Determinants of Time-to-recovery from Hypertension using a Comparison of Various Parametric Shared Frailty Models: In Case of Felege Hiwot Refral Hospital.

Nigist Mulu
Bahir Dar University
Yeshambel Kindu (kinduyeshambel555@gmail.com)
Debark University
Abay Kassie
Debark University

Research Article

Keywords: hypertension, recovery, survival analysis, accelerated failure time, frailty model

Posted Date: July 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-608412/v1

Determinants of time-to-recovery from hypertension using a comparison of various parametric shared frailty models: In case of Felege Hiwot Refral hospital.

Nigist Mulu Takele [1] Yeshambel Kindu Yihuna[2] Abay Kassie Lakew[3]

[1]Department of Statistics, College of Natural and Computational Sciences, P.O. Box: 79, Bahir Dar University, Bahir Dar, Ethiopia, E-mail nigistmu@gmail.com,

[2]Department of Statistics, College of Natural and Computational Sciences, P.O. Box: 90, Debark University, Debark, Ethiopia

[3]Department of Statistics, College of Natural and Computational Sciences, P.O. Box: 90, Debark University, Debark, Ethiopia

Abstract

Background*: Hypertension is a major public health problem that is responsible for morbidity and mortality. In Ethiopia hypertension is becoming a double burden due to urbanization. The study aimed to identify factors that affect time-to-recovery from hypertension at Felege Hiwot Referral Hospital. Retrospective study design was used at FHRH.*

Methods*: The data was collected in patient's chart from September 2016 to January 2018. Kaplan-Meier survival estimate and Log-Rank test were used to compare the survival time. The AFT and parametric shared frailty models were employed to identify factors associated with the recovery time of hypertension patients. All the fitted models were compared by using AIC and BIC.*

Results*: Eighty one percent of sampled patients were recovered to normal condition and nineteen percent of patients were censored observations. The median survival time of hypertensive patients to attain normal condition was 13 months. Weibull- inverse Gaussian shared frailty model was found to be the best model for predicting recovery time of hypertension patients. The unobserved heterogeneity in residences as estimated by the Weibull-Inverse Gaussian shared frailty model was θ=0.385 (p-value=0.00).*

Conclusion*: The final model showed that age, systolic blood pressure, related disease, creantine, blood urea nitrogen and the interaction between blood urea nitrogen and age were the determinants factors of recovery status of patients at 5% level of significance. The result showed that patients creantine >1.5 Mg/dl compared to creantine ≤1.5 Mg/dl and SBP were prolonged the recovery time of patients whereas patients having kidney disease, other disease and had no any disease compared*

to diabetic patients and the interaction BUN and age were shorten recovery status of hypertension patients.

Key words: *hypertension, recovery, survival analysis, accelerated failure time, frailty model*

Abbreviations

AFT: Accelerated Failure Time
AIC: Akaike's Information Criterion
BIC: Bayesian Information Criterion
BMI: Body Mass Index
BP: Blood Pressure
BUN: Blood Urea Nitrogen
CVD: Cardio Vascular Disease
DBP: Diastolic Blood Pressure
DM: Diabetes Mellitus
FBS: Fasting blood sugar
FHRH: Felege Hiwot Referral Hospital
HR: Hazard Ratio
Mg/dl: Milligrams per deciliter
MmHg: Millimeter Mercury
NCD: Non Communicable Disease
PH: Proportional Hazard
SBP: Systolic Blood Pressure
TVC: Time Vairing Covariates
UN: United Nation
UNICEF: United Nations Children's Fund
WHO: World Health Organization

1. Background

Hypertension is one of chronic disease and the most commonly health problems resulting from high blood pressure during circulation. Hypertension is a silent public-health problem and a modifiable risk factor for non -communicable disease (NCD) and mortality. Globally, WHO report shows that 40 million deaths and more than 70% of mortality had been caused by NCDs in

2015 and 2016 respectively(Bayray *et al.,* 2018). Annually, the number of worldwide deaths from NCDs is expected to increase to 52 million by 2030 (Organization, 2014).Globally, the prevalence of hypertension was 1.3 billion individuals which represents 31% of adults (Bloch, 2016). Among adults aged 18 years and above the prevalence of hypertension was nearby 22% in 2014 (Organization, 2014)**.** And also hypertension is estimated to the risk factor for 7.5 million deaths and 57 million disability individuals in each year. In addition it is responsible for an estimated 45% and 51% of deaths due to heart disease and stroke respectively (Organization, 2009). Hypertension has been in the past regarded as a disease of affluence in economically higher class society but this has changed radically in the last two decades, that developing nations are not furnished to handle due to the health community's focus on infectious diseases and lack of awareness (Roger *et al.,* 2011*).* In a sense, the prevalence is increasing amongst poor sections of society. The worldwide prevalence of hypertension in adults aged 25 years and above was 40%, with the highest is found in Africa (46%) while the lowest is found in America (35%), based on world health organization report. Generally, high income countries have a lower prevalence of hypertension (35%) than lower and middle income countries (46%) (Alwan, 2011). Among total deaths due to cardiovascular disease, 80% occur in low and middle-income countries, with the highest death rate reported in Africa (Oo *et al.,* 2017). Hypertension is one of the most commonly detected danger factor for cardiovascular disease in Sub-Saharan African countries (Hendriks *et al.,* 2012). In this area, countries are experiencing an unexpected rise in the incidence of hypertension. There was an estimated 74.7 million hypertension patients in Sub-Saharan African countries and by the year 2025, the estimated number of hypertensive patients will be projected to increase by 68% to 125.5 million (Ogah and Rayner, 2013)**.**Ethiopia is one of Sub-Saharan African countries in which, non- communicable diseases and their related risk factors are becoming a double burden due to urbanization and economic development. Hypertension is also one of the most serious non- communicable chronic diseases in Ethiopia (Twagirumukiza *et al.,* 2011)**.** For many developing countries, the prevention and control of hypertension have given attention. Though, control of blood pressure and awareness about treatment is very little among under developing countries including Ethiopia (Organization).During the years 2011-2025, economic victims associated with NCDs like diabetes mellitus, hypertension in under developed countries are expected to be over $7 trillion these may causes millions of people below the poverty line. Ethiopia has not methodically

studied the epidemiological situation of hypertension nor did they establish systematic programs for create awareness ,prevention and control hypertension yet (Bloom *et al.,* 2011). Hypertension control is very low; among ten patients only eight patients are aware of their hypertension not controlled. These control levels are attributable to poor compliance and health seeking behaviors which are affected by a number of factors. Strategies are very important to improve usage of time to attain respectable control of hypertension. Understanding controlled time for hypertension is important in designing programs for hypertension control time and enhancing quality standards in healthcare delivery(Musinguzi *et al.,* 2018).Even if health professionals try to control blood pressure, the changes over time of blood pressure level on different factors that accelerate blood pressure are still not well understood. Hypertensive patients who have had a strong family history of hypertension, higher heart rate and a greater cardiovascular response controlled their BP compared to the normotensive family history-negative control population (Falkner *et al.,* 1981).Baseline systolic BP and age were significantly associated with control of hypertension but not diastolic BP. Factors like increasing age, current smoking, higher initial systolic BP and serum creantine were associated with uncontrolled BP at the final visit that is do not control their BP in a short period of time. A larger drop in SBP with little increase in the number of antihypertensive drugs is important to control BP in a short period of time (Xiong *et al.,* 2013).In this study, parametric share frailty models were used to analyze correlated hypertension by assuming that patients within the same cluster (residence) shares similar risk factors. Frailty model can be parametric or semi-parametric. The choice of distribution for the base line hazard is very important than the choice of frailty distribution(Fine *et al.,* 2003). Unlike proportional hazards model, AFT models measured the direct effect of the explanatory variables on the survival time instead of hazard. They are also less affected by the choice of probability distribution(Lambert *et al.,* 2004). Hence, in this research weibul, log-normal and log-logistic distributions were used and compared their efficiencies. And for the frailty distribution, gamma and inverse Gaussian frailty distributions were assumed. The aim of this study is to determine time to recovery from hypertension and identify significant predictors at Felege Hiwot referral hospital (FHRH).

2 Method

2.1 Study area

The study was conducted at Felege Hiwot referral hospital which provides good services for the management of Hypertension in Bahir Dar city administration. Bahir Dar is found at a distance of 565 kilometers away from Addis Ababa and located on the Southern shore of Lake Tana, the source of the Blue Nile. According to the report of the chief executive officer of FHRH, the total population served by the hospital is about 12 million per year, and there are around 400 health care professionals.

2.2 Data type

The data was retrospective survival data and it is secondary data that were found at Felege Hiwot referral Hospital which was followed up by doctors from September 2016 to January 2018. The patients with incomplete recording of baseline data were excluded from the study.

2.3 Study design and Sample size determination

2.3.1 Study design

A hospital based cross-sectional study design using secondary data from the retrospective record was conducted based on data from the chronic illness medication and follow up at FHRH. The source of population consists of all hypertensive patients who were joined for follow up in this hospital from September 2016 to January 2018. From the total registered patients at the hospital, only 299 of them were included in the study. The selected patients were picked-up by using systematic random sampling. Patients' id from review charts of hypertensive patients were used to find response and independent variables.

2.3.2 Sample size determination

In survival analysis (time to event data) sample size is determined by using the formula, Sample Size Calculations for Survival Analysis below(Collett, 2003): To find the required number of sample size, first calculate the required number of events

$$event = \frac{4(z\%_2 + z_\beta)^2}{(\theta_R^2)} = \frac{4c(\alpha,\beta)}{\theta_R^2} \quad , \quad \psi(hazardratio) = \frac{\log S_N(t)}{\log S_S(t)} \quad \text{and} \quad \theta_R = lan(\psi)$$

Total number of sample size can then calculate as:

$$n = \frac{event}{pr\{event\}} \quad \text{..... (3.1)}$$

$$\text{Where,} \qquad pr\{event\} = 1 - \frac{1}{6}\left[\bar{S}(f) + 4\bar{S}(0.5a + f) + \bar{S}(a + f)\right]$$

$$\bar{S}(t) = \frac{S_S(t) + S_N(t)}{2}$$

Where, $z_{\alpha/2} and\ z_{\beta}$ are the upper $\alpha/2$ and upper β points of the standard normal distribution, c(α,β)=c(0.05,0.10)=10.51 from table value, θ_R is log hazard ratio, a=accrual period, f= flow-up period and $S_S(t) and\ S_N(t)$ are estimated values of survivor functions for standard and new treatment respectively. Typically, α =0.05 and β =0.1 was used to determine sample size.

