MODELING TIME TO RETURN OF VOLUNTEER BLOOD DONOR'S: IN NATIONAL BLOOD BANK, ADDIS ABABA, ETHIOPIA

By

SEID ALI TAREKE

A Thesis Submitted to the Department of Statistics, College of Natural Science, Jimma University as a Partial Fulfillment for the Requirements of Master of Science(MSc) Degree in Biostatistics

March 24, 2021

Jimma, Ethiopia

MODELING TIME TO RETURN OF VOLUNTEER BLOOD DONOR'S: IN NATIONAL BLOOD BANK, ADDIS ABABA, ETHIOPIA

By

SEID ALI TAREKE

Main Advisor: KADI ADIVEPPA (Prof.))

Co-Advisor: FIKADU ZEWDIE(MSc)

March 24, 2021

Jimma, Ethiopia

STATEMENT OF AUTHOR

I declare that this thesis is a result of my genuine work and all sources of materials used have been duly acknowledged. I have submitted this thesis to Jimma University in the partial fulfillment for the Degree of Master of Science in Biostatistics. The thesis can be deposited in the university library to be made available to borrowers for reference. I solemnly declare that I have not so far submitted this thesis to any other institution anywhere for that award of any academic degree, diploma or certificate.

Brief quotations from this thesis are allowed without requiring special permission provided that an accurate acknowledgement of the source is made. Requisites for extended quotations for the reproduction of the thesis in whole or in part may be granted by the head of the department of statistics when in his or her judgment the proposed use of the material is for a scholarly interest. In all other instances, however, permission must be obtained from the author.

Seid Ali Tareke

Date: _____

Signature: _____

Jimma University, Jimma, Ethiopia

DEPARTMENT OF STATISTICS, SCHOOL OF GRADUATE STUDIES JIMMA UNIVERSITY

As thesis research advisors, we her by certify that we have read the thesis prepared by Seid Ali under our guidance, which is entitled "Modeling Time to Return of Volunteer Blood Donor's: in National Blood Bank, Addis Ababa, Ethiopia", in its final format is found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including tables and figures are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Kadi Adiveppa (Professor)	...For...	23/03/2021
Main advisor	Signature	Date
Fikadu Zewudie (MSc.)	For...	23/03/21.
Co-advisor	Signature	Date

As the members of the board of examiners of MSc thesis open defense examination, we certify that we have read and evaluated the thesis and examined the candidate. Hence, we recommend that the thesis be accepted as it fulfills the requirements for the degree of Master of Science in Biostatistics.

Dr. Habte Tadesse		10/07/2013
External Examiner	Signature	Date
Jaleta Abdisa		23-03-2021
Internal Examiner	Signature	Date
KIBREALEM SISAY		23-03-21
Chairman, Department graduate committee	Signature	Date

ACKNOWLEDGEMENT

First and for most, I would like to extend my unshared thanks to Allah for providing me the opportunity for what I have achieved and for his mercy, wisdom, favor, faithfulness and protection.

I would like to express sincere appreciation to my advisor Prof. Kadi Adiveppa (Professor) and my co-advisor Mr. Fikadu Zewudie (MSc) for their kind support, advice, guidance and constructive comments, beginning from the proposal development up to the final work of this thesis and I will always be grateful for that. My special appreciation and acknowledgment go to Mizan-Tepi University for offering me the opportunity and financial support during my study time.

I would like to extend our gratitude to the National Blood Bank data officers for permitting us to use the data.

Lastly, I take this opportunity to express my gratitude to my beloved family who are the source of pride and encouragement throughout my life. I am thankful to my father Ali Tareke, my mother Abebech Adem, my wife Halima Walelgn, my brothers Idris Ali and Mohammed Jenberu, to all my sisters and all my best friends for your unlimited love, pray and supports.

ABSTRACT

Background: Blood is the most precious and unique gift that one human being can give to another. It is lifesaving fluid that cannot be created artificially, but is only collected from donors which are the precious resources.

Objectives: The study aimed to model time to return of volunteer blood donors' in National blood bank, Addis Ababa, Ethiopian.

Methods: A retrospective cohort study had been conducted at National Blood Bank, Addis Ababa, Ethiopia. In this study, a total of 6,019 voluntary blood donors were included and baseline data had been obtained from existing databases of all volunteer blood donors who donated blood from September 06, 2012 to September 11, 2013 and whether he/she returned or not were followed until September 2015. In this study Exponential, Weibull, log-logistic and lognormal as baseline hazard functions with the gamma and the inverse Gaussian frailty distributions were used.

Results: A total of 6,019 voluntary blood donors donated blood during a period of 06 Sept 2012 to 11 Sept 2013. The median returning time of the donors was 26 months. About 46.7% of the donors were returned to donate blood again during the study period. The clustering effect is significant on modeling time to return of volunteer blood donors. According to the output of the Lognormal Gamma frailty model, gender of the donors', age of the donors', weight, occupation, donation experience, and experiencing donors' reaction were the significant risk factors at 5% level of significance.

Conclusion and Recommendation: Lognormal-Gamma frailty model is the model that best described time to return of the donors' dataset. Being male donor, donation experience (repeat donor) and increasing weight (in Kg) significantly shorten/minimize the time-to-return of blood donors, while being in the age group (45-65), being a student and experiencing donors' reaction prolongs the time to- return of volunteer blood donors. For those groups whose return time were prolonged, policy makers and human resource managers are expected to make interventions, and design appropriate policy, programs and donor motivational strategies to improve their return time.

Key words: Parametric Shared Frailty Model, Heterogeneity, Donors' Return

ACRONYMS/ABBREVIATIONS

ADR	Adverse Donor Reaction
AIDS	Acquired Immune Deficiency Syndrome
AR(s)	Adverse Reaction(s)
BDRR(s)	Blood Donor Return Rate(s)
BP	Blood Pressure
BTS	Blood Transfusion Services
CI	Confidence Interval
DBP	Diastolic Blood Pressure
EC	Ethiopian Calendar
ENBBS	Ethiopian National Blood Bank Service
FMOH	Federal Ministry of Health
HBV	Hepatitis B Virus
HCV	Hepatitis C Virus
HIV	Human Immunodeficiency Virus
NBTS	National Blood Transfusion Services
SBP	Systolic Blood Pressure
SSA	Sub-Saharan Africa
TTIs	Transfusion Transmissible Infections
VVR	Vasovagal Reaction
WB	Whole Blood
WBDD	World Blood Donor Day
WHO	World Health Organization

Contents

List of Tables

List of Figures

1 INTRODUCTION

1.1 Background

Blood is our body fluid that delivers essential substances (nutrients and oxygen) to the cells and exports metabolic waste products away from cells(Nureye, *et al.,* 2019). It is the most precious and unique gift that one human being can give to another. It is lifesaving fluid that cannot be created artificially, but is only collected from donors which are the precious resources(Omura, *et al.,* 2014).

Blood donation is a procedure in which a voluntary and healthy individual has his or her blood drawn for transfusion to the one who needs it. It is considered to be one of the most valuable contributions of an individual to the community. Blood transfusion is vital and fundamental in medical practice since there is no efficient substitute to human blood when needed (Nureye & Tekalign, 2019).

The source for blood to be transfused relies mainly on voluntary non-remunerated blood donors (Jemberu et al., 2016). Even though over a million of blood units are collected every year, many more millions still need to be collected to meet the global demand (Abderrahman & Saleh, 2014). .

The demand for blood and blood products are increasing in all parts of the globe (MELAKU, 2019). Despite increasing demand for blood due to increases in the number of clinical procedures that require transfusions, the number of eligible blood donors is declining. Thus, in many countries there is a widespread shortfall between blood requirements and blood supplies and as a result, many patients die or suffer unnecessarily because they do not have access to blood and blood products (WHO, 2017b).

A study conducted on 4200 blood donors in Shiraz Iran found that, overall return rates of blood donors show a significant yearly decline of 51.2%, 45.1%, 37.6%, 38.8%, 28.7%, 22.7% and 22.1% over a 7 year follow up. The return rates for first time donors declined annually by 30%, 25%, 21%, 18.5%, 15.2%,12.3% and 11.8%(Kasraian et al., 2020).

A study by Schreiber et al. found that only 38.5% of first-time donors returned to donate during a 6-year follow-up period; those authors suggested that first-time donors who returned to donate soon after the first donation were more willing to become regular donors.

In developing countries like Brazil the estimated voluntary blood donors return rate is 40% within one year, 53% within two years, whereas 30% never return to donate (Lourençon, 2011). According to a retrospective study done in Iran, more than half (51.7%) of the first-time donors returned to donate again during the 3 years after the first donation (Kasraian & Tavassoli, 2012). Another study done in the same country which is based on cross-sectional follow-up study for 5 years' data shows that, rate of donors returned to donate again were 49·1% which is approximately the same result (Kheiri & Alibeigi, 2015).

Though WHO estimates that blood donation by 2%-3% of the country's population is needed to meet a nation's most basic requirements for blood, less than 0.5% of Ethiopian's population donates blood(WHO, 2010a). As there are limited repeat blood donors, National blood bank can only meet 52% of hospital demand which lets many patients in need of blood to die or suffer (*Federal Ministry of Health Ethiopia(FMOH), 2017*).

Taking Whole blood donations per 1,000 population as an indicator for the general availability of blood in a country, low-income countries including Ethiopia has a rate of 2.8 donations/1000 population (range 0.4 – 8.2) which is very low as compared to 36.4 donations/1000 population (range 13.3 – 64.6) in developed countries (Fantahun *et al.*, 2017).

Study conducted in Bahir Dar town among civil servants in June 2015 found that, the return rates were 18.1%(Mekonnen et al., 2015), which is not a good record but is better than a study conducted in Mekele town (12.4%).

Understanding return behavior and the factors that affect time to return of volunteer blood donors is essential in order to design new recruitment strategies(Jorge A. Rios, 2013).

