistics*, 1978, **6**, 461–464.

S. P. Brooks and A. Gelman, *Journal of computational and graphical statistics*, 1998, **7**, 434–455.

J. O. Berger, *Statistical decision theory and Bayesian analysis*, Springer Science & Business Media, 2013.

H. Rue, S. Martino and N. Chopin, *Journal of the royal statistical society: Series b (statistical methodology)*, 2009, **71**, 319–392.

A. E. Gelfand and B. K. Mallick, *Biometrics*, 1995, 843–852.

S. Depaoli, *Structural Equation Modeling: A Multidisciplinary Journal*, 2014, **21**, 239–252.

A. Bhattacharjee, *World journal of oncology*, 2014, **5**, 109.

H. Dezfuli, D. Kelly, C. Smith, K. Vedros and W. Galyean, 2009.

M. Ganjali and T. Baghfalaki, *JRSS*, 2012, **5**, 17–32.

R. Christensen, W. Johnson, A. Branscum and T. E. Hanson, *Bayesian ideas and data analysis: an introduction for scientists and statisticians*, CRC Press, 2011.

D. J. Spiegelhalter, K. R. Abrams and J. P. Myles, *Bayesian approaches to clinical trials and health-care evaluation*, John Wiley & Sons, 2004, vol. 13.

S. Watanabe, *Journal of Machine Learning Research*, 2010, **11**, 3571–3594.

K. Chaloner, *Biometrika*, 1991, **78**, 637–644.

J. Wakefield, *Bayesian and frequentist regression methods*, Springer Science & Business Media, 2013.

J. Piironen and A. Vehtari, *Statistics and Computing*, 2017, **27**, 711–735.

L. Pettit, *Journal of the Royal Statistical Society: Series B (Methodological)*, 1990, **52**, 175–184.

L. Held, B. Schrödle and H. Rue, in *Statistical modelling and regression structures*, Springer, 2010, pp. 91–110.

A. P. Dawid, *Journal of the Royal Statistical Society: Series A (General)*, 1984, **147**, 278–290.

D. W. Hosmer Jr, S. Lemeshow and S. May, *Applied survival analysis: regression modeling of time-to-event data*, Wiley-Interscience, 2008, vol. 618.

I. E. El-Hakim and M. A. Uthman, *International journal of dermatology*, 1999, **38**, 108–110.

Appendices

Appendix-1:- Some selective relevant summary tables and figures for HF data-set

Table 5.1: The Log-rank test of each the covariates

Covariates	**Chi-square**	**Degree of freedom**	**P-value**
Sex	4.8	1	0.03$*$
Age	21.7	2	2e-05$*$
Residence	0.4	1	0.5
Alcohol Consumption	2.5	1	0.1
History of Heart failure	5.7	2	0.06
Chronic kidney disease	84.8	1	$<$2e-16$*$
Diabetes Mellitus	102	2	$<$2e-16$*$
Etiology of Heart failure	19.7	4	6e-04$*$
Hypertension	75.8	1	$<$2e-16$*$
Anemia	58.5	1	2e-14$*$
Smoking Cigerette	63.1	1	2e-15$*$
Treatments Taken	3.4	4	0.5
Stage of Heart failure	41.2	3	6e-09$*$

$*$ **is $<$0.05 (significant)**.

Source: JUMC, Jimma, Ethiopia; from first January, 2016 to first January, 2019.

Table 5.2: Counts for patient status

Patients Status	Number of patients (%)
Censoring	245(59.90)
Death	164(40.10)