From the pilot survey estimate of survivor function for standard treatment; the median survivor time was 16 month, and the survival rate at seventeen, thirty-two and forty-seven months were 0.414, 0.265 and 0 respectively. The new treatment is expected to increase the survival rate at sixteen months from 0.69 under the standard treatment to, 0.79. In this study patients was recruited to the study over 30-month accrual period and there is to be a follow-up period of 17 months. From the above results the sample size is given by:

$$event = \frac{4(10.51)}{0.206} = 204.08 \approx 205\,, \qquad \text{pr(event)=0.687}$$

$$\text{Therefore,} \quad n = \frac{205}{0.687} = 298.399 \approx 299$$

2.4 Variables in the Study

The response variable of this study is the survival time of the hypertensive patients, that is the length of time from the start date of taking antihypertensive drugs until the date of recovery and the candidate predictors that included in this study were sex, age, residence, systolic blood pressure (mmhg), creantine(mg/dl), other related disease, blood urea nitrogen (mg/dl), diastolic blood pressure(mmhg) and number of medication.

2.5 Survival analysis

Survival analysis is a collection of statistical procedures that used to analysis data for outcome variable of interest in time until an event occurs. In a survival analysis, we usually refer to the time variable as survival time, because it gives the time that an individual has “survived” over some follow up period. We also typically refer to the event as a failure, because the event of

interest usually is death, disease incidence, or some other negative individual experience. However, survival time may be “time to return to work after an elective surgical procedure,” in which case failure is a positive event (Kleinbaum & Klein ,2005).The survival time of an individual is said to be censored when the end-point of interest has not been observed for that individual (Collett, 2015).

2.5.1 Survivor function S (t)

The survivor function is the probability that the survival time of a randomly selected subject is greater than or equal to some specified time t(G. Rodr´ıguez , 2010). Let T be a random variable associated with the survival times, 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 cumulative distribution function F (t), which represents the probability that a subject selected at random will have a survival time less than some stated value t, is given

$$F(t) = p(T < t) = \int_0^t f(u)du, t \geq 0 \quad (1)$$

Using the above the survivor function, S (t), can be given as

$$s(t) = p(T \geq t) = 1 - F(t), t \geq 0 \quad (2)$$

From equations (1) and (2) the relationship between f (t) and S (t) can be derived as

$$f(t) = \frac{dF(t)}{dt} = \frac{d}{dt}(1 - s(t)) = \frac{-ds(t)}{dt}, t \geq 0 \quad (3)$$

Survivor functions have the following characteristics that:

- They are non-increasing.
- The probability of surviving past time 0 is one.
- At time $t \rightarrow \infty$, $S(t) = S(\infty) = 0$

2.5.2 Hazard function h (t)

The hazard function h (t) gives the instantaneous potential for failing at time t, given that the individual has survived up to time t(G. Rodr´ıguez , 2010). The hazard function h (t) ≥ 0, is given

as h (t) $= \lim_{\Delta t \to 0} p\left(\frac{an\ indivdual\ failinin\ thetime\ \text{int}\ erval(t/t+\Delta t)\ given the survivalt}{\Delta t}\right)$

$= \lim_{\Delta t \to 0} p\,(t \le T \le t + \frac{\Delta t}{t}, T \ge t)/\Delta t$…….. (4)

By applying the theory of conditional probability and the relationship in equation, 3 the hazard function can be expressed in terms of the underlying probability density function and the survivor function becomes

$$h(t) = \frac{f(t)}{s(t)} = -\frac{d \ln s(t)}{dt} \quad \text{……} \quad (5)$$

The corresponding cumulative hazard function H (t) is defined as

$H(t) = \int_0^t h(u)du = -\ln s(t)$ then s(t)=exp(-H(t) and f(t)=h(t)s(t).from the above relation we can understand if Specifying one of the four functions f(t), S(t), h(t) or H(t) specifies the other three functions. Under the parametric approach, the baseline hazard function is defined as a parametric function and the vector of its parameters, say ψ , is estimated together with the regression coefficients and the frailty parameter(s).

2.6 Estimation of survivorship function

In practice, when using actual data, we usually obtain estimated survivor function and obtain curves that are step functions, rather than smooth curves.

2.6.1 The Kaplan-Meier estimate of the survival function

The number of observed events at t (j), j = 1… r. Then the K-M estimator of S(t) is defined as the Kaplan-Meier (KM) estimator is the standard non-parametric estimator of the survival function used for estimating the survival probabilities from observed survival times both censored and uncensored(Kaplan and Meier ,1958).

Suppose that r individuals have failures in a group of individuals, let $0 \le t_{(1)} \le \ldots < t_{(r)} < \infty$ be the observed ordered death times. Let $r_{(j)}$ be the size of the risk set $t_{(j)}$, where risk set denotes the

collection of individuals alive and uncensored just before $t_{(j)}$. Let $d_{(j)}$ be the number of observed events at $t_{(j)}$ $j=1\ldots r$. Then the K-M estimator of (t) is defined by

$$\hat{S}(t) = \prod_{j:\, t(j) < t} \left[1 - \frac{d_{(j)}}{r_{(j)}}\right] \quad \ldots\ldots (6)$$

This estimator is a step function that changes values only at the time of each birth interval. The cumulative hazard function of the KM estimator can be estimated as:

$\hat{H}(t) = -\ln\left[\hat{S}(t)\right]$, Where $\hat{S}(t)$ is KM estimator.

2.7 Comparison of Survivorship Functions

When comparing groups of subjects, it is always a good idea to begin with a graphical display of the data in each group. The figure in general shows if the pattern of one survivorship function lying above another which means the group defined by the upper curve had a more favorable survival experience than the group defined by the lower curve (Sudarno and A Prahutama ,2019). The general form of this test statistic is given by

$$Q = \frac{\left[\sum_{i=1}^{m} wi(d_{1i} - \hat{e}_{1i})\right]^2}{\sum_{i=1}^{m} wi^2 \hat{v}_{1i}} \quad \ldots\ldots (7)$$

In this expression

$$\hat{e}_{1i} = \frac{n_{1i} di}{ni} \quad \text{and} \quad \hat{v}_{1i} = \frac{n_{1i} n_{2i} di(ni - di)}{ni^2(ni-1)}$$

n_{1i} is the number at risk at observed survival time $t_{(i)}$ in group 1

$n_{2i}$is the number at risk at observed survival time $t_{(i)}$in the group 2

$d_{1i}$is the number of observed deaths in group 1

$d_{2i}$is the number of observed deaths in group 2

ni is the total number of individuals or risk before time $t_{(i)}$

di is the total number of deaths at $t_{(i)}$

wi is the total number of deaths at time $t_{(i)}$

The contribution to the test statistic depends on which of the various tests is used, but each may be expressed in the form of a ratio of weighted sums over the observed survival times.

Under the null hypothesis that the two survivorship functions are the same, and assuming that the censoring experience is independent of the group, and that the total number of observed events and the sum of the expected number of events is large, Q follows a chi-square distribution with one degree of freedom. We can also use the above test to compare k groups. In this study we used the log rank test which is special cases of Q.

2.7.1 Log-rank test

The log rank test, sometimes called the Cox-Mantel test, is the most well-known and widely used test statistic. This test is based on weights equal to one, i.e. wi=1

Therefore, the log rank test statistic becomes

$$Q_{LR} = \frac{\left[\sum_{i=1}^{m}(d_{1i}-\hat{e}_{1i})\right]^2}{\sum_{i=1}^{m}\hat{v}_{1i}} \quad \ldots\ldots (8)$$

2.8Accelerated failure time model

Parametric models are very applicable to analyze survival data; there are relatively few probability distributions for the survival time that can be used with these models. The Accelerated Failure Time model (AFT) is an alternative to the PH model for the analysis of survival time data(Frank Emmert-Streib et al.,2019). In AFT models we measured the direct effect of the explanatory variables on the survival time instead of hazard. This characteristic allows for an easier interpretation of the results because the parameters measure the effect of the correspondent covariate on the mean survival time. The members of the AFT model class include the exponential AFT model, Weibull AFT model, log logistic AFT model, log-normal AFT model, and gamma AFT model.

2.8.1 Weibul Accelerated Failure Time Model

The Weibul distribution is very flexible model for time-to-event data. It has a hazard rate which is monotone increasing, decreasing, or constant. The Weibul distribution (including the exponential distribution as a special case) can also be parameterized as an AFT model, and they are the only family of distributions to have this property. The results of fitting a Weibul model can therefore be interpreted in either framework (Klein & Moeschberger, 2003).

In terms of the log–linear representation of the model, if T_i has a Weibull distribution, then ε_i has a type of the extreme value distribution known as Gumbel distribution. This is an asymmetric distribution with survival function $S_{\varepsilon_i}(\varepsilon) = \exp(-e^{\varepsilon}), -\infty < \varepsilon < \infty$

Then the survivor function of $T_i = \exp(\mu + \alpha'\mathbf{X_i} + \sigma\varepsilon_i)$ is

$$S_i(t) = S_{\varepsilon i}\left(\frac{\log t - \mu - \alpha'\mathbf{X_i}}{\sigma}\right)$$

$$S_i(t) = \exp\left\{-\exp\left(\frac{\log t - \mu - \alpha_1 x_{1i} - \cdots - \alpha_p x_{pi}}{\sigma}\right)\right\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (9)$$

And the hazard function is

$$h_i(t) = \frac{1}{\sigma t}\exp\left\{\frac{\log t - \mu - \alpha_1 x_{1i} - \cdots - \alpha_p x_{pi}}{\sigma}\right\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (10)$$

2.8.2 log-logistic accelerated failure time model

Unlike Weibul hazard, Log-logistic distribution is not a monotonic function of time. However, situations in which the hazard function changes direction can arise. In cases where one comes across to censored data, using log-logistic distribution is mathematically more advantageous than other distributions.