In considering the different factors for donors return, the time to return of volunteer blood donors can be predicted and statistically estimated with the survival analysis. Survival data is a term used for describing data that measure the time to a given event of interest (Klein and Moeschberger, 1992). In this study, the event of interest was the time to return of volunteer blood donor's after ones they donate blood. Proportional hazard model popularized by (Cox, 1972) is the classical model for this kind of data. However, the correct inference based on Cox's models needs identically and independently distributed samples.

Moreover, the frailty model, introduced in the statistical literature by (Vaupel, et al., 1979), and

discussed in details by (Hougaard, 2000; Duchateau and Janssen, 2008; and Wienke, 2011), accounts for heterogeneity in baseline. This concept is an extension of the Cox's PH model in which the hazard function depends upon an unobservable random quantity, the so-called frailty that acts multiplicatively on the hazard function.

Furthermore, study population for this study was clustered and hence clustered donor's survival data may be correlated at the cluster (donation sites) level. In current study, shared frailty models explored assuming that donor's with in the same cluster (donation sites) shares similar risk factors. Ignoring the full dependence among observations might lead to understated standard errors and also parameter estimates that are both biased and inconsistent (Sastry, 1997).

Typically, the estimation of the frailty model can be parametric or semi-parametric. In the former case, parametric density is assumed for the event times, resulting in a parametric baseline hazard function. In the latter case, the baseline hazard is left unspecified and more complex techniques are available to approach that situation (Abrahantes, et al., 2007). The advantage of parametric method over the semiparametric method shows that having distribution may calculate the quantiles, simplicity and completeness are reasons for the popularity of parametric distributions (Kleinbaum & Klein, 2012).

This study was built on assessing the relationship between potential covariates and time to return of volunteer blood donors. Knowledge about donor return and the donor population enables establishments to adopt appropriate donor motivational strategies, pre-donation counseling, and developing guidelines and hemovigilance programme in countries.

1.2 Statement of the problem

Despite there exist a pressing need for blood transfusion in African countries related with numbers of clinical procedures, many countries in the region collect less than half of the blood needed to meet the transfusion requirements of their populations and predictably, access to blood remains a challenge(Abderrahman & Saleh, 2014).

Quarter million death in the world and 15% of child mortality in Africa is reported to due to obstetric bleeding and anemia which require blood transfusion(WHO, 2017). The high maternal and child mortality rates attributed to pregnancy related-complications and anemia, are evidence of the magnitude of the unmet need for blood transfusion in Africa (Jemberu et al.,

2016). As WHO reported in 2014, each year, about 25-40% of Ethiopian mothers who give birth die due to lack of enough blood from donors.

Annually 81 million units of blood are collected all over the world and 27 million of this is collected from low and middle-income countries, where 82% of the world's populations live. In SSA out of the estimated need of 18 million units of safe blood per year, only 15% was collected (Mirutse et al., 2014). The annual blood collection level rate in Ethiopia is very much below its demand(Ahmed, 2019). Concerning this, the joint press conference by the Ministry of Health and the National Blood Bank (2017) indicated that the blood deficit had reached critical stage over the period as less volume of donated-blood have been collected. The country requires 18,000 units of blood daily, yet the average daily amount collected is approximately 1,100 (UNODC, 2020). As a result, Blood Bank was not supplying the required volume of donated-blood to hospitals and health centers left their patients to suffer more (Ahmed, 2019). Thus, adequate and safe blood supply has remained a great challenge in Ethiopia (Eluke et al., 2017).

Getting adequate blood and blood products majorly depends on rate of blood donation (Nureye, 2019). The higher return rate of donations is, the shorter the time interval between donations. The shorter the time interval between donations is, the better supply of bloods that keeps up with the increasing demand of blood and blood products (Kheiri & Alibeigi, 2015b). So, to improve provision of safe and adequate blood and blood products, studying covariates of time to return (time interval between donations) of volunteer blood donors is found to be a predominant issue and it happened to be a reason to conduct this study.

Some studies have been conducted to identify covariates of return of blood donors' by using logistic regression(Arage et al., 2017; Dejen, 2015; Fantahun et al., 2017; Kheiri & Alibeigi, 2015a; Wiersum-osselton et al., 2014a) and Cox proportional hazard models(Agnihotri et al., 2012; Brian Custer, 2013; Custer et al., 2012; John et al., 2017; Lourençon, 2011). But Logistic regression does not account the censoring observations. Similarly, correct inference based on Cox's models needs identically and independently distributed samples. Cox proportional hazards model didn't take into account any extra heterogeneity present in the data. Ignoring this heterogeneity will produce biased parameter estimates and inconsistent standard errors in survival analysis (Wienke, 2010). Therefore, we have employed a shared frailty model to inves-

tigate the factors associated with time to return of voluntary blood donors taking into account the heterogeneity present in the data. Hence, this study focus on addressing the following research questions:

- What are the significant factors associated with time to return of the donors?
- What is the estimated median return time of the donors?
- Which baseline as well as frailty distribution best describes the time to return of the donors' data?

1.3 Objectives of the study

1.3.1 General of the study

The general objectives of this study is modeling time to return of volunteer blood donors in National blood bank, Addis Ababa, Ethiopian.

1.3.2 Specific objectives

The specific objective of the study are:

- To determine significant factors that affect the time to return for blood donation after once they donate blood.
- To estimate the median return time.
- To fit a best parametric shared frailty model for the donors dataset.

1.4 Significance of study

The finding of this study will

- helps to the national blood bank service to improve provision of safe and adequate blood and blood products by showing the gaps.
- helps to the health care system to have good number of donors for safe supply of blood that keeps up with the increasing demand of blood and blood products.

- provide an input to policy makers and human resource managers to design appropriate policy, programs and donor motivational strategies in countries.
- be useful to other researchers as a baseline while conducting further studies on similar problems.

2 LITERATURE REVIEW

In this section several related literatures to the present study had been examined, and the return behavior of voluntary blood donors is presented.

2.1 Overview of blood donation and donors return status

In spite of, extensive efforts and a number of blood donation programs being organized worldwide, availability of adequate and safe blood still remains short to meet the increased demand in developing countries including Ethiopian. Globally, around 92-million-unit blood donations are collected annually from blood donors (Jemberu et al., 2016). Despite of this fact, only 27 million are collected in low and middle-income countries, whereby 82% of the world's population lives (Mirutse et al., 2014).

Different countries have estimated their donor return rate so as to prepare an informed donor retaining strategy. Study conducted in northern Tanzania by using logistic regression analysis found that, the proportion of repeat donation were 63.9%, which is a good record (Mauka et al., 2015). A study done by on 24,684 first time voluntary blood donor's record using logistic regression analysis, the return rate was found to be 37.7% per two years, which is not a good record (Fantahun et al., 2017). Another cross-sectional study which was conducted in Madawalabu University using binary logistic regressions shows that, from the total respondents, only 18.4% of the donors donated blood before, whereas 81.6% were first time donors (had not donated blood in their life) (Darega et al., 2015).

There exists a significant reverse relationship between frequency of return to donation (becoming a regular repeat donor) and the time interval between donations. The shorter the time interval between donations is, the higher the frequency of donations. Thus, to improve provision of safe and adequate blood and blood products, studying the effect of different covariates on time to return of volunteer blood donors is essential (Kheiri & Alibeigi, 2015b).

2.2 Factors affecting the return of voluntary blood donors

This section mainly covers several related literatures to the factors that are associated with the time to return of voluntary blood donors.

Gender of the donors: - Several studies mention female to be more altruistic and for that mat-

ter become a first-time donor whereas other studies point out men to be committed and for this reason more likely to become regular donors(Bani & Giussani, 2010). Data from 118 countries on the gender profile of blood donors showed that, an overall 30% of the blood donations were given by female donors; with 18 of these countries reported less than 10% of donations by women(World Health Organization, 2016). However, women play a more substantial role by donating more than 45% of the total donations in some European countries and the United States(Bani & Giussani, 2010).

Cohort studies conducted in Netherlands (Wevers & Baaren, 2014) and US (Notari *et al*., 2009) suggested male donors to more likely return to donate after an index donation. Retrospective studies like the one done in Brazil reveled that, average number of donations among older men respondents were higher. It also concluded in studies done in Iran and Germany that more men tend to be repeat donors (Kasraian & Tavassoli, 2012). A study in Norway also identified women aged less than 45 years to be not as much of a regular donor but those who are above 45 years of age to have the same return pattern as men of the same age group (Darega *et al*., 2015). Opposing this finding a study done in china found among other factors being female to be linked with subsequent return for blood donation (Guo *et al*., 2012).

Even though gender influences donor returns in many literatures, study by (Tilburg, 2019) and (Lourençon, 2011) declared that the effect of gender on donor return is not statistically significant.

Donors' age:- Studies have shown that, age was significant factor of return status of the donors (Tilburg, 2019). Increasing age has a negative effect on time to return between donations: the younger the donor, the higher the chances of further donations, i.e. the time interval between two donations has been shorter (Kheiri & Alibeigi, 2015). This fact is also stated in a study done in China, subsequent return to blood donation were found to be higher among younger donors (Guo *et al*., 2012).

A cohort study conducted in America in which, first time donors were followed for 25 months, younger donors of age (16 & 17) found to be most likely to return (Notari *et al.*, 2009). In Germany also first-time donors aged 18 years were returned for successive donations comparing to older ones (Flegel & Wagner, 2000). A study done in Canadian metropolitan areas to investigate the correlation of Geographical variations with blood donor turnout, younger residents found to

be influential (Saberton *et al*., 2009).

In contrary to the presented literatures older age was found to positively associated with return of donors. A research done on Six REDS-II blood centers in the United States found that, increasing age was positively associated with subsequent visits. Likelihood of return with donors over 45 years were 2 times more likely to return than donors 16 to 22 years old. (Brian Custer, 2013). Among Netherlanders, the donors more likely to returned as they were older donors (Wevers & Baaren, 2014). The observed return rates dropped to the lowest among the age groups from 18 to 24 years and then gradually rose with age (Brian Custer, 2013).