Table 5.3: The Multivariable analysis of Cox PH for HF data set

Covariates	Categories	$\hat{\beta}$	SE[$\hat{\beta}$]	$\exp(\beta)$	z	$[95\%CI\exp(\beta)]$	p-value
Sex	Female	Ref					
	Male	-0.466	0.216	0.627	-2.158	[0.41, 0.95]	0.030$*$
	≤ 49	Ref					
Age	49-65	0.439	0.292	1.552	1.501	[0.874, 2.756]	0.133
	≥ 65	0.738	0.281	2.093	2.625	[1.205, 3.634]	0.0086$*$
Hist.of HF	New	Ref					
	HF Pati.	0.023	0.195	1.024	0.122	[0.698, 1.502]	0.902
	Medical	-0.031	0.228	0.969	-0.137	[0.618, 1.518]	0.891
Alcohol	No	Ref					
	Yes	-0.037	0.176	0.962	-0.21	[0.681, 1.361]	0.830
Hyper	No	Ref					
	Yes	0.702	0.20	2.01	3.51	[1.364, 2.989]	0.00044$*$
CKD	No	Ref					
	Yes	0.602	0.195	1.826	3.086	[1.245, 2.676]	0.002$*$
	IHD	Ref					
Etio. of HF	RVHD	0.672	0.297	1.958	2.257	[1.092, 3.512]	0.024$*$
	Cardio	0.087	0.294	1.091	0.298	[0.612, 1.944]	0.765
	HHD	0.473	0.283	1.605	1.671	[0.921, 2.797]	0.094
	Others	0.094	0.447	1.616	1.07	[0.673, 3.883]	0.28
	I	Ref					
Stages	II	1.105	0.521	3.021	2.11	[1.086, 8.402]	0.034$*$
	III	1.103	0.483	3.015	2.281	[1.168, 7.786]	0.022$*$
	IV	1.178	0.477	3.248	2.470	[1.275, 8.274]	0.013$*$
Scigarette	No	Ref					
	Yes	0.688	0.216	1.989	3.183	[1.302, 3.039]	0.0014$*$
	Not	Ref					
DM	TypeI	0.538	0.26	1.713	2.0	[1.012, 2.900]	0.044$*$
	TypeII	1.047	0.232	2.85	4.49	[1.805, 4.499]	6.88e-06$*$
Anemia	No	Ref					
	Yes	0.60	0.182	1.826	3.307	[1.278, 2.609]	0.000943$*$

$*$ **is <0.05 (significant).** $exp(\beta)$**:indicates Acceleration factor; 95%CI for** $exp(\beta)$**: 95% confidence interval for acceleration factor; SE: standard error; Ref:Reference**.
Source: JUMC, Jimma, Ethiopia; from first January, 2016 to first January, 2019.

Table 5.4: The analysis of maximum likelihood parameter estimates

Covariates	Categories	$\hat{\beta}$	SE[$\hat{\beta}$]	ϕ	[95%$CI\hat{\phi}$]	p-value
Intercept		4.8604	0.2077			
Age	$\leq$ 49	Ref				
	49-65	-0.242	0.109	0.7849	[0.6335 , 0.9726]	0.0268$*$
	$\geq$ 65	-0.3175	0.1043	0.7279	[0.5933 , 0.8931]	0.0023$*$
Hyper	No	Ref				
	Yes	-0.2891	0.0721	0.7489	[0.6502 , 0.8627]	6.1e-05$*$
CKD	No	Ref				
	Yes	-0.375	0.0708	0.6873	[0.5982, 0.7896]	1.2e-07$*$
Etiology of HF	IHD	Ref				
	RVHD	-0.2951	0.1109	0.7445	[0.599, 0.9252]	0.0078$*$
	Cardio	-0.1557	0.1079	0.8558	[0.6927, 1.0573]	0.1489
	HHD	-0.2484	0.1092	0.780	[0.6297 , 0.9662]	0.0229$*$
	Others	-0.3756	0.1519	0.6868	[0.510 , 0.925]	0.0134$*$
Stages	I	Ref				
	II	-0.3824	0.181	0.6822	[0.4784 , 0.9726]	0.0346$*$
	III	-0.4005	0.1676	0.6699	[0.4824 , 0.9305]	0.0169$*$
	IV	-0.4792	0.1642	0.6193	[0.4488, 0.8544]	0.0035$*$
Scigarette	No	Ref				
	Yes	-0.145	0.0692	0.865	[0.7553 ,0.9907]	0.0362$*$
DM	Not	Ref				
	TypeI	-0.2190	0.0954	0.8033	[0.6663, 0.9685]	0.0217$*$
	Type II	-0.4058	0.0815	0.6664	[0.568, 0.7819]	6.5e-07$*$
Anemia	No	Ref				
	Yes	-0.1434	0.0688	0.8664	[0.757 , 0.9915]	0.0373$*$

$*$ is $<$0.05 (significant). ϕ:indicates Acceleration factor; 95%CI for ϕ: 95% confidence interval for acceleration factor; SE: standard error; Ref:Reference.
Source: JUMC, Jimma, Ethiopia; from first January, 2016 to first January, 2019.