According to the study of Gupta *et* al. (1999), the log-logistic distribution is proved to be suitable in analyzing survival data conducted by Cox (1972), Cox and Oakes (1984), Bennet (1983).The cumulative distribution function can be written in closed form is particularly useful for analysis of survival data with censoring (Bennett, 1983). The log-logistic distribution is very similar in shape to the log-normal distribution, but is more suitable for use in the analysis of survival data. The log-logistic model has two parameters λ and ρ, where λ the scale parameter is and ρ is the shape parameter. The probability density function of Log-logistic distribution given by $f(t) = \frac{\lambda\rho\, t^{\rho-1}}{(1+\lambda t^{\rho-1})^2}$, $\lambda \in R$, $\rho > 0$ the survivor function and the hazard functions of Log-logistic distribution are $\frac{1}{1+\lambda\, t^{\rho}}$ and $\frac{\lambda\rho\, t^{\rho-1}}{1+\rho\, t^{\rho-1}}$

respectively.

Suppose the variable ε_i in the log-linear formulation has a log-logistic distribution with zero mean and variance$\frac{\pi^2}{3}$. The survival function of ε_i is $S_{\varepsilon_i}(\varepsilon) = \frac{1}{1+e^{\varepsilon}}$

The survival function of T_i is then

$$S_i(t) = S_{\varepsilon i}\left(\frac{logt - \mu - \alpha' X_i}{\sigma}\right)$$

$$S_i(t) = \left\{1 + exp\left(\frac{logt - \mu - \alpha_1 x_{1i} - \cdots - \alpha_p x_{pi}}{\sigma}\right)\right\}^{-1} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (11)$$

And the hazard function of T_i for the i^{th} individual is then

$$h_i(t) = \frac{1}{\sigma t}\left\{1 + exp\left[-\left(\frac{logt - \mu - \alpha_1 x_{1i} - \cdots - \alpha_p x_{pi}}{\sigma}\right)\right]\right\}^{-1} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (12)$$

2.8.3 Log-normal accelerated failure time

The lognormal distribution is also defined for random variables that take positive values and so may be used as a model for survival data. If the survival times are assumed to have a log-normal distribution, the baseline survival function and hazard function respectively are given by.

$S_0(t) = 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)$ and $h_0(t) = \frac{\phi\left(\frac{\log t}{\sigma}\right)}{\left[1-\Phi\left(\frac{\log t}{\sigma}\right)\right]\sigma t}$, Where μ and σ are parameters, (t) is the probability densty function and (t) is the cumulative density function. The survival function for the i^{th} individual is

$$S_i(t) = 1 - \Phi\left(\frac{\log t - \boldsymbol{\alpha' X_i} - \mu}{\sigma}\right)$$

2.8.4 Parameter estimation

Parameters of AFT models can be estimated by maximum likelihood method. The likelihood of n observed survival time, t1, t2, t3… tn, the likelihood function for right censored data is given by:

$$L(\partial,\mu,\delta)=\prod_{j=1}^{n} f_i(t_i)^{\delta_i} * s_i(t_i)^{1-\delta_i} \quad \text{.....(13)}$$

Where fj(tj) the density function of the t^{th} individual at time ti , Si(ti) is the survival function of the t^{th} individual at time ti, δ_i is indicator variable. The logarithm of the above equation yields;

$$\log L(\partial,\mu,\delta)=\sum_{j=1}^{n}\{-\delta_i \log(\delta t_i+\delta_i \log f_i(x_i)+(1-\delta_i)\log S_i(W_i))\}$$

.....14

Where $W_j=\{\log t_i-\frac{\mu+\alpha_{1i}...+\alpha_{pi}x_{pi}}{\delta}\}$, Z= $\{z_{ji}\}$ is a vector of covariates for the j^{th} subject. The Maximum likelihood parameter estimates are found by using a Newton-Raphson procedure which can be done by software.

2.9 Shared frailty model

Multivariate or shared frailty model is a conditional independence model in which frailty is common to all subjects in a cluster. The concept of frailty provides a suitable way to introduce random effects in the model to account for association and unobserved heterogeneity. In its simplest form, a frailty is an unobserved random factor that modifies multiplicatively the hazard function of an individual or cluster of individuals. (Vaupel&Manton et al., 1979) introduced the term frailty. Clayton. (1978) promoted the model by its application to multivariate situation on chronic disease incidence in families. A random effect model takes into account the effects of unobserved or unobservable heterogeneity, caused by different sources. Multivariate frailty model is an extension of the univariate frailty model which allows the individuals in the same cluster to share the same frailty value. When frailty is shared, dependence between individuals who share frailties is generated.

Suppose we have k observations and i subgroups. Each subgroup consists of ni observations and $\sum_{i=1}^{r} n_i=n$ where n is the total sample size. The hazard rate for the k^{th} individual in the i^{th} subgroup is given by:

$$h_{ik}(t)=h_0(t)\exp(z_{ik}'\beta+u_i) \qquad \text{i}=1,2,3.....\text{rand k}=1,2.....n_i \quad \text{.....(13)}$$

Where h_0 (t) is the baseline hazard function and β is a vector of fixed effect parameters to be estimate. If the proportional hazards assumption does not hold, the accelerated failure time frailty model which assumes

$$h_{ik}(t) = h_0(\exp(z'_{ik}\beta + u_i)t)\exp(z_{ik}'\beta + u_i) \quad \text{.....(14)}$$

The frailties u_i are assumed to be distributed to be identically and independently with a mean of zero and a variance of unity. If the number of subjects ni is 1 for all groups, the univariate frailty model is obtained (Wienke, 2011); otherwise the model is called the shared frailty model (Hougaard, 2000) because all subjects in the same cluster share the same value of frailty.

The main assumption of a shared frailty model is that all individuals in cluster i share the same value of frailty Zi (i = 1, 2, 3 ..., n), and this is why the model is called the shared frailty model. The lifetimes are assumed to be conditionally independent with respect to the shared frailty. This shared frailty is the cause of dependence between lifetimes within the clusters.

Extensive research has been devoted to the frailty issue in survival analysis and generalized linear model (GLIM). Recently, investigators have recognized that ignoring individual heterogeneity may lead to inaccurate conclusions. (Oakes, 1989) proposed frailty models for bivariate survival times and introduced several possible frailty models. For reasons of convenience, analysts often choose parametric representations of frailty models that are mathematically tractable. (Hougaard 1986) used several distributions for frailty including gamma, inverse Gaussian, positive stable distributions and claimed that these two distributions are relevant and mathematically tractable as a frailty distribution for heterogeneous populations. Vaupel et al.(1979) assumed that frailty is distributed across individuals as a gamma distribution. In this study we used gamma distributions. These two distributions are the main frailty distributions widely used in the literature because of its simplicity and mathematical tractability. From an analytical and computational view gamma is a very convenient distribution. Arguably, this is the most popular frailty model due to its mathematical tractability, by (De Sherbinin et al., 2008).

2.9 The Gamma shared Frailty Distribution

The gamma distribution has been widely applied as a mixture distribution (Vaupel et al., 1979).From a computational and analytical point of view, it fits very well to failure data. It is widely used due to mathematical tractability (Wienke, 2010).

The density of a gamma-distributed random variable with parameter θ is given bY

$$f_z(z) = \frac{z_i^{\frac{1}{\theta}} \exp(\frac{-z_i}{\theta})}{\theta^{\frac{1}{\theta}} \Gamma(\frac{1}{\theta})} \qquad \theta > 0$$

Where $\Gamma(.)$ is the gamma function, it corresponds to a Gamma distribution Gam (μ, θ) with μ fixed to 1 for identifiability. Its variance is then θ, with Laplace transform.

$$L(u) = (1+\frac{u}{\theta})^{-\theta}$$

The conditional survival function of the gamma frailty distribution is given by (Gutierrez ,2002)

$$s_\theta(t) = [1-\theta \ln(s(t)]^{\frac{-1}{\theta}}$$

And the conditional hazard function is given by:

$$h_0(t) = h(t)[1-\theta \ln(s(t)]^{-1}$$

Where S (t) and h (t) are the survival and the hazard functions of the baseline distributions. For the Gamma distribution, the Kendall's Tau (Hougaard,2000), which measures the association between any two event times from the same cluster in the multivariate case, can be compute by:

$$\Gamma = \frac{\theta}{\theta+2} \in (0,1)$$

2.10 Parameter estimation

For right-censored clustered survival data, the observation for subject $j \in J_i = \{1,2,3.....n_i\}$ from the cluster $i \in I = \{1,2,3.....r\}$ is the couple $((z_{ij}, \delta_{ij})$ where $z_{ij=} \min(t_{ij}, c_{ij})$ is the minimum between the survival time t_{ij} and the censoring time c_{ij}, and $\delta_{ij} = I(t_{ij} \le c_{ij})$ is the event indicator. When covariate information is being collected the observation will be $(z_{ij}, \delta_{ij}, x_{ij})$ where X_{ij} denote the vector of covariates for the ij-th observation. In the parametric setting, estimation is based on the marginal likelihood in which the frailties have been integrated out by averaging the conditional likelihood with respect to the frailty distribution. Under assumptions of non-informative right-censoring and of independence

between the censoring time and the survival time random variables, given the covariate information, the marginal log-likelihood of the observed data can be written as.

$$\text{Im}\arg(\mathrm{h}_O,\beta,\theta;z,x)=\prod_{i=1}^{s}\left[(\prod_{j=1}^{ni}(h_o(yij\exp(xij^{T}\beta))^{\delta ij})\times\int_0^{\infty}zi^{di}\exp\left(-zi\sum_{j=1}^{ni}H_O(yij)\exp(xij^{T}\beta)\right)f(zi)dzi\right]$$

$$=\prod_{i=1}^{s}\left[(\prod_{j=1}^{ni}(h_o(yij\exp(xij^{T}\beta))^{\delta ij})(\times(-1)^{di}L^{(di)}\left(\sum_{j=1}^{ni}H_O(yij)\exp(xij^{T}\beta\right)\right]$$

Taking the logarithm, the marginal likelihood is

$$I_{m\arg}(h_O(.),\beta,\theta,z,x)=\sum_{i=1}^{s}\{[\sum_{j=1}^{n_i}\delta ij(\log(h_0(yij))+xij^{T}\beta)]+\log[(-1)^{di}L^{(di))}([\sum_{j=1}^{n_i}H_0(yij)\exp(xij^{T}\beta)]\}$$

Where $d_i=\sum_{j=1}^{n_i}\delta_{ij}$ is the numbers of events in the i-th cluster, and $L^{(q)}(.)$ the q-th derivative of the Laplace transform of the frailty distribution defined as,

$$L(s)=\mathrm{E}[\exp(-zs)]=\int_0^{\infty}\exp(-z_is)f(z_i)dz_i\,,\ s>0$$

$$L^{(q)}(s)=(-1)^q\int_0^{\infty}z^q\exp(-zs)f(z)dz,q\geq 0$$

Where $h_O(.)$ represents a vector of parameters of the baseline hazard function β, the vector of regression coefficients and θ the variance of the random effect. Estimates of $h_O(.)$, β, θ are obtained by maximizing the marginal log-likelihood above. This can be done if one is able to compute higher order derivatives $L^{(q)}$ (.) of the Laplace transform up to $q=\max\{d1,d2,\ldots,ds\}$.