Occupation: - Study conducted by Kheiri *et al*., (2015) shows that the occupation of donors affected the time intervals between two donations. According to this study, job classes of clerks, workers, the self-employed and university students have had a higher chance of donation compared with housekeepers, i.e., they have returned to donation sooner than housekeepers. A study was conducted in Madawalabu University students to assessed the practice of voluntary blood donation and associated factors. This study revealed that, even though university students are in the range of age of huge pool for blood donation, proportion of students have ever donated blood were low (Darega *et al*., 2015).

Body weight:- Studies have shown that body weight had an effect on return to donation(Kheiri & Alibeigi, *et al.,* 2015). The higher the weight, the more the chances of return to donation and the shorter the time interval between donations. Odds ratio for body weight was equal to 1·035, i.e., with increasing 1 kg of body weight, the ratio of chances for returning to not returning to donation increases by 3-5% (Kheiri & Alibeigi, 2015). A research done on Six REDS-II blood centers in the United States found that, repeat donors were heavier than first-time donors (Brian Custer, 2013).

Volume of blood donation: - Volume of blood donation has been identified as important risk factors for the return status of blood donors. A retrospective case-control study conducted to investigate the risk factors for adverse events after blood donation using logistic regression analysis found that, donation of 500Ml(donating larger volume of blood) was associated with increased likelihood of adverse event, which leads to decrement of the likelihood to be returned for subsequent donation(Wang et al., 2019).

In the contrary, Low estimated blood volume (3.5L) was associated with significantly lower

odds of return compared to donors with estimated blood volumes of 5 liters or more(Brian Custer, 2013). Study done by Dagmawit Fantahun on 24,684 first time voluntary blood donor's record using logistic regression analysis discovered that, donors who donated larger volume of blood were more likely to be returned for subsequent donation(Fantahun et al., 2017). As of this finding in china a study found that larger volume donors were 1.4 times more likely to return compare to low volume donors (*Guo et al*., 2012).

Experiencing donors reaction: - Although, blood donation has relatively low risk rate and donors undergo meticulous screening for any contraindications prior donation, some adverse events arise occasionally during or after the process (Raina, 2015). A study on Japanese Red Cross Tokyo Blood Centre done on randomly selected days in two consecutive years using logistic regression and chi-square test found that 10.07% of donors were experience vasovagal reaction (Takanashi *et al.*, 2012).

Negative experiences have a more severe impact on return than other factors. Previous studies had shown that donor return is severely impacted by the experience of adverse reactions during donation. A cohort study to assess the effect of complications on the donor return rate using logistic regression found that, complications reduce donor return for both first-time and more experienced donors. This study shows that among first-time donors, the return rate within 1 year was 82% following an uncomplicated first donation, but 55% and 61% following vasovagal reactions and needle-related complications respectively (Wiersum-osselton *et al.*, 2014b).

A study in US examined 1000 whole blood donors return pattern with a negative binomial regression analysis and found the return rate were 1.32 visits per year for those who didn't experience adverse reaction. Whereas, for those experience adverse reaction, return rate found to be 0.87 visits per year (Newman et al., 2006). In a study of 69,289 whole blood donors, Eder et al showed that 27% of donors who had any type of reaction returned to donate compared to 35% of those with no reaction. Only 18% of donors who experienced syncope made a repeat donation (Eder & Edward, 2009).

The severity of reaction is closely related to the likely-hood of returning to try to donate again. In the research of Custer et al., distinction is made between minor and major adverse reactions. The researchers found that for donors who experienced major adverse reactions such as loss of consciousness and needle-related injuries (i.e., nerve irritation and arterial puncture), it took

longer for them to return for a second donation. Whereas, minor reactions, namely nausea, dizziness, hyperventilation and bruising, have a small impact on donor return (Custer *et al*., 2016). No more than 30% of persons with major reactions returning. Similarly, for persons with minor reactions no more than 45% returned during follow-up. Whereas, for donors without reactions, 60% to 70% returned during follow-up period(Custer *et al*., 2012).

Major reactions will impact donor return regardless of previous donation history, but a positive donor experience increases the likelihood of return after a complication (Gemelli & Waller, 2017). If previous donation experience is positive, the likelihood of return after a complication is also higher (Brian Custer, 2013).

Although most of the researchers agreed that, adverse reactions will lead to lower donation return rate, a study done in Australia to asses' negative experiences and donor return using chi-square test found that, experiencing donor reactions may not universally deter donors from continuing to donate (Masser *et al*., 2016).

Donation experience: - For donors as whole pervious return status has an impact on continuous donation. A study done in Iranian Blood Transfusion Organization, using a cross-sectional follow-up design of 5-year duration shows a significant reverse relationship between frequency of return to donation and the time interval between donations. The shorter the time interval between donations is, the higher the frequency of donations (Kheiri & Alibeigi, 2015).

Odds of returning to donate increase in relation to previous donation frequency (Guo *et al*., 2012). In the study done in Netherlands those who had higher previous return rate seemed to return more often. Those who were repeat donors return more often (Wevers & Baaren, 2014). Similarly, a study conducted in Malaysia also found that, donors who practice donation has a higher aspiration to do it again (Zainie et al., 2013). Even donors with one previous donation had an odds of future return 3.7 times higher than that of first time donors, which indicates that once a donor is returned, the more likely to return in the future(Guo *et al*., 2012).

In US, frequency of first year donation was found to determine commitment of donors (Kheiri & Alibeigi, 2015). A study done in Brazil also found that, those who donated five or more times, 43.8% returned for the next donation at an interval of less than six months (de Araújo et al., 2010). This fact is also stated in a study conducted in Iran that the return rate first time donors to be directly correlated with the number of donations in the first year since their first

donation period (Kasraian & Tavassoli, 2012).

Blood group: -No significant difference was found in non-return rates regarding blood type/group (Lourençon, 2011).

Rh factor: - The Rh blood factor is one of the known human blood group systems. It is the second most important blood group system, after the ABO blood group system(Landsteiner & Wiener, 1937).

Rh factor has been found to be an important risk factors for the return status of blood donors. A higher percentage of donors with the Rh negative blood factor returned within six months(de Araújo et al., 2010). But, according to many scholars, Rh hadn't a significant impact on time to return of the donors (Fantahun *et al*., 2017; Kheiri & Alibeigi, 2015; Wevers & Baaren, 2014).

Donation site (fixed or mobile): - A well-publicized fixed blood center should attract a good number of donors if it is well-sited and is close to the center of a town with good transport links and parking. Donors can come whenever it suits them and there is generally little peer pressure placed on them to donate. On the other hand, mobile donor sessions can help to overcome many of the problems associated with inaccessible locations, particularly if they take place in the workplaces, educational institutions and communities in which people spend their time(WHO, 2010a).

A study conducted in the United States to assess adverse reactions and other factors that impact subsequent blood donation visits in Six REDS-II blood centers using logistic regression found that, collection sites other than fixed locations were affect a return of donors(Custer *et al*., 2012). In Tanzania, nearness of donation site was identified as a factor for repeat donation (Mauka *et al*., 2015). The return rate was higher at fixed donation site than mobile donation site (Fantahun *et al*., 2017).

In the study carried out in china, surprisingly found that donors who donated at mobile collection vehicles were higher to return than donors who donated at blood centers. Those who donated in a blood collection vehicle were 4 times more likely to return than those who donated at a blood center. This may be due to mobile collection vehicles in China usually operate at same "fixed" locations in the most crowded streets of a city, where they can be accessed by more people. They also have more flexible working hours, and work on weekends to allow

people to donate while shopping. Mobile(Guo *et al*., 2012).

2.3 2.2 Overview of Survival Analysis

Research on the statistical analysis of survival data from related individuals began in the mid-1970s with (Clayton, 1978) and (Holt and Prentice, 1974). Models have the advantage over other regression methods of being able to accommodate censored observations and time-varying covariates in analyses of event-history data. Clustered survival data also provide us an opportunity to estimate the association among subjects belonging to same cluster that persists even after controlling for observed covariates (Sastry, 1997).

2.3.1 Overview of frailty models

Frailty models (Clayton and Cusick, 1985) are increasingly popular for analyzing clustered survival data, where frailties often enter into the baseline hazard multiplicatively to model the correlation among observations within the same cluster (Li, 2000). Frailty models have been applied to the analysis of event- history data in a number of research areas, including the study of unemployment durations (McCall, 1994), consumer purchase behavior (Srinivasan, 1993), spells on welfare (Blank, 1989), migration (Lindstrom, 1996), fertility (Larsen and Vaupel, 1993), and marriage and divorce (Lillard, et al., 1995).

Wienke (2010) considered Halle Lung Cancer study data and applied two different parametric shared gamma frailty models with exponential and Weibull baseline hazards. The results showed that the exponential hazard function was not flexible enough and Weibull model shows a significantly better fit to the data with respect to the likelihood ratio test.

Duchateau and Janssen (2008) fit the inverse Gaussian frailty model with Weibull hazard to the udder quarter infection data. The better the model, the less unobserved heterogeneity there will be. It is argued that the less heterogeneity in the model, the more appropriate it is to interpret any observed duration dependence in substantive terms (Zorn, et al., 2000).

The gamma distribution has been widely applied as a mixture distribution (Hougaard, 2000, Duchateau and Janssen, 2008). The most common reason for using the gamma distribution is its mathematical convenience. This is due to the simplicity of the derivative of the Laplace transform, meaning that traditional maximum likelihood procedures can be used for parameter

estimation (Hougaard, 2000 and Locatelli et al., 2003). Its flexible shape is another reason given for selection of the gamma distribution as the frailty distribution (Sastry,1997; Manda, 2001). Although it may be the most commonly used frailty distribution for the mathematical reasons, (Hougaard, 1995) emphasized that there are no biological reasons for choosing the gamma distribution.