Table 5.5: Indicating the comparisons of AFT model using AIC and BIC

Distribution	**Loglikelihood**	**AIC**	**BIC**
Exponential	-724.7	1489.453	1547.915
Log-normal	**-628.3**	**1298.620**	**1365.253**
Weibull	-630.8	1303.567	1369.596
Log-logistic	-630.5	1302.935	1367.806

Source: JUMC, Jimma, Ethiopia; from first January, 2016 to first January, 2019.

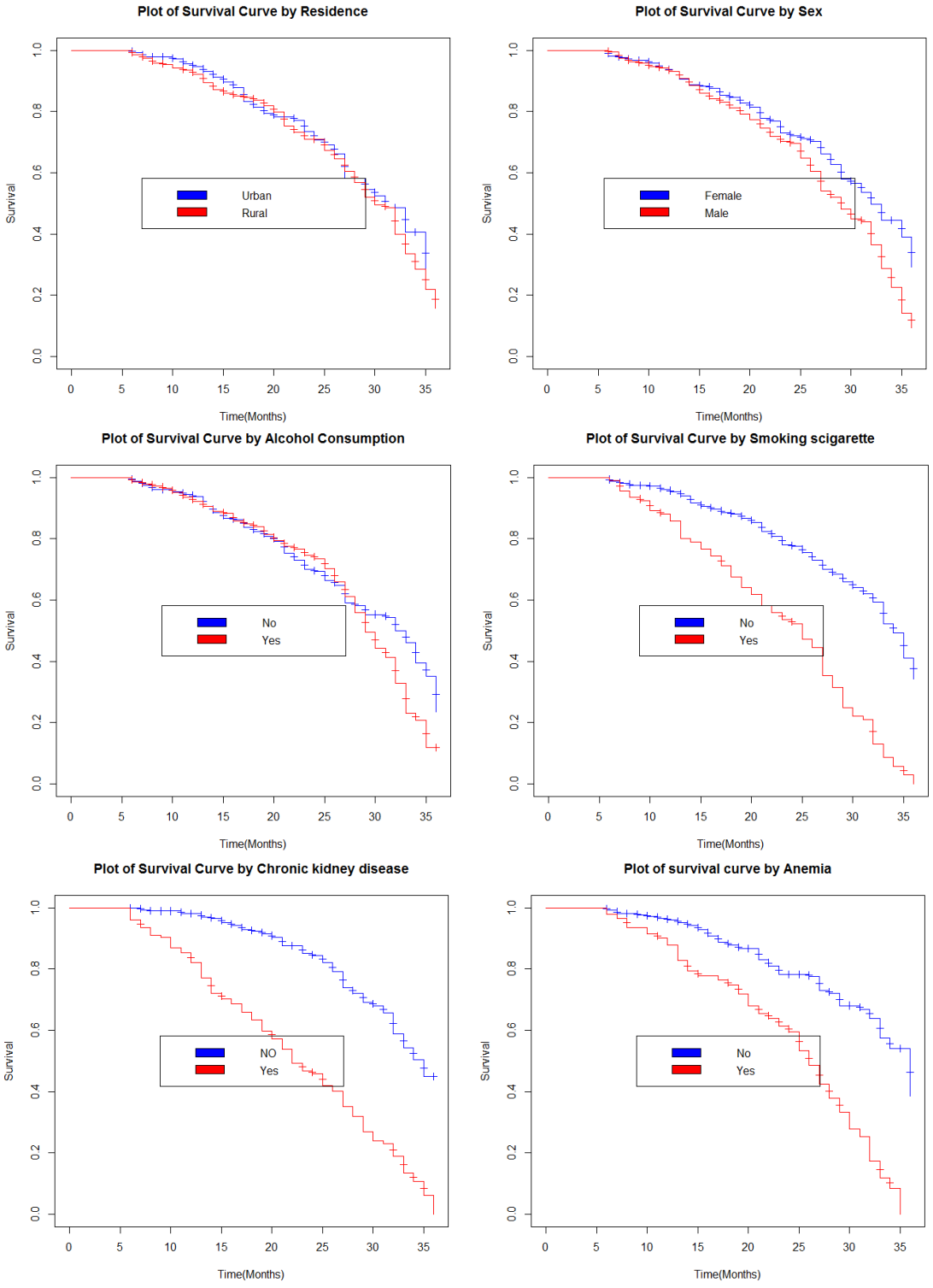


Figure 5.1: Kaplan-Meier estimates of the survival curves of HF data set for Residence, Sex, Chronic kidney disease, Smoking cigarette, Alcohol and Anemia