2.11. Comparison of models

Model selection and comparison are the most common problems of statistical practice, with numerous procedures for selecting among a set of candidate models (Kadane and Lazar ,2004).There are several methods of model selection like AIC, BIC and LRT. However, in some circumstances, it might be useful to easily obtain AIC value for a series of candidate models (Munda *et al*., 2012). In this study, AIC criteria were used to compare various candidate models and the model with the smallest AIC value is considered as a better fit. AIC can be obtained by

$$AIC=-2\log L+2(c+k+1)$$

Where L is the maximum likelihood value, K is the number of covariates and C the number of model specific distributional parameters. In addition to this criteria, also BIC and likelihood ratio test are important in order to compare models, LRT is a best model selection technique, when models are nested (one model contains a subset of the explanatory variables in another), however, BIC is available for non-nested model.

2.12 Model checking and diagnostics

After a model has been fitted, the adequacy of the fitted model needs to be assessed. The methods for assessment of model checking for this study were used evaluation of the Parametric Baselines and the Cox Snell Residuals.

2.12.1 Evaluation of the parametric baselines

The appropriateness of model with the weibul baseline has a property that the log (-log(S (t)) is linear with the log of time, where S (t) = $\exp(-\lambda t^{\gamma})$ Hence, Substitute the K-M estimate of the survivor function $\hat{s}(t)$ for s(t) .If the Weibul assumption is tenable $\hat{s}(t)$ would be close to s(t) and the plot of $\log\{-\log(s(t)\}$ against $\log(t)$ would then give an approximately straight line ((Datwyler, 2011).

The log-failure odd versus log time of the log-logistic model is linear. Where the failure odds of log-logistic survival model can be computed as:

$\frac{1-s(t)}{s(t)} = \frac{\frac{\lambda t^{\gamma}}{1+\lambda t^{\gamma}}}{\frac{1}{1+\lambda t^{\gamma}}} = \lambda t^{\gamma}$ Therefore the log-failure odds can be written as:

$\log(\frac{1-s(t)}{s(t)})) = \log(\lambda t^{\gamma}) = \log\lambda + \gamma\log(t)$,therefore the appropriateness of model with the log logistic baseline can graphically be evaluated by plotting $\log((\frac{1-\hat{s}(t)}{\hat{s}(t)}$ vs log time where $\hat{s}(t)$ is Kaplan-Meier survival estimate (Duchateau & Janssen, 2008). If the plot $\phi^{-1}[1-s(t)]$ against log (t) is linear, the lo-normal distribution is appropriate for the given data set.

3.12.2 Cox-Snell residuals

The residual that is most widely used in the analysis of survival data is the Cox-Snell residual, the Cox-Snell residual is given by (Cox and SNELL). For the parametric regression problem,

analogs of the semi parametric residual plots can be made with a redefinition of the various residuals to incorporate the parametric form of the baseline hazard rates (Klein, 2012).

The Cox-Snell residual for the k^{th} individual with observed survival time t_k is defined $r_k = \hat{H}(T_k / X_k) = -\log \hat{s}(T_k / X_k)$

Where $\hat{H}$ is refer to the cumulative hazard function of the fitted model. If the model fits the data, then the r_k's should have a standard (λ =1) exponential distribution, so that a hazard plot r_k versus the Nelson–Aalen estimator of the cumulative hazard of the r_k's should be a straight line with intercept zero and slope unity.

Note: the Cox-Snell residuals will not be symmetrically distributed about zero and cannot be

Negative.

3 Results

3.1 summary statistics

Random samples of 299 hypertensive patients were included in this study. Of these, 242 (81%) were recovered and 57(19%) were censored observations during the follow up period. Among those sampled patients, 173 (57.9%) were female and 175 (58.5%) come from urban areas. Furthermore, 60(20%) of the patients had diabetes, 58(19.5%) had kidney disease, 36(12.0%) had other disease and 145(48.5%) had no any disease. the results had been summarized in **table1** below.

Table 1: Descriptive Summary of hypertension patients at felege Hiwot referral hospital

Factors	categories	status		
		Censored (%)	Event (%)	Total (%)
Sex	male	12(4.0)	144(38.0)	126(42.0)
	female	45(15.2)	128(42.8)	173(58.0)
residence	rural	18(6.0)	106(35.5)	124(41.5)
	urban	39(13.0)	136(45.5)	175(58.5)

Creantine(mg/dL)	≤1.5	18(6.0)	112(37.5)	130(43.5)
	>1.5	39(13.0)	130(43.5)	169(56.5)
Other related disease	Diabetes	28(9.4)	32(10.7)	60(20.0)
	kidney	18 (6.0)	40(13.5)	58(19.5)
	Other	4 (1.3)	32(10.7)	36(12.0)
	none	7(2.3)	138(46.2)	145(48.5)
Blood urea nitrogen(mg/dL)	≤35	14(4.7)	120(40.1)	134(44.8)
	>35	43(14.4)	122(40.8)	165(55.2)
Number of medication	one	53(17.7)	177(59.2)	230(76.9)
	two or more	4 (1.4)	65(21.7)	69(23.1)

Table 2: Summary table of continuous variables for time-to-recovery from hypertension

The mean, median, minimum, maximum and standard deviation of continuous variables included in this study is given **table2** below.

Covariates	Mean	median	minimum	Maximum	standard deviation
Age of patients	56.49	58.00	20	106	17.755
Systolic blood pressure	162.62	160.00	110	240	24.872
diastolic blood pressure	94.28	90.00	70	140	12.743
Time	15.63	13.00	3	45	10.408

3.2.1 The Kaplan- Meier Survival Estimate for Recovery Time of hypertension Patients

Non-parametric survival analysis is very important to visualize the survival of time-to-recovery of patients from hypertension at FHRH under different levels of covariates. Moreover, it gives information on the shape of the survival and hazard functions of hypertension data set given figure 1 below. Survival time of distributions for time-to-recovery is estimated for each group using the K-M method and in order to compare the survival curves of two or more groups, log-rank and Wilcoxon tests.

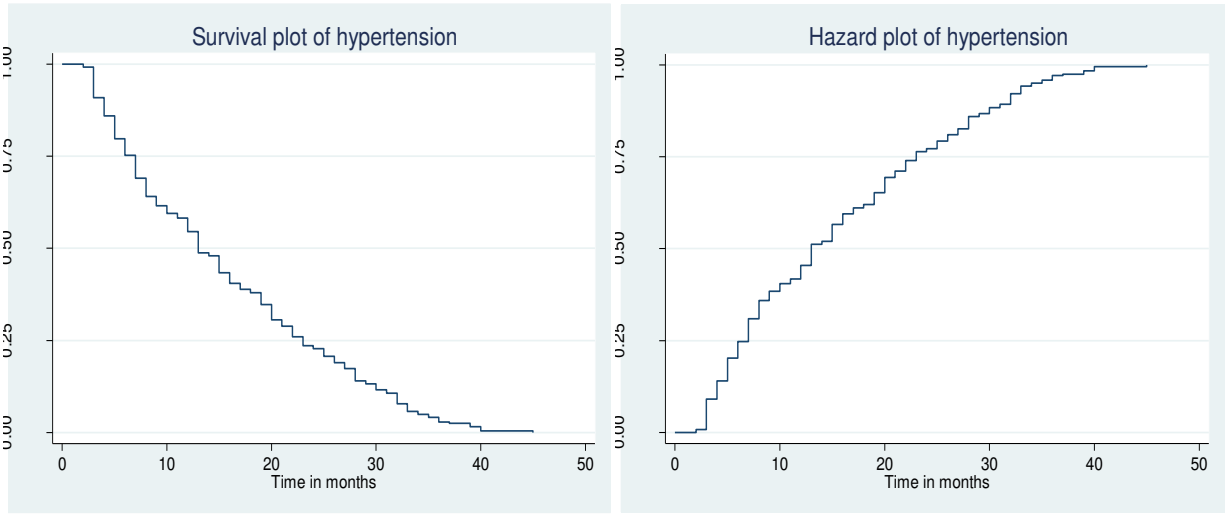

Figure 1: The K-M plots of Survival and hazard functions of hypertension patients

3.2.2 Survival Function of Different Categorical variables

Descriptive graphs of the survivor function were used for the purpose of comparing the event experiencing times of two or more groups and the survival quantities of covariates to describe the survival experience of an individual at specific times. Separate Kaplan-Meier survivor functions were constructed for different covariates to see for possible existence of differences in survival experience between the indicated categories. In general, the pattern of one survivorship function lying above another means the group defined by the upper curve had a better survival than the group defined by the lower curve. In this case the event is recovery time, group defined by the lower curve attend normal condition of hypertension faster (have a better recovery time).