The inverse Gaussian (inverse normal) distribution was introduced as a frailty distribution alternative to gamma distribution by (Hougaard, 1984) and was used, for example, by (Manton, et al., 1986), (Klein, 2006), (Duchateau and Janssen, 2008). The choice of a family of frailty distributions should therefore be accompanied by an assessment of fit.

3 METHODOLOGY

3.1 Description of the study area and Data set

The study had been conducted in Ethiopian national blood bank service which is located in Addis Ababa, the capital city of Ethiopia. The National Blood Transfusion Services was established in 1969 by the Ethiopian Red Cross society. Since 2012, the Federal Ministry of Health of Ethiopia is overseeing the national blood bank entrusting with the responsibility of community mobilization & education on voluntary blood donation, managing of blood donors, collection, testing and transfusion of blood and blood products, promotion of appropriate clinical use of blood, research and capacity building in BTS(Gadisa Kebede,*et al.,* 2016). It is an autonomous body established with the aim of providing patients of Ethiopia with adequate, safe and effective blood products in an equitable and sustainable manner.

This data is secondary data recorded at the National Blood Bank, Addis Ababa, Ethiopia. The event for this study was donors return to donate blood after ones they donate blood. Baseline data has been extracted from existing databases of all volunteer whole blood donors who donated blood from September 06, 2012 to September 11, 2013 and whether he/she returned or not were followed until September 2015.

The time taken to return was the time interval between the first and the second donation, which had been rounded to the nearest month. Donors, who do not donate blood at least two times during the study time are considered as censored. Total number of donors considered in the study was 6019.

3.2 Study design

A retrospective cohort study has been conducted at National Blood Bank, Addis Ababa. In Ethiopia, any healthy person aged from 18-65 years and weighting not less than 45 kilograms may become a donor. A donor should weigh at least 45 kg to donate 350ml and 50 kg to donate 450ml.

This study considered parametric frailty models to investigate the relationship between different potential covariates and time to return of volunteer blood donors for clustered survival data with random right censoring. Hence, in this study Exponential, Weibull, log logistic and Lognormal

baseline hazard functions were used. On the other hand, among frailty distribution, Gamma and Inverse-Gaussian distributions were used. For comparison of different models, the AIC criteria was used.

3.3 Inclusion and exclusion criteria: -

In this study, all volunteer whole blood donors who donated blood from September 06, 2012 to September 11, 2013 and whether he/she secondly returned or not were followed until September 2015, had been included. But, blood donors with insufficient information had not been eligible.

3.4 Study Variables

3.4.1 Response variable

The outcome variable considered in this study was the time to return of volunteer blood donors after once donate blood (i.e. the time interval between the first and the second donation), which is measured in months.

3.4.2 Explanatory variables

The study considers the following explanatory variables that are may be factors of time to return of volunteer blood donors.

Table 3.1: Explanatory variables with codes and description

Variable	Description	Categories and codes
Age	Age of the donor (in years)	0 = 18-24, 1 = 25-44, and 2=45-65
Gender	Gender of the donor	0=Female and 1=Male
Weight	Weight of the donor (in kg.)	45-49, 50-59, 60-69, 70-79 and $\geq$ 80
Blood group	Blood group of the donors	0=A, 1= B, 2=AB, 3=O
Rh factor	Rh factor of the donors'	0=Negative, 1=Positive

Occupation	Occupation of the donor	0=Civil servant, 1= Private/ NGO worker,2= Student, 3 =Unemployed and 4= Others
Volume of blood donated	Volume of blood donated (in ml)	0 = 350, 1 = 450
Experiencing donors reaction	Occurrence of adverse reaction	0= no, 1= yes
Donation experience	Previous donation status of the donors'	0=First time donor, 1= Repeat donor
Donation site	Blood donation sites	0=Mobile,1= Fixed center

Note: Donation site had been considered as a clustering variable in all frailty models.

3.5 Method of data analysis

3.5.1 Survival analysis

Survival analysis is an important statistical technique used to describe and model time to- event data. According to Collett (2003), it is used to describe the analysis of data that corresponds to the time from a well-defined time origin until the occurrence of some particular event or end-point. Survival data are not amenable to standard statistical procedures used in data analysis mainly due to censoring. One of the features of survival data that renders standard methods inappropriate is that survival times were frequently censored. The use of survival analysis as opposed to the use of other statistical method is most important when some subjects are lost to follow up or when the period of observation is finite (certain patients may not experience the event of interest over the study period). In this latter case, one cannot have complete information for such individuals. These incomplete observations are referred to as being censored.

In reality, such event can occur due to the following reasons:

- A person does not experience the event before the study ends
- A person is lost to follow-up during the study period and
- A person withdraws from the study for unknown/ known reasons

There are three categories of censoring, (Klein and Moeschberger, 1992),

1. **Right censoring**: Survival time is said to be right censored when it is recorded from its beginning to a defined time before its end time.

2. **Left censoring:** Survival time is said to be left censored if an individual develops an event of interest prior to the beginning of the study.

3. **Interval censoring:** Survival time is said to be interval censored when it is only known that the event of interest occurs within an interval of time but the exact time of its occurrence is not known.

3.5.2 Descriptive methods for Survival data

An initial step in the analysis of a set of survival data is to present numerical or graphical summaries of the survival times in a particular group. In summarizing survival data, the two common functions applied are the survivor function and the hazard function (Hosmer and Lemeshow, 1999).

Survival Function: - The basic quantity employed to describe time-to-event phenomena is the survival function, the probability of an individual surviving or being event-free beyond time t (experiencing the event after time t). Moreover, the distribution of survival time is characterized by three functions: survivorship function, probability density function, and hazard function.

Let T be a random variable associated with the survival times, t be the realization of the random variable T and f (t) be the underlying probability density function of the survival time t. The cumulative hazard function H(t), which represents the probability that a subject selected at random will have a survival time (in this case, survival time to return) less than some stated value t, is given by:

$$F(t) = P(T \leq t) = \int_0^t f(u)du, t \geq 0 \tag{1}$$

The survival function is defined as the probability that the survival time is greater or equal to t.

$$S(t) = P(T > t) = 1 - F(t), t \geq 0 \tag{2}$$

$$S(t) = 1 - F_T(t), \tag{3}$$

Theoretically, as t ranges from 0 to infinity, the survivor function can be graphed as a smooth curve. This survival function gives the probability of surviving or being event free beyond time t. Because Survival functions (S(t)) is probability, it is characterized by:

1. They are non-increasing function.

2. At time, t = 0; S (t) = S (0) = 1. That is, at the start of the study no one has experienced the event yet, the probability of surviving past time 0 is one (1).

3. As time $t \to \infty; S(t) \to 0$.. That is, theoretically, if the study period increased without limit, eventually nobody would survive, so the survivor curve must eventually converge to zero.

Hazard Function: - The hazard function is a measure of the probability of failure during a very small interval, assuming that the individual has survived at the beginning of the interval. The hazard function describes the concept of the risk of an outcome (e.g., death, failure, hospitalization, return) in an interval after time t, conditional on the subject having survived to time t. It is the probability that an individual dies somewhere between t and $t + \triangle t$, divided by the probability that the individual survived beyond time t.

The hazard function h(t) can be formulated as:

$$h(t) = \lim_{\triangle t \to 0} \frac{P(t \leq T \leq t + \triangle t / T \geq t)}{\triangle t} = \frac{f(t)}{S(t)} \tag{4}$$

$$h(t) = \frac{f(t)}{s(t)} \tag{5}$$

The survival and cumulative hazard functions can be given in terms of the hazard function as:

$$h(t) = \int_0^t h(u)du \tag{6}$$

Using the above expressions the hazard function h(t) can also be given as:

$$h(t) = -\frac{\log S(t)}{dt} = \frac{dH(t)}{dt} \tag{7}$$

3.5.3 Estimation of survivorship function

In survival analysis, it is always a good idea to present numerical or graphical summaries of the survival times for the individuals. In general, survival data are conveniently summarized through estimates of the survival function and hazard function. This method is non-parametric or distribution-free, since they require no specific assumptions to be made about the underlying distribution of the survival times (Hosmer and Lemeshow, 1999).

Among the other estimators of the survivor function the Kaplan-Meier estimator is the most

common one. The Kaplan-Meier estimator of the survivorship function (Kaplan and Meier, 1958) also called product limit estimator, is the estimator used by most software packages. This estimator incorporates information from all of the observations available, both uncensored and censored, by considering survival to any point in time as a series of steps defined by the observed survival and censored times.

Suppose we have a sample of independent observations, their survival times denoted by $t_1, t_2, \ldots t_n$ and indicators of censoring denoting by $\delta_1, \delta_2 \ldots \delta_n$ where

$$\begin{cases} \delta_i = 1, & if\ the\ donorn\ returns\ to\ donate \\ \delta_i = 0, & otherwise \end{cases}$$

Thus, the survival data are denoted by (ti ,δ_i); i = 1,2,... n. The first step to obtain the Kaplan-Meier estimator of the survival function is to order the survival times as $t_1, t_2, ...t_n$. Assume that among the n observations $m \leq n$ event occurred at distinct m times. The main quantity of interest is the probability that an event would not occur by time t: $S(t) = P(T > t)$. Kaplan and Meier (1958) develop an estimator for the survival function.

$$\hat{S}_{KM}(t) = \prod_{t(i) \leq t} (\frac{ni - di}{ni})^{\delta i} = \prod_{t(i) \leq t} (1 - \frac{di}{ni})^{\delta i} \tag{8}$$

Where, *di* number of donors with event at *ti* and n_i is number of donors at risk before t_i.