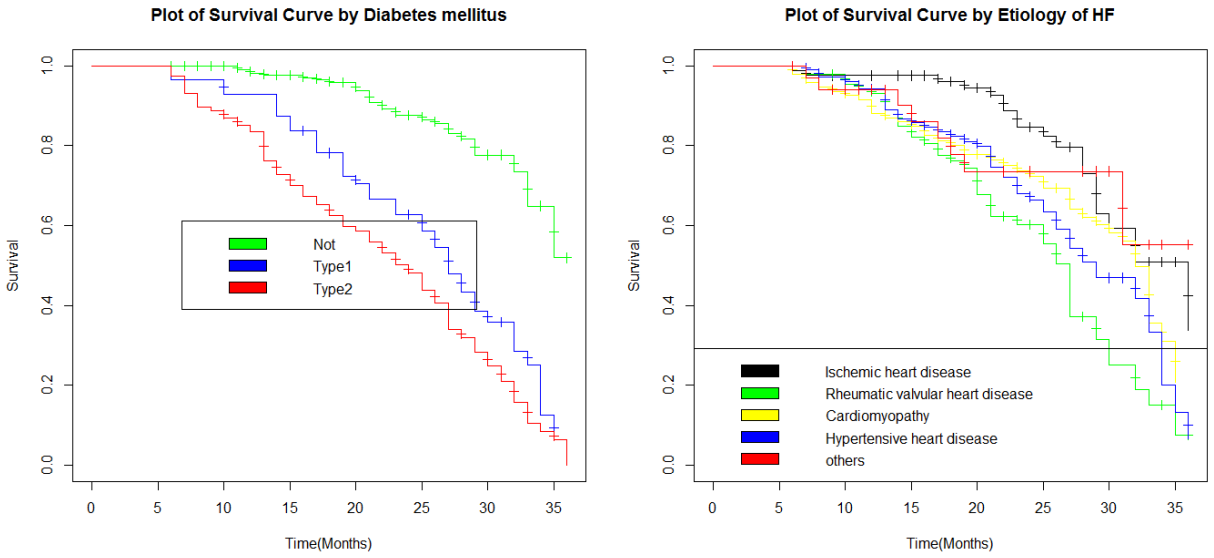


Figure 5.2: Kaplan-Meier estimates of the survival curves of HF data set for Diabetes, and Etiology of heart failure