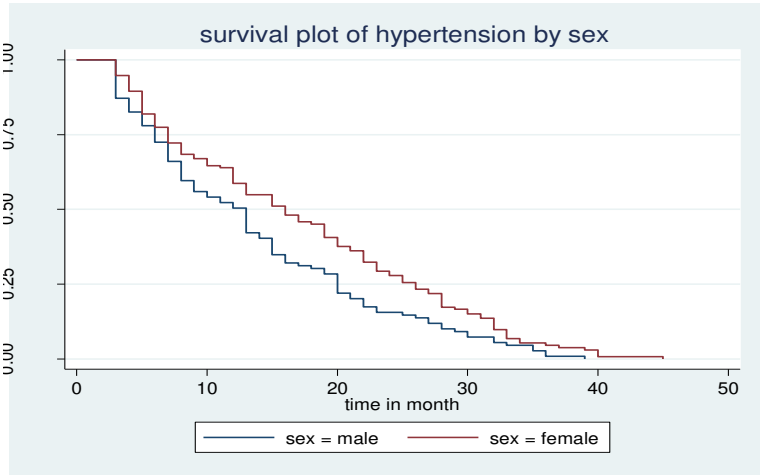

Figure 2: K-M survival plot of hypertension data set by sex

The survival curves in Figure 2 above show the survival distribution of survival time of hypertension patients by sex. The curve of female patients lying above as compared to male patients indicates that male patients attained good control of hypertension faster than female patients. This means, the probability of prolonging recovering time at a given time for female patients is greater as compared to male patients.

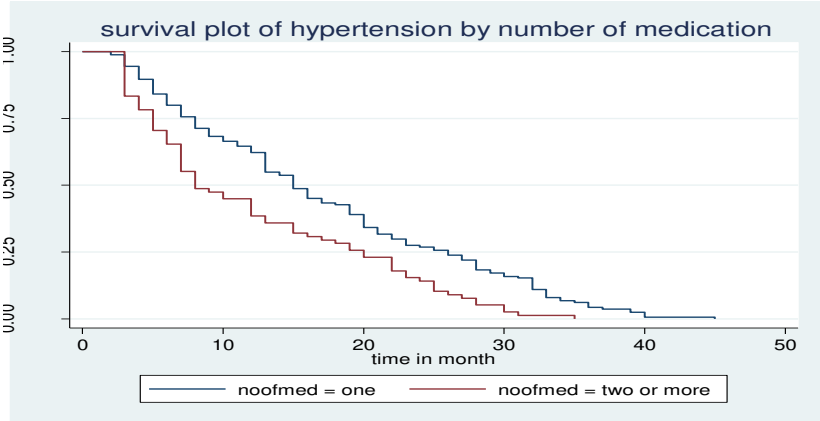

Figure 3: K-M survival plot of hypertension data set by number of medication

The survival of the patients who were using one medication is greater as compared to the survival of the patients who were used two or more medications. That is, the probability of prolonging recovering time at a given time for patients received one medication is greater as compared to those who received two or more medication group. (Figure 3)

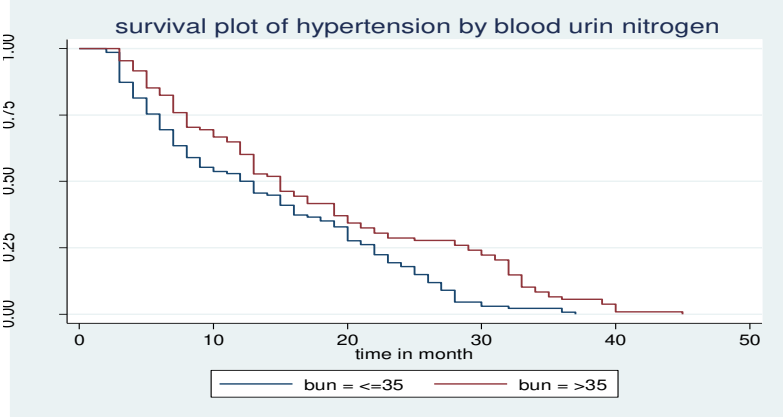

Figure 4: K-M survival plot of hypertension patients by blood urea nitrogen

The survival function against survival time for recovery time of hypertension patients by blood urea nitrogen is shown in the Figure 4 above. This plot suggested that the length of recovery time was greater for patients whose blood urea nitrogen is >35 (mg/dL) compared to those whose

BUN is ≤35 (mg/dL) (i.e. patients whose blood urea nitrogen ≤35 (mg/dL) have shorter recovery time compared to those have blood urea nitrogen >35(mg/dL).

3.2.3. Comparison of Survival Experiences of Recovery Time

It is vital to do some statistical tests that will be used as initiation to our subsequent findings. Here we start with the test of equality of probabilities across different groups of categorical variables using Log Rank and Breslow tests. The results of both log-rank and Breslow test for survival difference were highly significant showed **table3** below. By using both tests there were significant differences in survival experience among groups of sex, residence, related disease, creantine, blood urea nitrogen and number of medications. These showed that all categorical variables had statistically significant differences in survival probabilities.

Table 3: Results of log-rank and Wilcoxon test for each categorical variable of hypertension

		Log-rank test		Generalized Wilcoxon (Breslow)	
Covariates/Factors	DF	Chi-Square	p-value	Chi-square	p-value
Sex	1	13.83	0.002	33.67	0.000
Residence	1	255.93	0.000	216.56	0.000
Related disease	3	256.74	0.000	188.52	0.000
Creantine	1	82.26	0.000	81.06	0.000
BUN	1	222.3	0.000	209.71	0.000
Number of medication	1	40.2	0.000	46.69	0.000

3.3 Cox proportional hazards model

3.3.1 Univariable Analysis

A univariable analysis was performed in order to see the effect of each covariate on recovery time of hypertension patients and to select variables to be included in the multivariable analysis. The result of univariable analysis indicated that age, sex, residence, SBP, DBP, creantine, blood urea nitrogen and number of medications were significant for Cox proportional hazards model at 25% level of significance. Therefore all predictor variables were included in the multivariable analysis.

3.3.2 Multivariable analysis

The multivariable analysis of recovery time of hypertension patients using Cox proportional hazards model was fitted by including all the covariates significant in the univariable analysis at 25% level of significance. Covariates which become insignificant in the multivariable analysis were then removed one by one from the model starting from the largest p-value by using purposeful variable selection technique. Accordingly, the covariates age, sex, residence, systolic blood pressure, related disease, creantine, blood urea nitrogen and the interaction between age and sex were significant at 5% level of significance (Appendix B).

3.3.3 Assumption checking for Cox proportional hazard model

While the Goodness of fit testing approach is employed for Cox PH model, it provides a test statistic and p-value for assessing the PH assumption for a given predictor of interest. Then the PH assumption was checked by using graphical method, adding time dependent covariates in the Cox model and tests based on the Schoenfeld residuals.

3.3.3.1 Test of proportional hazard assumption by generating time varying covariates

Generating time varying covariates by creating interactions of the predictors and a function of survival time is another method of checking proportional hazard assumption. The output in appendix B Table 5 indicates that the p-value of the covariates sex, related disease, creantine, and blood urea nitrogen was less than 0.05, indicating that the proportional hazard assumption was not satisfied.

3.3.3.2 Test of proportional hazard assumption by schoenfeld residual

The Schoenfeld residual is one of the methods used to check the PH assumption. In this study the p-value was checked for testing the assumption is fulfilled or not. The global test result in (Appendix B Table 6) shows that p-value is significant (p-value=0.021). This implied that there is evidence to contradict the proportionality assumption. So the proportionality assumption is not fulfilled.

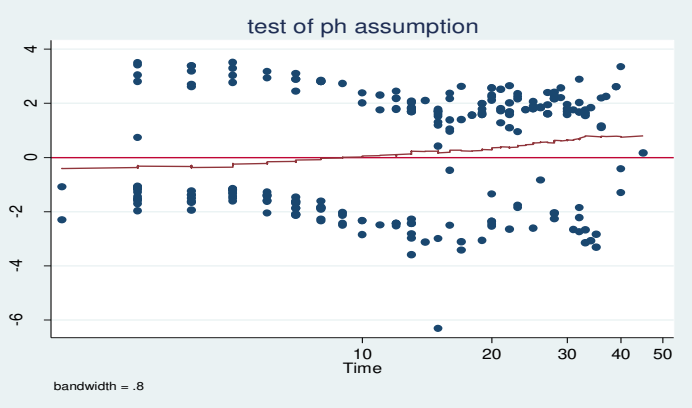

Figure 5: Plots of Scaled schoenfeld residuals by sex to check proportional hazard assumption.

Figure 5 given above displayed that, lines that appeared to be not parallel (cross each other) implying that the proportional hazards assumption has been violated.

3.4. Accelerated Failure Time Model

Since proportional hazards assumption was not satisfied, the accelerated failure time model is an alternative model for the analysis of this data. We fitted the data using accelerated failure time model with Weibull, Lognormal and Log-logistic as a baseline distribution. For each model fit, we did both univariable and multivariable analysis.

3.4.1 Univariable Analysis

A univariable analysis was performed to select variables to be included in the multivariable analysis. The univariable analysis was fitted for each covariate by AFT model using different baseline distributions. From the outputs in univariable analysis (Appendix A), age, sex, systolic blood pressure, diastolic blood pressure, related disease, creantine, blood urea nitrogen and number of medication were significant at 25% level of significance in all base line distributions. Hence, based on the univariable analysis these variables were included in the multivariable analysis.

3.4.2 Multivariable analysis and Model Comparison

Multivariable analysis of Weibull, lognormal, and log-logistic models was done by using all significant covariates at univariable analysis. The output of multivariable analysis in appendix B indicates that age, related disease, creantine, blood urea nitrogen and the interaction between age and blood urea nitrogen were significant at 5% level of significant in all AFT models. The covariate systolic blood pressure was significant in Weibull and lognormal AFT models. While,

diastolic blood pressure and the interaction between blood urea nitrogen and systolic blood pressure were significant in Log logistic model. Model comparison was done using the covariates that are significant in multivariable analysis. AIC and BIC were used to compare models. The best model is the one with the lowest value of AIC and BIC. From **Table 4** below the AIC and BIC values of Weibull were the smallest (170.58 and 205.46) implying that Weibull accelerated failure time model was the best model to describe the hypertension data.

Table 4: Comparison of AFT models using AIC and BIC criteria for hypertension data set

AFT Models	AIC	BIC
Weibull	170.58	205.46
Log-logistic	182.71	221.09
Log-normal	201.72	250.57

AIC= (Akaike's Information Criteria) and BIC= (Bayesian Information Criteria)

An acceleration factor greater than one (positive coefficient) indicates extending the recovery time while an acceleration factor less than one (negative coefficient) indicates shortened recovery time. Based on **Table 5** the predictors: age, systolic blood pressure, blood urea nitrogen>1.5 mg/dL and creantine>35 mg/dL had positive coefficient indicating that these factors extend recovery time. Patients having kidney disease, other disease and have no disease had acceleration factor less than one (negative coefficient) indicating that they have shorter recovery time as compared to diabetic patients (ref). The interaction between blood urea nitrogen and age had acceleration factor less than one indicate that this factor shorten the recovery time.