3.5.4 Median survival time

Median survival time is the time beyond which 50% of the individuals in the population under study are expected to survive and is given by that value t(50) which is such that$\$_{t(50)} = 0.5$. Due to the fact that the non - parametric estimates $S(t)$ are step functions, it will not usually be possible to realize an estimated survival time that makes the survival function exactly equal to 0.5. Instead, the estimated median survival time, is defined to be the smallest observed survival time for which the value of the estimated survival function is less than 0.5

$$t(50) = min\{t : \hat{S}(t(j)) < 0.5\} \tag{9}$$

Where t(i) is the observed survival time for the i^{th} individual, i=1,2,...,n and t(j) is the j^{th} ordered event time, j=1,2,...,r+

3.6 Frailty models

The notion of frailty provides a convenient way to introduce random effect, association and unobserved heterogeneity into models for survival data. Frailty models are the survival data analog to regression models, which account for heterogeneity and random effects.

For example, in many cases it is impossible to measure all relevant covariates related to the disease of interest, sometimes because of economic reasons, sometimes the importance of some covariates is still unknown. The frailty approach is a statistical modeling concept which aims to account for heterogeneity, caused by unmeasured covariates. In statistical terms, a frailty model is a random effect model for time-to-event data, where the random effect (the frailty) has a multiplicative effect on the baseline hazard function. This random effect explains the dependence in the sense that had we known the frailty, the events would be independent. In other words, the life times are conditional independent, given the frailty.

A frailty is a latent multiplicative effect on the hazard function and is assumed to have unit mean and variance θ, which is estimated along with the other model parameters.

3.6.1 Shared Frailty Models

A shared frailty model is a random effects model where the frailties are common (or shared) among groups of individuals or spells and are randomly distributed across groups.

A natural extension of the univariate frailty model would be a multivariate survival model where individuals are allowed to share the same frailty value. Sharing a frailty value also generates dependence between those individuals who share frailties, whereas conditional on the frailty those individuals are independent. They are conditional independence model in which frailty is common to all subjects in a cluster. It is also known as a mixture model because the frailties in each cluster are assumed to be random. It assumes that, given the frailty, all event times in a cluster are independent.

Shared frailty model was introduced by Clayton (1978) without using the notion frailty and extensively studied in Hougaard (2000), Therneau and Grambsch (2000), Duchateau, *et al.* (2002, 2003), and Duchateau and Janssen,*et al.* (2004).

Multivariate frailty model is an extension of the univariate frailty model which allows the indi-

viduals in the same cluster to share the same frailty value. When frailty is shared, dependence between individuals who share frailties is generated.

Conditional on the random term, called the frailty denoted by w_i, the survival times in cluster $i(1 \leq i \leq n)$ are assumed to be independent, the proportional hazard frailty model assumes:

$$h_{ij}(t/X_{ij}, wi) = h_o(t)exp(\beta^t X_{ij} + wi) \tag{10}$$

where w_i the random term of all the subjects in cluster.

Assuming $Z_i = exp(wi)$, above equation can be rewritten as

$$h(t_{ij}/Z_i) = Z_i * h_0(t_{ij})exp(\beta^{'} X_{ij}) \tag{11}$$

In this model the variability has two different sources: the natural variability, included in the baseline hazard function and the other which is given by a frailty term that represents the unobserved variability from the covariates (Wienke, 2010). In this model it is assumed that given frailty term, the risk of each survival time follows a proportional hazard model, where the frailty term has a multiplicative effect on the baseline hazard function and also the covariates. For that reason, we have to specify the assumed distribution for baseline hazard function and frailty term.

The core assumption of a shared frailty model is that all individuals in cluster i share the same value of frailty $Z_{(i)}$ (i = 1,...,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 (common) frailty. This shared frailty is the cause of dependence between lifetimes within the clusters.

3.6.2 Frailty distributions

The frailty distributions that have been studied most belong to the power variance function family, a particular family of distributions introduced first by Tweedy (1984) and later independently studied by Hougaard (1986). The gamma, inverse Gaussian, positive stable, and compound Poisson distribution are all members of this family. Laplace transforms are adequate representations of the frailty distribution, as many characteristics of the frailty model can be expressed in terms of the Laplace transform. Among several number of frailty distributions, for this study our focus will be on Gamma and Inverse-Gaussian frailty distributions.

3.6.2.1 Gamma frailty distribution

Gamma frailty model belongs to the power variance function family (Hougaard, 1986) and can be expressed in terms of its Laplace transform from which properties such as mean and variance are easily derived (Duchateau and Janssen, 2008). From a computational and analytical point of view, it fits very well to failure data. It is widely used due to mathematical tractability (Wienke, 2011). Assuming a two-parameter gamma density with$\sigma > 0$ and $\gamma > 0$ as shape and scale parameters respectively, the density function is given by

$$f_z(Z) = \frac{\gamma^\sigma z_i^{\sigma-1} exp(-\gamma zi)}{\Gamma(\sigma)} \tag{12}$$

With $\sigma > 0$ and $\gamma > 0$ and where $\Gamma(.)$ is the Gamma function. The corresponding Laplace transformation is: $L(s) = \gamma^\sigma (s+\gamma)^{-\sigma}$

In gamma frailty models, restriction $\sigma = \gamma$ is used, which results in expectation of 1. The variance of the frailty variable is then 1. Assuming that the frailty term z_i is a gamma with E(Z) = 1 and Var(Z)= θ then, $\gamma = \frac{1}{\theta}$

Larger values of θ, indicate that there is a higher degree of heterogeneity among groups and strong association within groups (Abdulkarimova, 2013).This entails no loss of generality because the average level of frailty can always be absorbed into the baseline hazard. Then density of a gamma-distributed random variable frailty term z_i with parameter θ is:

$$f_z(Z) = \frac{z_i^{\frac{1}{\theta}-1} exp(\frac{-z_i}{\theta})}{\theta^{\frac{1}{\theta}}\Gamma(\frac{1}{\theta})} \tag{13}$$

Where: $\Gamma(.)$ is the gamma function; it corresponds to a Gamma distribution $Gam(\mu, \theta)$ with μ fixed to 1 for identifiability and its variance is θ.

$Z_{(i)} > 1$, indicates that individuals in group i are more frail, where as
$Z_{(i)} < 1$, indicates that individuals are less frail and have lower risk.

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

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

And the conditional hazard function is given by:

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

Where: S(t), and h(t) are the survival and the hazard functions of the baseline distributions. The variance θ of the frailty term represents the heterogeneity among clusters while the mean is constrained to 1 in order to make the average hazard identifiable (Duchateau *et al.,* 2002; Nguti, 2003; Glidden and Vittinghoff, 2004; Duchateau and Janssen, 2008). Larger variance indicates a stronger association within groups.

If hazard is constant but individuals are heterogeneous and has gamma frailty distribution, the population hazard then is given as:

$\frac{\lambda}{1+\lambda\theta t}$ where λ is average individual hazard and $\theta = \sigma^2$, is the variance of frailty.

3.6.2.2 Inverse-Gaussian frailty distribution

The inverse Gaussian (inverse normal) distribution was introduced as a frailty distribution alternative to the gamma distribution by (Hougaard, 1984). Similar to the gamma frailty model, simple closed-form expressions exist for the unconditional survival and hazard functions, this makes the model attractive. The probability density function of an inverse Gaussian distributed random variable Z with parameter $\theta > 0$ is given by:

$$f_z(Z) = \frac{1}{\sqrt{2\pi\theta z^3}} exp(\frac{-(z-1)^2}{2\theta z}) \tag{14}$$

It has a mean 1 and variance θ, and the Laplace transformation is given by:

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

For the inverse Gaussian frailty distribution conditional survival function is given by:

$S_\theta(t) = exp(\frac{1}{\theta}[1 - 2\theta ln(S(t))]^{\frac{1}{2}}), \theta > 0$

And the conditional hazard function is given by: $h_\theta(t) = h(t)[1 - \theta ln(S(t))]^{\frac{-1}{2}}, \theta > 0$, where S(t) and h(t) are the survival and the hazard functions of the baseline distributions.

Kendall's tau: is a global measure of the association between any two event times from the

same cluster in the multivariate case (Hougaard, 2000). It is an overall measure of dependence and independent of transformations on the time scale and the frailty model used.

For **Gamma distribution** Kendall's tau can be expressed in terms of the Laplace transform

$\tau = 4\int_0^\infty sL(s)L^2(s)ds - 1$

Using the Laplace transform of the gamma frailty, we obtain

$$\tau = \frac{\theta}{\theta + 2} \tag{15}$$

With $SE(\tau) = \frac{2SE(\theta)}{2+\theta)^2}$ Where $\tau\varepsilon(0,1)$

For **invers-Gaussian** distributed With multivariate data, frailty yields a Kendall's tau given by:-

$$\tau = \frac{1}{2} - \frac{1}{\theta} + \frac{2exp(\frac{2}{\theta})}{\theta^2}\int_{\frac{2}{\theta}}^{\infty}\frac{exp(-w)}{w} \tag{16}$$

Where $\tau\varepsilon(0,\frac{1}{2})$

3.6.3 Baseline hazard functions for parametric frailty models

As in the proportional hazards model, parametric or non-parametric forms of baseline hazard can be assumed in frailty models. If non-parametric form is assumed for h(t), then semi parametric proportional hazards model is considered and the estimates are usually obtained by using Expectation-Maximization (EM) algorithm.

Under the parametric approach, the baseline hazard function is defined as a parametric function and the vector of its parameters, say ϕ and it is estimated together with the regression coefficients and the frailty parameter(s). For simplicity, this study considers only the parametric forms of baseline hazard. Using parametric baseline hazards, not only makes the estimation easier, but it can also describe explicitly the effect of the frailty on hazard ratios over time. The survival time T is assumed to follow a distribution with density function f (t), then the survival function is given by:

$S(t) = P(T > t) = \int_0^t f(u)du$ and hazard function $h(t) = \frac{f(t)}{S(t)} = \frac{\frac{-d}{dt}S(t)}{S(t)}$

The relationship between the survival and the hazard function is given by

$S(t) = exp(-H(t))$= $exp(-\int_0^t h(u)du)$

The cumulative hazard function is given by $H(t) = \int_0^t h(u)du$ Specifying one of these four functions f(t),S(t),h(t) or H(t) can specifies the other three functions. The parameter λ is re-parameterized in terms of predictor variables and the regression parameters. Typically for parametric models, it is known that the shape parameter ρ is held fixed.