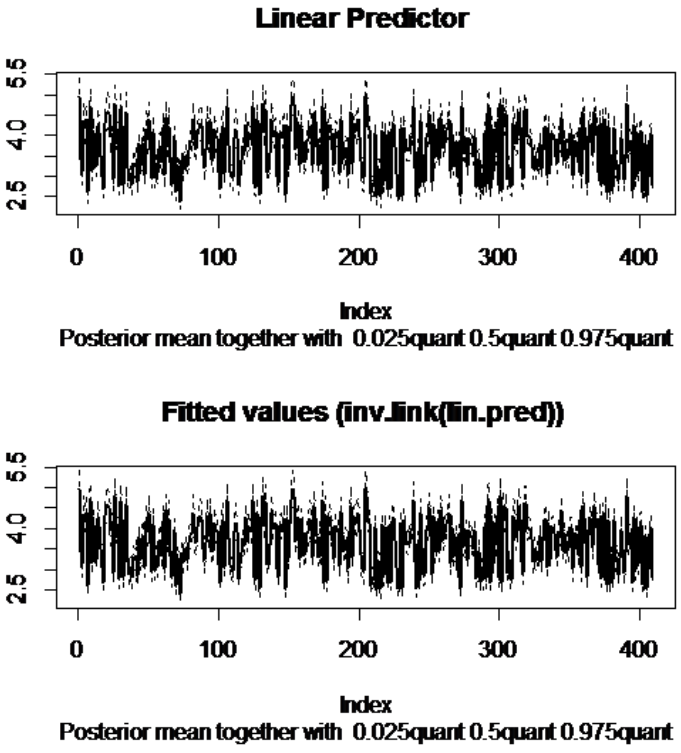


Figure 5.3: Plots for Linear predictors

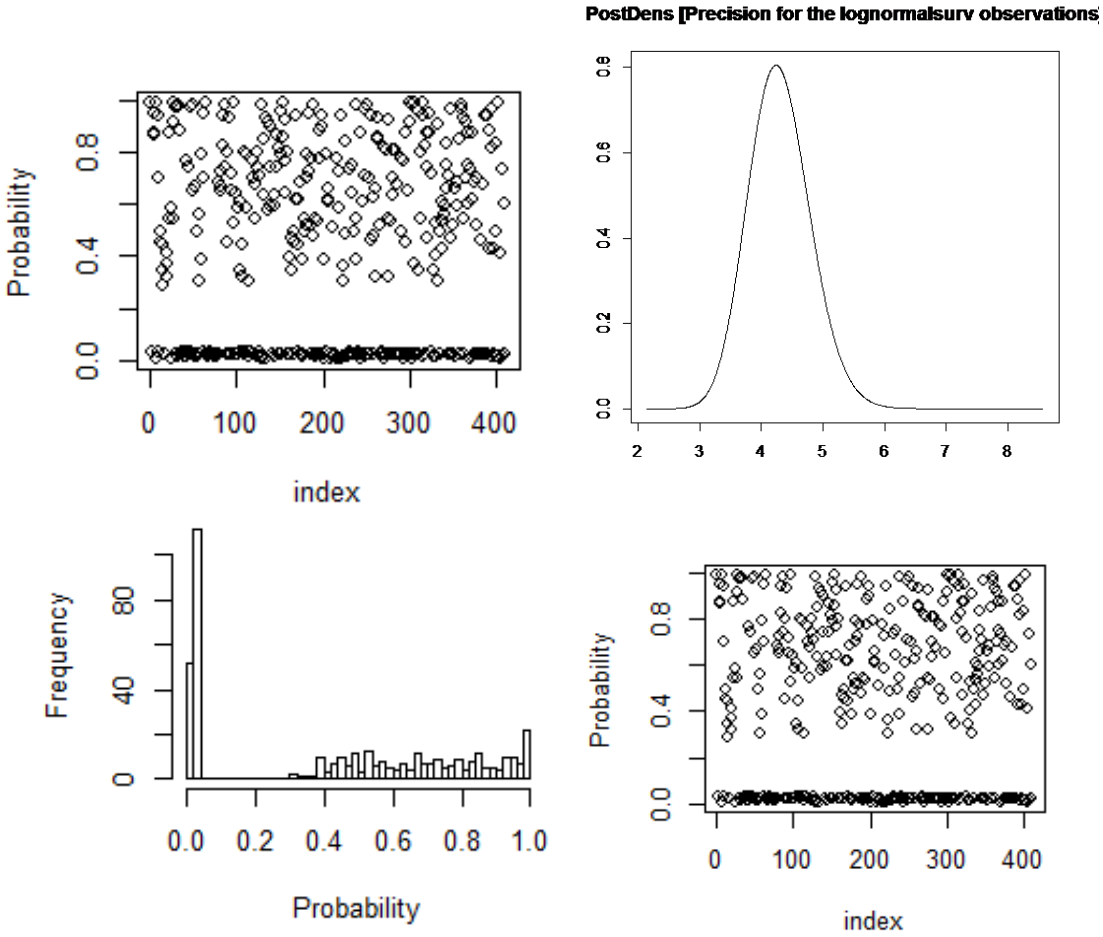
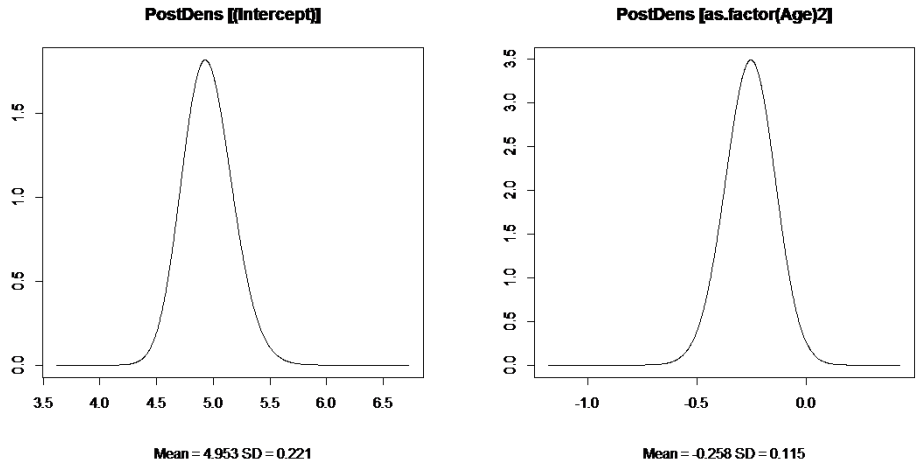