A statistical test for H_0: $\log(p) = 0$ yields a p-value of 0.000.We would reject the null hypothesis and decide p is not equal to 1, suggesting that the exponential model is not appropriate.

Table 5: Multivariable result of Weibull AFT model for recovery time of hypertension patients

Covariates	Categories	Coef	Φ	95%CI	SE($\hat{\beta}$)	P-value
Age		.014	1.014	(1.01,1.018)	.002	.000
Systolic blood pressure		.002	1.002	(1.00,1.004)	.001	.038

Related disease	Diabetes	Ref				
	kidney	-.148	.862	(.756,.983)	.067	.027
	other	-.458	.633	(.549,.730)	.073	.000
	none	-.628	.534	(.456,.624)	.08	.000
Creantine	≤1.5	Ref				
	>1.5	.213	1.237	(1.141,1.341)	.041	.000
Blood urea nitrogen	≤35	Ref				
	>35	1.153	3.168	(2.339,4.29)	.155	.000
Blood urea nitrogen with age	≤35					
	>35	-.013	.987	(.982,.992)	.003	.000
log(p)		1.269				.000
P=3.556	λ=.003			AIC=170.58		

P=Shape parameter, λ= scale parameter, Coef: estimated coefficient, ϕ Indicates Acceleration factor; 95%CI: 95% confidence interval for acceleration factor, S.E (Coef.): standard error for the coefficient, Ref: reference

3.5 Parametric Shared Frailty Model Results

In the previous section, three AFT models were fitted and compared to analyze the survival time of hypertension patients. To identify baseline distribution this study used Akaike Information Criteria and Bayesian Information Criteria. As a resultshowed **in table 6**, Weibull accelerated failure time model was selected as a baseline distribution for parametric shared frailty model since it has minimum AIC and BIC. This study was checked the heterogeneity using residence as frailty term and investigate risk factors associated with the recovery time of hypertension patients. The variance of the frailty was significant for all baseline distributions with gamma shared frailty distribution and inverse Gaussian shared frailty distribution with Weibull baseline distribution at 5% level of significance. This shows the presence of heterogeneity and necessitates for the frailty models.

The variance parameter θ in a shared frailty models indicated a measure of the degree of correlation and provides information on the variability in the population of clusters. The Kendall's tau (τ) is used to measure the association within residence and it is directly related with variance of random effect (θ). The Weibull-inverse Gaussian shared frailty model had the smallest AIC (151.08) and BIC (189.47). This indicated that Weibull-inverse Gaussian shared

frailty model was the most efficient model to describe the hypertension dataset with the value of variance parameter θ was 0.385.

Table 6: Comparison of gamma and Inverse Gaussian shared frailty model with different baseline distribution

Baseline Distribution	Frailty Distribution	AIC	BIC
Weibull	Gamma	151.19	189.56
	Inverse Gaussian	151.08	189.46
Lognormal	Gamma	192.01	233.87
log logistic	Gamma	165.46	207.33

The result of Weibull-Inverse Gaussian shared frailty model is given below in **Table 7**. From this result the frailty term θ=0.385 indicates that there is presence of heterogeneity between residence (urban and rural). A variance of zero (θ= 0) would indicate that the frailty component does not contribute to the model. A likelihood ratio test for the hypothesis θ= 0 is shown in Table below indicates a chi-square value of 21.48 with one degree of freedom resulting a highly significant with p-value of 0.000. This implied that the frailty component had significant contribution to the model. The Kendall's tau (τ) is used to measure the association within the residence. From the results of this study the values of Kendall's tau (τ) for the Weibull-Inverse Gaussian frailty was 0.141. The estimated value of shape parameter of Weibull-Inverse Gaussian frailty was (p=3.803). This indicates that the shape of hazard functions is increases up as time increase since its value is greater than one.

Table 7: Results of Weibull-Inverse Gaussian shared frailty model

Covariates	Categories	Coef	Φ	95%CI	SE($\hat{\beta}$)	P-value
Age		.013	1.013	(1.009,1.017)	.002	.000
Systolic blood pressure		.002	1.002	(1.00,1.004)	.001	.022
related disease	Diabetes	Ref				
	kidney	-.152	.859	(.759,.971)	.063	.015
	other	-.465	.628	(.549,.718)	.068	.000
	none	-.586	.557	(.478,.649)	.078	.000
Creantine	≤1.5	Ref				

	>1.5	.133	1.142	(1.051,1.241)	.042	.002
Blood urea nitrogen	≤35	Ref				
	>35	.968	2.632	(1.97,3.517)	.148	.000
Blood urea nitrogen with age	≤35	Ref				
	>35	-.012	.988	(.983,.993)	.002	.000

Θ=.385 τ= .141 *P=3.803* λ=.001 AIC=151.08 BIC=189.4
Likelihood-ratio test of theta=0: chibar2(01) = 21.48 Prob>=chibar2 = 0.000

θ= variance of the random effect, τ = Kendall's tau p =shape parameter, λ=scale parameter, AIC= Akaike Information Criteria, prob=probability, chibar2= Chi-square.

3.6 Comparison of Weibull AFT and Weibull-Inverse Gaussian shred frailty models

In this study, in order to compare the efficiency of the models AIC and BIC were used. As showed **Table 8** below, the AIC and BIC values of Weibull-Inverse Gaussian shared frailty model (AIC=151.08 and BIC=189.47) was smaller than Weibull AFT (AIC=170.58 and BIC=205.46). These indicate that the Weibull-Inverse Gaussian shared frailty model is best model for hypertension dataset. Weibull-Inverse Gaussian shared frailty model take in to account the clustering effect (θ is significant at 5% level of significant), for this data Weibull-Inverse Gaussian shared frailty is better fit than Weibull accelerated failure time model.

Table 8: AIC and BIC value of Weibull AFT and Weibull-Inverse Gaussian shared frailty models

AFT Models	AIC	BIC
Weibull AFT	170.58	205.46
Weibull-Inverse Gaussian	151.08	189.47

3.7 Model diagnosis

After the model has been fitted, it is important to determine whether a fitted parametric model adequately describes the data or not. The study used two different methods to diagnose the parametric baselines.

3.7.1 Diagnostic Plots of the Parametric Baselines

To check the adequacy of the baseline accelerated failure time models with baseline Weibull, baseline log logistic and baseline log-normal can be graphically evaluated by plotting log(−log

$(S(t)))$, $\log((\frac{1-\hat{s}(t)}{\hat{s}(t)}$ vs log time and $\phi^{-1}[1-s(t)]$ against log (t) respectively. If the plot of parametric model is approximately linear, the given baseline distribution is appropriate for the given dataset. Then based on the plots given in figure 6 below, the plot for the Weibull baseline distribution is approximately a straight line compare to that of log-logistic and lognormal baseline distribution. This evidence also support the decision made comparing baseline AFT models by using AIC value that Weibull baseline distribution is appropriate for the given dataset.

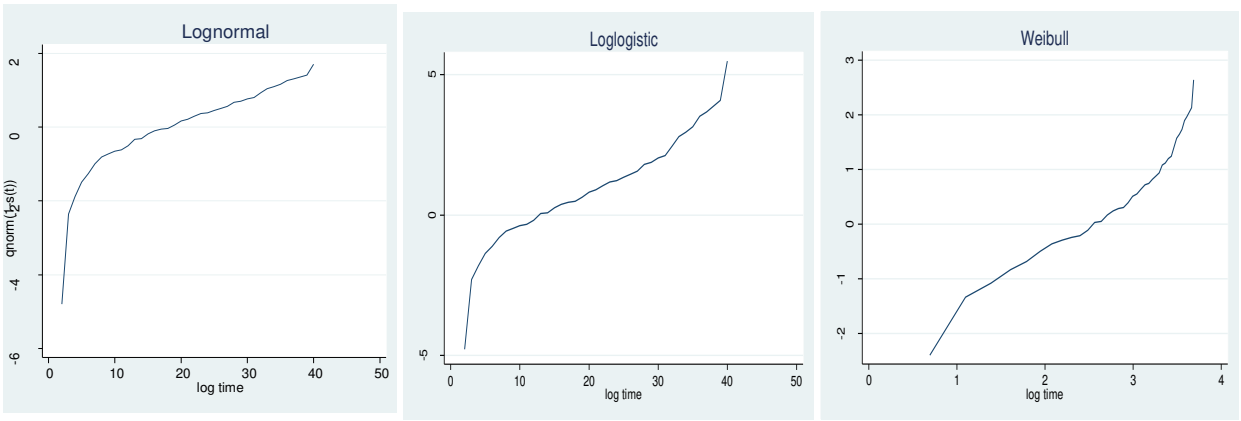

Figure 6: Graphical evaluation of the lognormal, log-logistic and Weibull, baselines

3.7.2 Cox-Snell residual plots

The Cox-Snell residuals method can be applied for parametric models. The Cox-Snell residuals are important to investigate how well the model fits the data. In this study we used the Cox-Snell residuals to check the overall goodness of fit for various parametric models. The Cox- Snell plot of Figure7 given below indicates that the Weibull AFT model fit the data better compare to log normal and log logistic AFT models. Nevertheless, the results of Cox-Snell were consistent with the results based on AIC and BIC. The plot makes approximately straight line through the origin for Weibull baseline distribution suggesting that it is appropriate for hypertension dataset.