Table 3.2: Parametric distributions for the baseline hazards considered in this study

Distribution	f(t)	S(t)	h(t)	Parameter space
Exponential	$\lambda \exp(-\lambda t)$	$exp(-\lambda)$	λ	$\lambda > 0$
Weibull	$\lambda\rho t^{\rho-1} exp(-\lambda t^{\rho})$	$exp(-\lambda t^{\rho})$	$\lambda\rho t^{\rho-1}$	$\lambda, \rho > 0$
Log-normal	$\frac{1}{t\sigma\sqrt{2\pi}} exp(-\frac{(logx-\mu)^2}{2\sigma^2})$	$1-\Phi(\frac{logt}{\sigma})$	$\frac{\Phi(\frac{logt}{\sigma})}{1-\Phi(\frac{logt}{\sigma})\sigma t}$	$\mu \varepsilon R, t > 0, \sigma > 0$
Log-logistic	$\frac{\lambda\rho t^{\rho-1}}{(1+\lambda\rho t^{\rho})^2}$	$\frac{1}{(1+\lambda\rho t^{\rho})}$	$\frac{\lambda\rho t^{\rho-1}}{(1+\lambda\rho t^{\rho})}$	$\lambda \varepsilon R, \rho > 0$

3.7 Method of parameter estimation

Frailty models account for the clustering present in grouped event time data. The data for our study case are right-censored clustered survival data, that the observation for subject j, $J_j = 1, \ldots, n_i$ from cluster i, I = 1,...,s is the couple (y_{ij}, δ_{ij}), where, $y_{ij} = min(t_{ij}, c_{ij})$ is the minimum between the survival time, t_{ij} and the censoring time c_{ij} , and where, $\delta_{ij} = I(t_{ij} \leq c_{ij})$ is the event indicator, while $\delta_{ij} = 0$ for a censored observation. When covariate information are collected the observation will be $(y_{ij}, \delta_{ij}, X_{ij})$, 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 the assumption of right-censoring and of independence among the censoring time and the survival time of random variables, given the covariate information, the marginal log-likelihood of the observed data can be given by (Duchateau and Janssen, 2008):-

$$Lmarg(\varphi, \beta, \theta; Z, X) = \sum_{i=1}^{s} \{\delta_{ij}(log(h_o(y_{ij})) + X_{ij}^T) + log[(-1)^{(di)} L^{(di)}(\sum_{i=1}^{ni} H_o(y_{ij}) exp(X_{ij}^T))]\} \tag{17}$$

Where $di = \sum_{i=1}^{ni} \delta_{ij} = 1$ is the number of events in the i^{th} clusters. and $L^{(q)}(.)$ is the q^{th} derivative of the Laplace transform of the frailty distribution Z is defined as:-

$L(s) = E[exp(-Zs)] = \int_0^\infty exp(Zs) f(Z_i) dz_i, s > 0$ *and*
$L^{(q)}(s) = (-q) \int_0^\infty Z^q exp(-Zs) f(Z_i) dz_i, q \geq 0$

Where φ represents a vector of parameters of the baseline hazard function, the β vector of regression coefficients and θ the variance of the random effect. Estimates of φ,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 $q = max(d_1, ...d_s)$.

3.8 Comparison of models

One of the most commonly used model selection criteria is Akaike Information Criterion (AIC). A data-driven model selection method such as an adapted version of AIC (Akaike, 1974) is used to find the truncation point of the series. In some circumstances, it might be useful to easily obtain AIC value for a series of candidate models (Munda, 2012). In this study AIC criteria used to compare various candidates of parametric frailty models. For comparing models that are non-nested type, the Akaike's information criterion (AIC) which is defined as:
AIC= $-2Log(L) + 2(k + c + 1)$
Where k is the number of covariates, c the number of model specific distributional parameters. The preferred model is the one with the lowest values of the AIC.

3.9 Model diagnostics checking

After a model is fitted, the adequacy of the fitted model needs to be assessed. The methods that involved for using in this study are checking adequacy of the Parametric Baselines and the Cox-Snell Residuals (Cox DR and Snell EJ., 1968).

3.9.1 Checking the adequacy of parametric baselines

Graphical methods can be used for model diagnostics to see whether or not the distribution fits the observed data. Appropriateness of assumed distributions baseline hazard function is evaluated as follows:

- The suitability of model with the exponential baseline can graphically be evaluated by plotting: $(-log(\hat{S}(t)))$ versus t. Where:- $\hat{S}(t)$ is Kaplan-Meier survival estimate and the plot should be linear (Klein, 1992).

- For Weibull Baseline plot of $log(-log(\hat{S}(t)))$ versus $log(t)$ is used, since model with the Weibull baseline has a property that the $log(-log(\hat{S}(t))$ is linear with the log of time, where $\hat{S}(t) = exp(-\lambda t^{\rho})$. Hence, $log(-log(\hat{S}(t))) = log(\lambda) + \rho log(t)$ is linear with log of time. This property allows a graphical evaluation of the appropriateness of a Weibull model by plotting $log(–log(\hat{S}(t)))$ versus $log(t)$.

- For log-normal baseline plot of $\Phi^{-1}(1 - exp(-\hat{H}(t))) = \Phi^{-1}(1 - \hat{S}(t))$ versus $log(t)$ should be linear, if the log-normal distribution is appropriate.

- For Log-logistic baseline plot $log(\frac{1-\hat{S}(t)}{\hat{S}(t)})$ versus $log(t)$ This should be linear with slope ρ. The log-failure odd versus log time of the log-logistic model is linear. Where log failure odds can be written as: $log(\frac{1-\hat{S}(t)}{\hat{S}(t)})$= $log(\lambda\rho t^{\rho}) = log(\lambda) + \rho log(t)$ Where $\hat{S}(t)$ is Kaplan-Meier survival estimate (Christoph, *et al.,* 2011).

3.9.2 The Cox Snell residuals

The Cox-Snell residuals method can be applied to any parametric model and the residual plots can be used to check the goodness of fit of the model. 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 and Moeschberger, 2003). The first such residual is the Cox–Snell residual that provides a check of the overall fit of the model. The Cox–Snell residual,r_j is defined by $r_j = \hat{H}(T_j/X_j)$, where $\hat{H}$ is the cumulative hazard function of the fitted model. If the model fits the data then $r'_j s$ should have a standard $(\lambda = 1)$ exponential distribution, so that a hazard plot of r_j versus the Nelson–Aalen estimator of the cumulative hazard of the $r'_j s$ should be a straight line with slope 1.

3.9.3 The Quantile - Quantile plot

An initial method for assessing the potential for an AFT model is to produce a quantile-quantile plot. A quantile-quantile or q-q plot is made to check if the accelerated failure time (AFT) model

provides an adequate fit to the data. The plot is based on the fact that, for the accelerated failure-time model, $S_1(t) = S_0(\phi t)$, Where S_0 and S_1 are the survival functions in the two groups and ϕ is the acceleration factor. Let t_{op} and t_{1p} be the p^{th} percentiles of groups 0 and 1 respectively that is $t_{kp} = s_k^{-1}(1-p), k = 0,1$. Using the relation $S_1(t) = S_0(\phi t)$ we must have $S_o(t_{op}) = 1 - p = S_1(t_{1p}) = S_0(\phi t_{1p})$ for all t. If the accelerated failure time model holds, $t_{op} = \phi t_{1p}$.

To check this assumption we compute the Kaplan–Meier estimators of the two groups and estimate the percentiles t_{1p}, t_{0p}, for various values of p. If we plot the estimated percentile in group 0 versus the estimated percentile in group 1 (i.e., plot the points t_{1p}, t_{0p} for various values of p), the graph should be a straight line through the origin, if the accelerated failure time model holds. If the curve is linear, a crude estimate of the acceleration factor ϕ is given by the slope of the line (Klein, 1992).

4 RESULT AND DISCUSSION

4.1 Descriptive analysis

In this study, a total of 6,019 voluntary blood donors who donated blood during a period of 06 Sept 2012 to 11 Sept 2013 were considered. The response was time to return of volunteer blood donors, which is measured in months. Of all 6,019 voluntary blood donors, only 2811 (46.7%) were experienced the event (returned to donate blood again during the study period) and 3208 (53.3%) were censored (never returned to donate blood) (Table 4.1).

From the total number of voluntary blood donors, 2172(36.1%) were females and 3847(63.9%) were males. Among these, 39.6% of females and 50.7% of males were returned to donate blood.

The age distribution of donors that donates blood (at least one times) was the following: 2693(44.7%) were 18-24 years old, 3104(51.6%) were 25-44 years old and about 222(3.7%) were 45-65 years old. Of this, the number of blood donors who gives blood two or more times from age group 18-24, 25-44 and 45-65 were 1447(53.7%), 1247(40.2%), and 117(52.7%) respectively.

From the totality, donors of 258(4.3%), 1879(31.2%), 1807(30.0%), 1248(20.7%), and 827(13.7%) had weights 45-49,50-59, 60-69, 70-79 and $\geq$ 8080 respectively. Concerning return status, as weight increases return status also linearly increases from 28.3%, to 64.2%.

Number of blood donors with blood group A, B, AB. and O were 1713(28.5%), 1428(23.7%), 373(6.2%), and 2505(41.6%) respectively. From these, persons with blood group A had best return status (49.1%) followed by those had blood O (46.5%). Donors with Rh positive were 5580(92.7%), from which 2607(46.7%) were returned; however, donors with Rh negative were 439(7.3%), of which 204(46.5%) were returned.