Figure 5.4: The scatter plots of the posterior mean, plot of the precision for the Bayesian log-normal model, and Histograms of the posterior predictive p-value



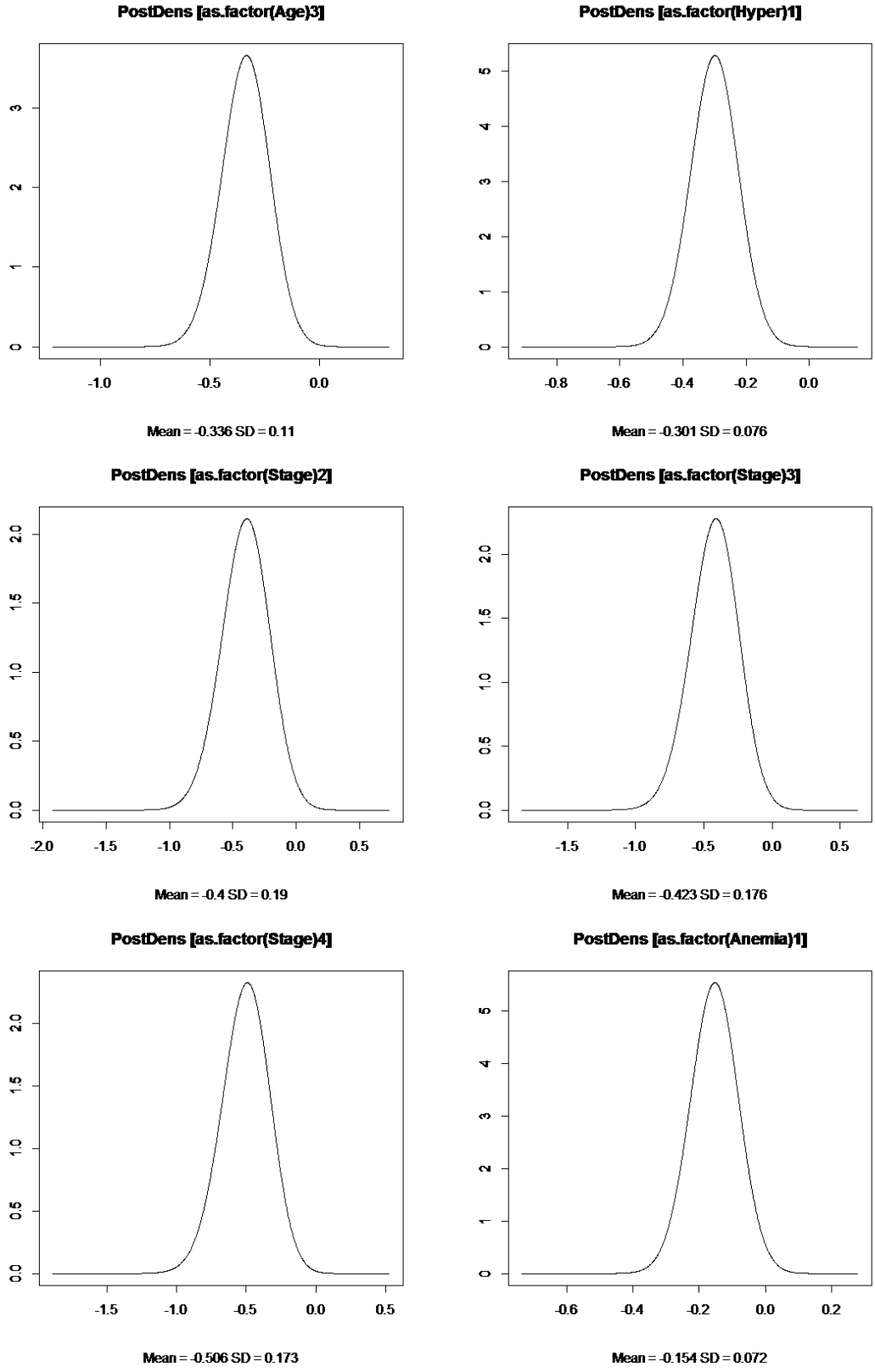


Figure 5.5: The Marginal distribution of Bayesian log-normal AFT model using INLA method.

Appendix-2:- Information Sheet and Data Extraction Form

Information Sheet

Introduction:- This information sheet is prepared for Jimma University Medical Center, Jimma, Ethiopia. The aim of the form is to make clear about the purpose of thesis, data collection procedures and to get permission for data collection.

Objective: The aim of this study is to identify factors affecting the survival time of heart failure patients in JUMC, Jimma, Ethiopia using Bayesian Survival Models.

Data Collection Procedure:- In order to achieve the above objective, information, which is necessary for the study, will be taken from the registration log book and patients registration card; if any inadequate information is countered it is checked from the file and excluded from analysis if proven to be inadequate. In order to come up with the above mentioned findings, total document of program clients enrolled during first January, 2016 up to 2019 will be seen and a review of the required information from the records are made by using the checklist.

Risk:- Since the study will be conducted by taking appropriate information from medical chart, it will not inflict any harm on the patients. The name or any other identifying information will not be recorded and all information taken from the chart will be kept strictly confidential and in a safe place. The information extracted will be kept secured and the information retrieved will only be used for the study purpose.

Benefits:- the thesis has no direct benefit for those whose document/ record is included in this thesis. However, indirectly the result of this study might be used to improve awareness on the factors that triggers the death of heart failure patients. It also enables to provide scientific information about the finding to Ministry of health in Ethiopia that helps policy makers to enhance the awareness of the society about factors that increase the probability of death due to heart failure which is protectable and curable if it is screened and treated in its earlier stage with appropriate treatment.