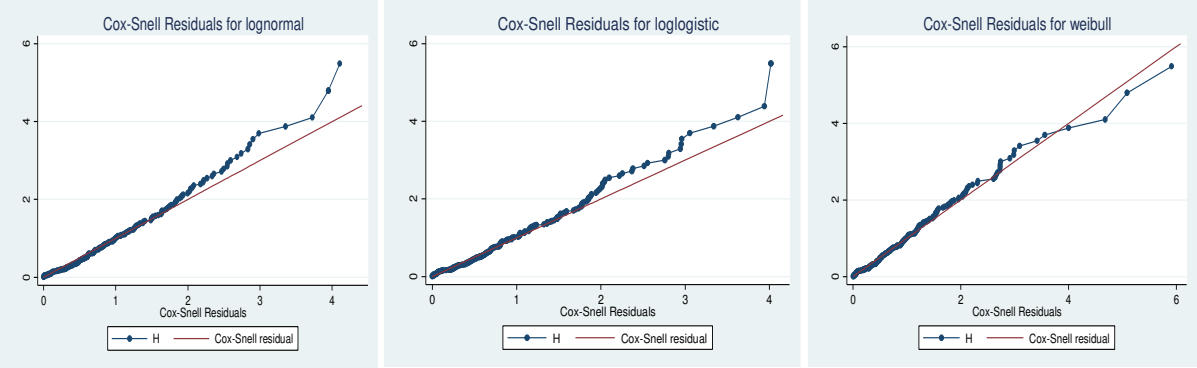

Figure 7: Cox-Snell residual plots for log normal, log logistic and Weibull AFT mode

3.8 Interpretation of Weibull-Inverse Gaussian shared frailty model results

From Weibull-Inverse Gaussian shared frailty model results in Table 7 above, controlling for other variables, Acceleration factor, 95% CI and p-value of the variable systolic blood pressure were 1.002, (1.00, 1.004) and .022 respectively. These indicated that systolic blood pressure was significant factors for the recovery time and for a one mmHg increase in systolic blood pressure the recovery time of hypertensive patients were prolonged by 1.002 times in month.

Acceleration factor and its 95% CI for related disease category of kidney, others and none were (ϕ =.859, 95% CI=0.760, 0.971), (ϕ= .628, 95% CI=0.550, 0.718) and (ϕ=.557, 95% CI=0.478, 0.649) respectively. This showed that patients having kidney disease, other and have no related disease shortened the recovery time by a factor of .859, .628 and .557 respectively compared to diabetes patients (Reference).The 95% CI did not include one showed that related disease was significantly important factor for the recovery time at 5% level of significance.

Acceleration factor and its 95% CI for the variable creantine were 1.142 and (1.051, 1.241) respectively. These indicated that creantine was significant factors for the recovery time and Showed that the recovery time of patients, who have had creantine level >1.5 mg/dL, were prolonged by 1.142 times as compared to patients those who have had creantine level ≤1.5 mg/dL.

The interaction between blood urea nitrogen and age had statistically significant effect on the recovery time, suggesting that as age increased by one year, the acceleration factor for patients who have had blood urea nitrogen >35 mg/dL was prolonged by exp (.013-.012) =1.001 with p value of .000 as compared to those patients who have had blood urea nitrogen ≤35 mg/dL.

4 Discussions

Hypertension remains a major public health problem that is associated with morbidity and mortality and the prevalence is still increasing (Chobanian, 2015). Therefore, this study attempted to identify factors that affect time-to-recovery from hypertension patients at FHRH, Bahir Dar using survival models. Factors concerned for this study were age, sex, residence, systolic blood pressure, diastolic blood pressure, related disease, creantine, blood urea nitrogen and number of medication. Lognormal, Weibull and log-logistic AFT distributions and gamma and inverse Gaussian frailty distributions were used to analyze this survival data.

Both univariable and multivariable analysis was performed to examine factors that affect recovery time of hypertension patients. The univariable analysis given in Appendix A revealed that variables age, sex, SBP, DBP, related disease, creantine, BUN and number of medication were significant at 25% level of significance. All significant covariates in univariable analysis were included in multivariable analysis.

This research showed that there was a clustering (frailty) effect on recovery time of hypertension patients which might be due to the heterogeneity in residence. Then, patients coming from the same residence share similar risk factors related to recovery time, and indicating that it was important to consider the cluster effect.

Comparison of models for different multivariable analysis was done by using AIC and BIC criteria. In this study, Weibull-inverse Gaussian shared frailty model had the smallest AIC and BIC i.e., 150.08 and 189.47 respectively, which is appropriate model for describing recovery time of patients with hypertension. The obtained results were discussed as follows:

The finding of this study showed that systolic blood pressure had positive significant effect on recovery time that is as systolic blood pressure increase the recovery time of hypertension patients also prolonged. This finding was consistent with (Gesese, 2017; Xiong *et al.,* 2013).

Related disease is another prognostic factor that significantly predicts the recovery time of patients with hypertension. And showed that patients having kidney disease, other and have no related disease shortened the recovery time by a factor of .859, .628 and .557 respectively compared to diabetes patients (Reference). This study is supported by (Workie *et al.,* 2017).

Another potential risk factor that accelerates recovery status of patients from hypertension was creantine and showed that the recovery time of hypertensive patients, who have had creantine

>1.5 mg/dL, were prolonged by 1.142 times compared to patients those who have had creantine ≤1.5 mg/dL. This study was consist with (Zhang *et al.,* 2018; Xiong *et al.,* 2013).

The results of this study also suggested that the interaction between blood urea nitrogen and age had statistically significant effect on the recovery time, suggesting that as age increase by one year, the acceleration factor for patients who have had blood urea nitrogen level >35 mg/dL was prolonged by 1.001 as compared to those who have had blood urea nitrogen ≤35 mg/dL.

The related literatures have not been done by using interaction effect of variables in the area of hypertension. So, this study is unique that interaction effect of variables was assumed to see whether the effect one variable is different or not on the acceleration factor of another variable.

In this study, the time of recovery for patients with hypertension did not depend on sex which is similar with the study of (Belachew *et al.,* 2018; Angaw *et al.*, 2015; Mungati *et al.,* 2014) suggested that sex was not statistically associated with hypertension using logistic regression model in Northwest Ethiopia , Addis Ababa and Zimbabwe respectively.

5 Conclusions

This study was based on hypertension patient dataset which is obtained from FHRH, Bahir Dar with an aim to identify predictors of recovery time of patients by applying survival models.

Parametric shared frailty models with different parametric baseline distributions and AFT models were applied. Among this using AIC and BIC, Weibull-inverse Gaussian shared frailty model is the best fitted model for the recovery time of patients with hypertension. There was a frailty (clustering) effect on time-to-recovery from hypertension that arises due to differences in distribution of timing of recovery among residence in Felege Hiwot referral hospital. This indicates the presence of heterogeneity and necessitates for the frailty models.

The result of Weibull-inverse Gaussian shared frailty model showed systolic blood pressure, related disease, creantine and the interaction between blood urea nitrogen and age were found significant predictors for the recovery time of patients among patients with hypertension in Felege Hiwot referral hospital. Among these significant predictors, systolic blood pressure and creantine prolong timing of recovery while, related disease and the interaction between blood urea nitrogen and age shorten timing of recovery for patients with hypertension. On the other hand, sex, diastolic blood pressure and number of medication were not significant predictors for the recovery time of patients with hypertension.

Ethics approval and consent to participate

The authors got an ethical approval certificate from Bahir Dar University Ethical approval committee, Bahir Dar, Ethiopia with Ref≠ BDU/54/2018 to use the secondary data related to patients. The Ethical approval certificate obtained in this committee can be attached up on request. Since the data used in the current investigation was secondary, there was no verbal or written consent from the participants.

Consent for publish

Not applicable.

Availability of data and materials

All data supporting the findings are contained in the manuscript. Anyhow, data are available from the corresponding author on rational request.

Competing interests

The author declares that they have no competing interests

Funding

No funding was received to provision the writing of this paper.

Authors' contribution

Nigist mulu and Yeshambel kindu wrote the proposal, analysis the data and draft the paper. Yeshambel Kindu prepared the first draft of the manuscript. Yeshambel Kindu and Abay Kassie reviewed the manuscript .all authors read and approved the final manuscript.

Acknowledgements

I would like to acknowledge for support from Felege Hiwot Refral hospital pediatric clinic nurses and doctor's for their invaluable guidance and support

Author details

Nigist Mulu Takele [1] Yeshambel Kindu Yihuna[2] Abay Kassie Lakew[3]

[1] Department of Statistics, College of Natural and Computational Sciences, P.O. Box: 79, Bahir Dar University, Bahir Dar, Ethiopia, E-mail nigistmu@gmail.com, Phone number +251-945-280-724

[2]Department of Statistics, College of Natural and Computational Sciences, P.O. Box: 90, Debark University, Debark, Ethiopia, yeshambelyk123@gmail.com Phone number +251-925-981-375

[3]Department of Statistics, College of Natural and Computational Sciences, P.O. Box: 90, Debark University, Debark, Ethiopia abay.kassie@yahoo.com

REFERENCE

Abebe, S. M., et al. (2015). "Prevalence and associated factors of hypertension: a crossectional community based study in Northwest Ethiopia." PloS one **10**(4): e0125210.

Abrahantes, J. C., et al. (2007). "Comparison of different estimation procedures for proportional hazards model with random effects." Computational statistics & data analysis **51**(8): 3913-3930.

Alwan, A. (2011). Global status report on noncommunicable diseases 2010, World Health Organization.

Angaw, K., et al. (2015). "Prevalence of hypertension among federal ministry civil servants in Addis Ababa, Ethiopia: a call for a workplace-screening program." BMC cardiovascular disorders **15**(1): 76.

Animut, Y., et al. (2018). "Blood pressure control status and associated factors among adult hypertensive patients on outpatient follow-up at University of Gondar Referral Hospital, northwest Ethiopia: a retrospective follow-up study." Integrated blood pressure control **11**: 37.

Anteneh, Z. A., et al. (2015). "Prevalence and correlation of hypertension among adult population in Bahir Dar city, northwest Ethiopia: a community based cross-sectional study." International journal of general medicine **8**: 175.

Bayray, A., et al. (2018). "Magnitude and risk factors for hypertension among public servants in Tigray, Ethiopia: A cross-sectional study." PloS one **13**(10): e0204879.

Beard, R. E. (1959). Note on some mathematical mortality models. Ciba Foundation Symposium-The Lifespan of Animals (Colloquia on Ageing), Wiley Online Library.

Bell, K., et al. (2015). "Hypertension: The silent killer: updated JNC-8 guideline recommendations." Alabama Pharmacy Association: 1-8.

Benetos, A., et al. (2016). "An expert opinion from the European Society of Hypertension–European Union Geriatric Medicine Society Working Group on the management of hypertension in very old, frail subjects." Hypertension **67**(5): 820-825.