Number of donors with occupation of Civil servant, private/NGO worker, student, unemployed and others were 177(4.6%), 3674(61.0%), 1982(32.9%), 36(0.6%), and 50(0.8%) respectively and from this 177(63.9%), 1893(51.5%), 684(34.5%), 31(86.1%), and 26(52.0%) were returned to donate blood respectively.

Donors who donate 350 (ml) were 4296(71.4%), from which 1760(41.0%) were returned; however, who donates 450 (ml) were 1723(28.6%), of which 1051(61.0%) were returned. although most of the donors donate 350 ml, their return status is less than that of who donates 450ml.

Table 4.1: Descriptive summaries of donors' data set

Variable	Categories	Return Status			Median
		Censored(%)	Event(%)	Total(%)	Time
Gender of donors	Female	1312(60.4)	860(39.6)	2172(36.1)	36
	Male	1896(49.3)	1951(50.7)	3847(63.9)	24
Age (in years)	18-24	1246(46.3)	1447(53.7)	2693(44.7)	21
	25-44	1857(59.8)	1247(40.2)	3104(51.6)	36
	45-65	105(47.3)	117(52.7)	222(3.7)	23
Weight (in Kg)	45-49	185(71.7)	73(28.3)	258(4.3)	36
	50-59	1204(64.1)	675(35.9)	1879(31.2)	36
	60-69	959(53.1)	848(46.9)	1807(30.0)	26
	70-79	564(45.2)	684(54.8)	1248(20.7)	20
	≥ 80	296(35.8)	531(64.2)	827(13.7)	12
Blood group	A	872(50.9)	841(49.1)	1713(28.5)	25
	B	781(54.7)	647(45.3)	1428(23.7)	28
	AB	215(57.6)	158(42.4)	373(6.2)	36
	O	1340(53.5)	1165(46.5)	2505(41.6)	26
Rh Factor	Negative	235(53.5)	204(46.5)	439(7.3)	27
	Positive	2973(53.3)	2607(46.7)	5580(92.7)	26
Occupation	Civil servant	100(36.1)	177(63.9)	277(4.6)	12
	Private/NGO worker	1781(48.5)	1893(51.5)	3674(61.0)	24
	Student	1298(65.5)	684(34.5)	1982(32.9)	36
	Unemployed	5(13.9)	31(86.1)	36(0.6)	7
	Others**	24(48.0)	26(52.0)	50(0.8)	25
Volume of blood donated (ml)	350	2536(59.0)	1760(41.0)	4296(71.4)	36
	450	672(39.0)	1051(61.0)	1723(28.6)	15
Donation experience	1st time donor	913(51.9)	846(48.1)	1759(29.2)	28
	Repeat donor	2295(53.9)	1965(46.1)	4260(70.8)	25
Donation site	Mobile	1994(68.8)	904(31.2)	2898(48.1)	36
	Fixed center	1214(38.9)	1907(61.1)	3121(51.9)	14
Donors reaction	No donor reaction	3086(52.9)	2753(47.1)	5839(97.0)	26
	Donor reaction	122(67.8)	58(32.2)	180(3.0)	28

Donors who donate in mobile site were 2898(48.1%), from which 904(31.2%) were returned; however, donors who donates in fixed site were 3121(51.9%), of which 1907(61.1%) were returned.

From the totality of blood donors who donated blood during a period of 06 Sept 2012 to 11 Sept 2013, 180 (3%) were exposed (had donor reactions during blood donation). The rest had no donors' reaction. The return status of those who faces adverse reaction were 58(32.2%) much less than that of reaction free donors 2753(47.1%).

4.1.1 Survival by experiencing adverse reaction

The survival time to return for donors who experience adverse reaction (red line) is greater than those donors who didn't experience adverse reaction (green line). This indicates that survival time to return of donors who didn't experience adverse reaction is better(less) than that of donors who experience adverse reaction (Figure 4.1). The effect of this group was significantly different from the reference groups (Table 4.3). KM curve of the remaining significant variables was presented in the appendix.

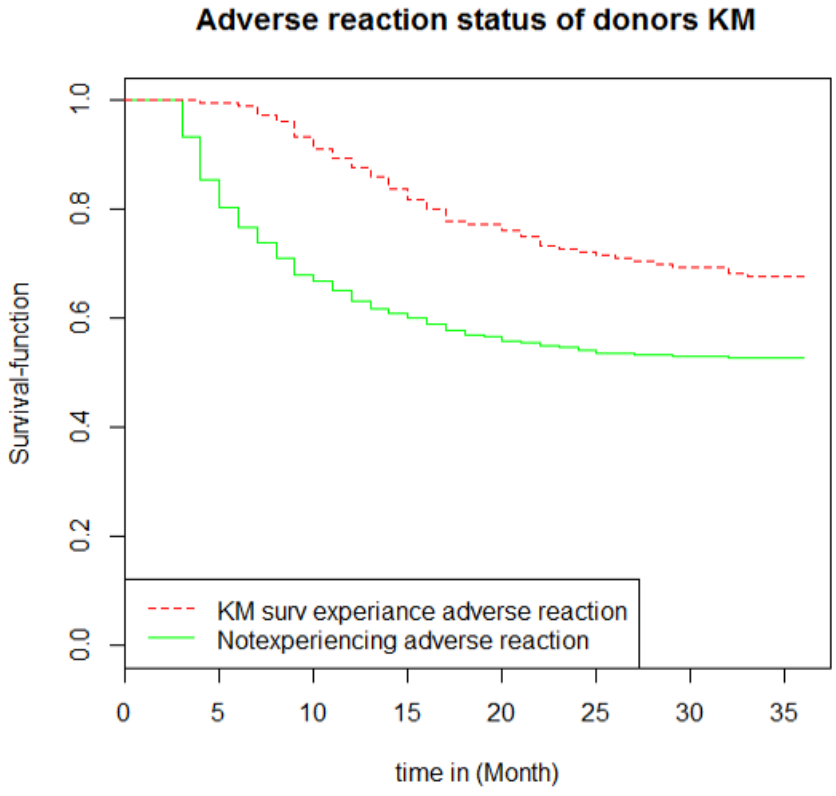

Figure 4.1: The survival functions of adverse reaction status of donors

4.2 Univariable Analysis

In the univariable analysis, covariates with p-value less than or equal to 25% were considered for multivariable analysis. Then, the multivariable models were fitted including all the potential covariates that were significant at 25% level of significance at the univariate level (see Annex 1).

From the univariable analysis we observed that the covariate gender of the donor, age of the donor, weight (in Kg), blood group, occupation of the donor, volume of blood donated (ml), donation experience and donors' reaction were significant almost in the entire models used. This indicates that they are significant prognostic factor for the time to return of blood donors.

However, Rh factor of the donor was not a significant factor for the time to return of blood donors according to all the candidate models (i.e., Exponential-Gamma, Weibull-Gamma, Lognormal-Gamma, Log logistic–Gamma, Exponential-Inverse Gaussian, Weibull- Inverse Gaussian, Lognormal-Inverse Gaussian and Log logistic– Inverse Gaussian). Therefore, based on this result, it is better to ignore these covariates and shall do our multivariable analysis using the significant factors. Hence, the effects of gender of the donor, age of the donor, weight (in Kg), blood group, occupation of the donor, volume of blood donated (ml), donation experience and donors' reaction on time to return of blood donors shall better be interpreted using the multivariable analysis.

4.3 Multivariable analysis and model comparison

For time to return of volunteer blood donors, the multivariable survival models of the Exponential, Weibull, Log-logistic and Lognormal for the baseline hazard function; and the Gamma and the Inverse Gaussian frailty distributions were fitted again by assuming all the significant covariates in the univariable analysis at 25% level of significance. The output of the Lognormal-Gamma multivariable frailty model is presented in Table 4.3; and the output of the other multivariable frailty models were similarly drawn (see Annex 2).

The variance of the random effect or frailty θ is significant for all baseline frailty models at 5% level of significance. The Kendall's tau τ is used to measure the dependence within the clusters (donation sites) and it is higher for the higher variance of random effect θ values. Accordingly, the dependence within the clusters for the log-normal-gamma frailty model ($\tau = 0.523$) is the maximum. In this study, we used AIC criteria to compare various candidate models of

parametric shared frailty models. A model having minimum AIC value considered as a better fit. Accordingly, the AIC value of the Lognormal- Gamma parametric shared frailty model (AIC=25626.23), was the minimum from all the other AIC values of the alternative models (Table 4.2). Hence, Lognormal- Gamma shared frailty model selected as best fit the donors' data set.

Table 4.2: AIC values of the models used in the study

Model		
Baseline hazard function	**Frailty distribution**	**AIC**
Exponential	Gamma	26260.97
	Inverse-Gaussian	26261.01
Weibull	Gamma	26138.96
	Inverse-Gaussian	26138.93
Lognormal	Gamma	**25626.23**
	Inverse-Gaussian	25630.24
Log logistic	Gamma	25822.55
	Inverse-Gaussian	25822.53

Analysis based on Lognormal- Gamma frailty model shows that the gender of the donor, age of the donors', weight, occupation, donation experience, and experiencing donors reaction were significant at 5% level of significance (Table 5). This indicates that they were the contributing factor for the return status of blood donors. However, according to this model the volume of blood donated and blood group of the donors had no significant effect on the return status of blood donors.

When the effect of other factor kept fixed, male donors had significantly different return time than the reference groups (female donors) with acceleration factor of 0.481. The respective 95% confidence interval was [0.435 0.731]. Therefore, male donors had shorter return time for the next donation by a factor of 0.481 than the reference group (female donors).

The result of this study suggested that age of donors had significant effect on the return status

of blood donors. Donors with age (45-65) years had significantly different return time than the reference group (18-24) years with acceleration factor (ϕ = 1.283). Therefore, donors with age (45-65) years had prolonged time to return for the second donation by a factor of 1.283 than reference group (donors with age (18-24)).