Confidentiality:- To ensure confidentiality the data on the chart will be collected by those individuals who are working in Hospital unit nurse and information will be collected without the name of the clients. The information collected from this thesis will be kept confidential and will be stored in a file. In addition, it will not be revealed to anyone except the investigator and it will be kept in key and locked system with computer password.

Person to contact:- This thesis will be reviewed and approved by college of Natural science, Jimma University post graduate research coordinate and by the department.

Permission:- Lastly but not least, you are kindly requested to permit and forward your permission to concerned body in your organization so that I can get cooperation from the data

clerks and other responsible bodies in place.

Data Extraction Form

Data extraction form, for the Bayesian Survival Analysis of Heart Failure Patients: A Case Study in Jimma University Medical Center, Jimma, Ethiopia, of Heart failure dataset (Starting from first January, 2016 up to first January, 2019).

1. Sex: Female ☐ Male ☐
2. Age of Patient (in year): ≤ 49 ☐ 49-65 ☐ ≥ 65 ☐
3. Residence: Urban ☐ Rural ☐
4. History of Heart Failure: New ☐ Heart failure patient before ☐ Medical OPD ☐
5. Alcohol consumption: No ☐ Yes ☐
6. Chronic kidney disease: No ☐ Yes ☐
7. Etiology of Heart Failure: Ischemic heart disease ☐ Rheumatic valvular heart diseases ☐ Cardiomyopathy heart disease ☐ Hypertensive heart disease ☐ others ☐
8. Smoking Cigarette: No ☐ Yes ☐
9. Stages of heart failure: I ☐ II ☐ III ☐ IV ☐
10. Treatments: Digoxin ☐ Spironolactone ☐ Atorvastatin ☐ Others ☐ Combination of the two or more ☐
11. Hypertension: No ☐ Yes ☐
12. Diabetes Mellitus: Not ☐ Type I ☐ Type II ☐
13. Anemia: No ☐ Yes ☐
14. Patient Status: ☐ Censored (alive , lost follow up, transfer & the like) ☐ Died