Bennett, S. (1983). "Log-logistic regression models for survival data." Applied Statistics: 165-171.

Blencowe, H., et al. (2012). "National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications." The Lancet **379**(9832): 2162-2172.

Bloch, M. J. (2016). "Worldwide prevalence of hypertension exceeds 1.3 billion." Journal of the American Society of Hypertension: JASH **10**(10): 753.

Bloom, D. E., et al. (2011). From burden to" best buys": reducing the economic impact of non-communicable disease in low-and middle-income countries, Program on the Global Demography of Aging.

Boutayeb, A. and S. Boutayeb (2005). "The burden of non communicable diseases in developing countries." International journal for equity in health **4**(1): 2.

Chamberlain, G., et al. (1985). "Longitudinal analysis of labor market data." Heterogeneity, Omitted Variable.

Chobanian, A. V. (2015). "Time to reassess blood-pressure goals." New England Journal of Medicine **373**(22): 2093-2095.

Chobanian, A. V., et al. (2003). "Seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure." Hypertension **42**(6): 1206-1252.

Choi, Y.-H., et al. (2014). Prediction of hypertension based on the genetic analysis of longitudinal phenotypes: a comparison of different modeling approaches for the binary trait of hypertension. BMC proceedings, BioMed Central.

Clayton, D. G. (1978). "A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence." Biometrika **65**(1): 141-151.

Collett, D. (2003). modeling survival data in medical research

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

Cox, D. R. (1972). "Regression models and life-tables." Journal of the Royal Statistical Society: Series B (Methodological) **34**(2): 187-202.

Cox, D. R. and D. Oakes (1984). "Analysis of survival data (Vol. 21)." Chapman & Hall/CRC **1**: 984.

Cox, D. R. and E. J. Snell (1968). "A general definition of residuals." Journal of the Royal Statistical Society. Series B (Methodological): 248-275.

Datwyler, C. a. S., T. (2011). "Parametric Survival Model. Handout.".

Duchateau, L. and P. Janssen (2008). "the frailty model (Statistics for Biology and Health)." Springer: New York **10**: 978-970.

Enki, D. G., et al. (2014). "A time-varying shared frailty model with application to infectious diseases." The Annals of Applied Statistics: 430-447.

Falkner, B., et al. (1981). "Cardiovascular characteristics in adolescents who develop essential hypertension." Hypertension **3**(5): 521-527.

Fine, J. P., et al. (2003). "A simple estimator for a shared frailty regression model." Journal of the Royal Statistical Society: Series B (Statistical Methodology) **65**(1): 317-329.

Frank Emmert-Streib ., et al. (2008).” Introduction to Survival Analysis in Practice.” Mach. Learn. Knowl. Extr. 2019, 1, 1013–1038

Gaziano, T. A., et al. (2010). "Growing epidemic of coronary heart disease in low-and middle-income countries." Current problems in cardiology **35**(2): 72-115.

G. Rodr´ıguez. “Survival Models Revised September”, 2010

Gesese, M. T. "Survival Analysis of Time to Cardiovascular Disease Complication of Hypertensive Patients at Felege Hiwot Referal Hospital in Bahir-Dar, Ethiopia: A Retrospective Cohort Study."

Gesese, M. T. (2017). "Survival Analysis of Time to Cardiovascular Disease Complication of Hypertensive Patients at Felege Hiwot Referal Hospital in Bahir-Dar, Ethiopia: A Retrospective Cohort Study." **10(3): 1-9**.

Gupta, R. C., et al. (1999). "A study of log-logistic model in survival analysis." Biometrical Journal: Journal of Mathematical Methods in Biosciences **41**(4): 431-443.

Henderson, R. and P. Oman (1999). "Effect of frailty on marginal regression estimates in survival analysis." Journal of the Royal Statistical Society: Series B (Statistical Methodology) **61**(2): 367-379.

Hendriks, M. E., et al. (2012). "Hypertension in sub-Saharan Africa: cross-sectional surveys in four rural and urban communities." PloS one **7**(3): e32638.

Hosmer, D., et al. (2008). "Model development." Applied survival analysis: regression modeling of time-to-event data. 2nd ed. Hoboken (NJ): Wiley-Interscience: 132-168.

Hougaard, P. (1984). "Life table methods for heterogeneous populations: distributions describing the heterogeneity." Biometrika **71**(1): 75-83.

Hougaard, P. (1986). "Survival models for heterogeneous populations derived from stable distributions." Biometrika **73**(2): 387-396.

Hougaard, P. (2000). "Analysis of Multivariate Survival Data Springer-Verlag New York Google Scholar."

Kadane, J. B. and N. A. Lazar (2004). "Methods and criteria for model selection." Journal of the American statistical Association **99**(465): 279-290.

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

Kaplan, N. M. (2010). Kaplan's clinical hypertension, Lippincott Williams & Wilkins.

Kibret, K. T. and Y. M. Mesfin (2015). "Prevalence of hypertension in Ethiopia: a systematic meta-analysis." Public Health Reviews **36**(1): 14.

Kleinbaum, D. G. and M. Klein (2005). "Competing risks survival analysis." Survival Analysis: A self-learning text: 391-461.

Lambert, P., et al. (2004). "Parametric accelerated failure time models with random effects and an application to kidney transplant survival." Statistics in medicine **23**(20): 3177-3192.

Lancaster, T. (1992). The econometric analysis of transition data, Cambridge university press.

McAloon, C. J., et al. (2016). "The changing face of cardiovascular disease 2000–2012: An analysis of the world health organisation global health estimates data." International journal of cardiology **224**: 256-264.

McConnell, T. H. (2013). The nature of disease: pathology for the health professions, Lippincott Williams & Wilkins.

Misganaw, A., et al. (2017). "National mortality burden due to communicable, non-communicable, and other diseases in Ethiopia, 1990–2015: findings from the Global Burden of Disease Study 2015." Population health metrics **15**(1): 29.

Mittal, B. V. and A. K. Singh (2010). "Hypertension in the developing world: challenges and opportunities." American Journal of Kidney Diseases **55**(3): 590-598.

Moeschberger, M. L. (2003). Survival analysis: techniques for censored and truncated data, Springer.

Molla, M. (2015). "Systematic reviews of prevalence and associated factors of hypertension in Ethiopia: finding the evidence." Sci J Public Health **3**(4): 514-519.

Munda, M., et al. (2012). "Parfm: parametric frailty models in R." Journal of Statistical Software **51**(11): 1-20.

Mungati, M., et al. (2014). "Factors affecting diagnosis and management of hypertension in Mazowe District of Mashonaland Central Province in Zimbabwe: 2012." BMC cardiovascular disorders **14**: 102-102.

Musinguzi, G., et al. (2018). "Factors Influencing Compliance and Health Seeking Behaviour for Hypertension in Mukono and Buikwe in Uganda: A Qualitative Study." International journal of hypertension **2018**.

Ogah, O. S. and B. L. Rayner (2013). "Recent advances in hypertension in sub-Saharan Africa." Heart **99**(19): 1390-1397.

Onyango, M. J., et al. (2017). "A study to determine the prevalence and factors associated with hypertension among employees working at a call centre Nairobi Kenya." The Pan African medical journal **27**.

Oo, W. M., et al. (2017). "Health literacy-is it useful in prevention of behavioral risk factors of NCDs?" International Journal of Research in Medical Sciences **3**(9): 2331-2336.

Organization, W. H. WHO. Global Action Plan for the Prevention and Control of NCDs 2013-2020 [Internet]. WHA66. Geneva: World Health Organization; 2013 [cited 2014 Jan 27]. 55 p.

Organization, W. H. (2009). Global tuberculosis control: epidemiology, strategy, financing: WHO report 2009, World Health Organization.

Organization, W. H. (2013). Global action plan for the prevention and control of noncommunicable diseases 2013-2020, World Health Organization.

Organization, W. H. (2014). Global status report on noncommunicable diseases 2014, World Health Organization.

Organization, W. H., et al. (2005). Preventing chronic diseases: a vital investment, World Health Organization.

Sastry, N. (1997). "Family-level clustering of childhood mortality risk in Northeast Brazil." Population Studies **51**(3): 245-261.

Sudarno ,and A Prahutama 2019 J. Phys.: Conf. Ser. 1217 012111

Tadesse, K., et al. (2018). "Prevalence of Hypertension among Patients with Type 2 Diabetes Mellitus and Its Socio Demographic Factors in Nigist Ellen Mohamed Memorial Hospital Hosanna, Southern Ethiopia." J Diabetes Metab **9**(4): 792.

Twagirumukiza, M., et al. (2011). "Current and projected prevalence of arterial hypertension in sub-Saharan Africa by sex, age and habitat: an estimate from population studies." Journal of hypertension **29**(7): 1243-1252.

Vaupel, J. W., et al. (1979). "The impact of heterogeneity in individual frailty on the dynamics of mortality." Demography **16**(3): 439-454.

Vaupel, J. W. and A. I. Yashin (1983). "The deviant dynamics of death in heterogeneous populations."

Weldearegawi, B., et al. (2013). "Emerging chronic non-communicable diseases in rural communities of Northern Ethiopia: evidence using population-based verbal autopsy method in Kilite Awlaelo surveillance site." Health policy and planning **28**(8): 891-898.

Who, A. (2013). "Global brief on hypertension." World Health Organization.

Wienke, A. (2003). "Frailty models." Max Planck Institute for Demographic Research, MPIDR Working Papers: No. WP-2003-32.

Wienke, A. (2010). Frailty models in survival analysis, Chapman and Hall/CRC.

Workie, D. L., et al. (2017). "Bivariate longitudinal data analysis: a case of hypertensive patients at Felege Hiwot Referral Hospital, Bahir Dar, Ethiopia." BMC research notes **10**(1): 722.

Xiong, X., et al. (2013). "Chinese herbal formulas for treating hypertension in traditional Chinese medicine: perspective of modern science." Hypertension Research **36**(7): 570.

Zhang, Y.-F., et al. (2018). "Postoperative survival analysis and prognostic nomogram model for patients with portal hypertension." World journal of gastroenterology **24**(39): 4499.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- APPENDICEStables.docx
- appendicesfigures.docx