Given the effect of other factor kept fixed, large weight donors (60-69,70-79 and $\geq$ 80) had significantly different return time than the reference groups (45-49) with acceleration factor of 0.697, 0.663 and 0.56 respectively. Their respective 95% confidence interval was [0.541, 0.898], [0.507, 0.866] and [0.423, 0.740]. Therefore, donors with weight (in Kg) 60-69,70-79 and $\geq$ 80 had shorter return time for the next donation by a factor of 0.697, 0.663 and 0.56 respectively than the reference group (40-49). From this we understand that as weight increases, the time required for next donation decreases.

According to the result, the type of occupation that the donors had known to be significant covariate. Donors with occupation of student had significantly different return time for the next donation than the reference group (civil servants) with acceleration factor of 1.471 and 95% confidence interval [1.173,1.845]. This result suggested that students had prolonged time to return for the next donation than the reference group (civil servants).

Donation experience of blood donors had significant effect on the return status of blood donors. According to the result, repeat blood donors had significantly different return time than the 1st time donor with acceleration factor of 0.405. This indicates repeat blood donors had shorter return time than first time donors.

Provided the effect of other factor kept fixed, donors who experience donors' adverse reaction had significantly different return time for next donation than the reference group (who didn't experience donors' adverse reaction) with acceleration factor of 5.215 and 95% confidence interval [3.909, 6.957]. This result suggested that donors who experience donors' adverse reaction had extremely large (5.215 times) time to return for the next donation than the reference group (who didn't experience donors' adverse reaction).

Table 4.3: Lognormal- Gamma multivariable analysis

Covariates	Category	Coef	S.E	ϕ	95% CI	p-value
Gender of	Female	Ref		1		
donors	Male	-0.731	0.0519	0.481	[0.435 0.731]	0.016
	18-24	Ref		1		
Age	25-44	0.0931	0.0559	1.098	[0.984 1.225]	0.096
(in years)	45-65	0.2493	0.1252	1.283	[1.004 1.640]	0.047*
	40-49	Ref		1		
	50-59	-0.1263	0.1262	0.881	[0.688 1.129]	0.32
Weight	60-69	-0.3607	0.1294	0.697	[0.541 0.898]	0.005*
(in Kg)	70-79	-0.4111	0.1364	0.663	[0.507 0.866]	0.002*
	≥ 80	-0.5801	0.1427	0.56	[0.423 0.740]	$< .001***$
	A	Ref		1		
Blood	B	0.0999	0.0623	1.105	[0.978 1.249]	0.11
group	AB	0.1489	0.1006	1.161	[0.953 1.414]	0.14
	O	0.0523	0.0542	1.054	[0.947 1.172]	0.33
	Civil servant			1		
	Private/NGO worker	0.0799	0.1049	1.083	[0.882 1.330]	0.45
Occupation	Student	0.3857	0.1156	1.471	[1.173 1.845]	$< .001***$
	Unemployed	-0.1736	0.2815	0.841	[0.484 1.460]	0.54
	Others**	0.3319	0.2602	1.394	[0.837 2.321]	0.2
Volume of blood	350			1		
donated (ml)	450	0.0498	0.0567	1.051	[0.940, 1.175]	0.38
Donation	1st time donor			1		
experience	Repeat donor	-0.9027	0.0621	0.405	[0.359 0.458]	$< .001***$
Donors	No donor reaction			1		
reaction	Donor reaction	1.6516	0.1470	5.215	[3.909 6.957]	$< .001***$
θ=1.1 , λ=1.54, τ= 0.523, AIC = 25626.23						

*Source: National Blood Bank, Addis Ababa, Ethiopia; donates blood from September 06, 2012 to September 11, 2013 and whether he/she returned or not would followed until September 2015, *Statistically significant at 5% level, ϕ=Acceleration factor, θ=Variance of the random effect, τ= Kendall's tau, CI=confidence interval for ϕ, Ref=Reference, λ=Scale*

4.4 Model diagnostics

4.4.1 Diagnostic plots of parametric baselines

The final step in the model assessment is to see the overall goodness of fit. Therefore, it is desirable to determine whether a fitted parametric model adequately describes the data or not.

To check the adequacy of our baseline hazard: Exponential is plotted by $log(\hat{S}(t))$ versus t; Weibull is plotted by $log(–log(\hat{S}(t)))$ with the logarithm of time of the study; the log-logistic is plotted by log odds of failure or $log(\frac{1-\hat{S}(t)}{\hat{S}(t)})$ with the logarithm of time and the log-normal is plotted by the qnorm(1-survival) or $\Phi^{-1}[1-\hat{S}(t)]$ with the logarithm of time (Figure 4.2). The plot of Lognormal is more linear than the other plots. The pattern suggests that the Lognormal hazard function is appropriate in the model.Plot of the remaining was presented in the appendix.

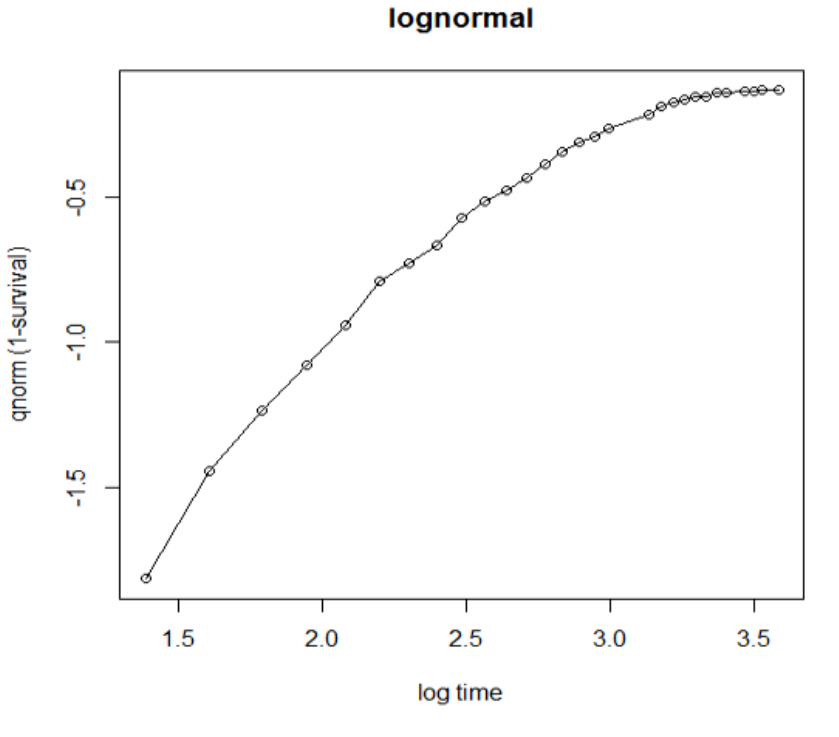

Figure 4.2: Graphical evaluation of the Log-normal assumptions

4.4.2 The Cox Snell residual plots

The Cox-Snell residuals are one way to investigate how well the model fits the data. In this case we used the Cox-Snell residuals to check the overall goodness of fit for different parametric models. The Cox- Snell residuals obtained from fitting the Lognormal model to our data via

maximum likelihood estimation. By comparing with Exponential, Weibull and Log Logistic, plot of the Cox-Snell residuals of the Lognormal models were nearest to the line through the origin, indicating that this model describes the donors' dataset well.

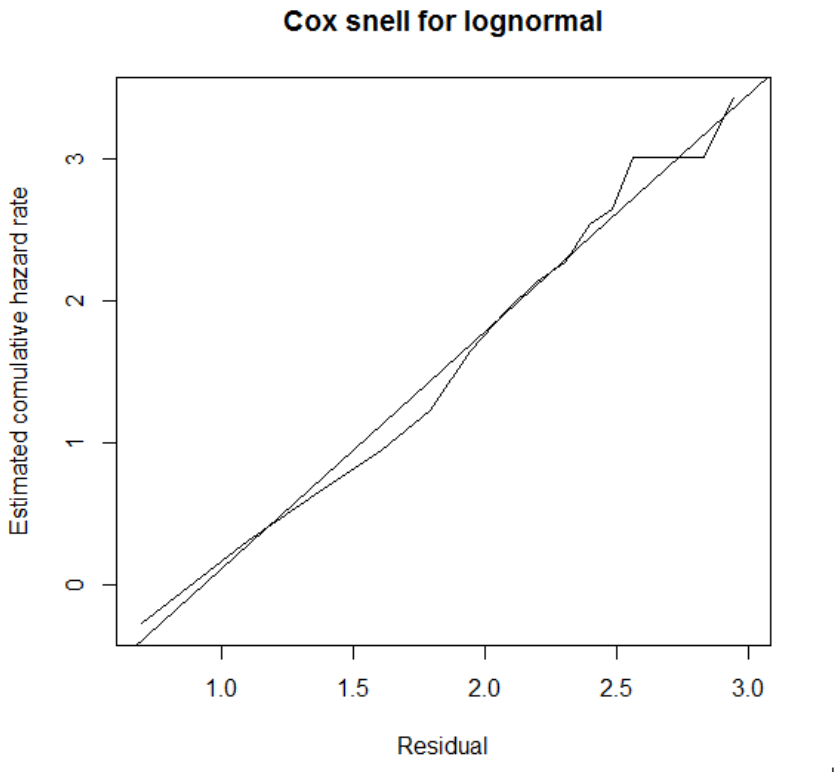

Figure 4.3: Cox-Snell residuals obtained by fitting Log-normal to the donors' dataset.

4.4.3 Adequacy of accelerated failure time

A quantile-quantile or q-q plot is made to check if the accelerated failure time provided an adequate fit to the data using two different groups of population. We shall graphically check the adequacy of the accelerated failure-time model by comparing the significantly different reaction status (experience adverse reaction, not experiencing adverse reaction), donation experience(1st time or repeat donors'), gender of the donors' (Figure 4.3). The q-q plot for remaining significantly different groups of variables also has drawn (See Annex 4). The figures appear to be approximately linear for all covariates.

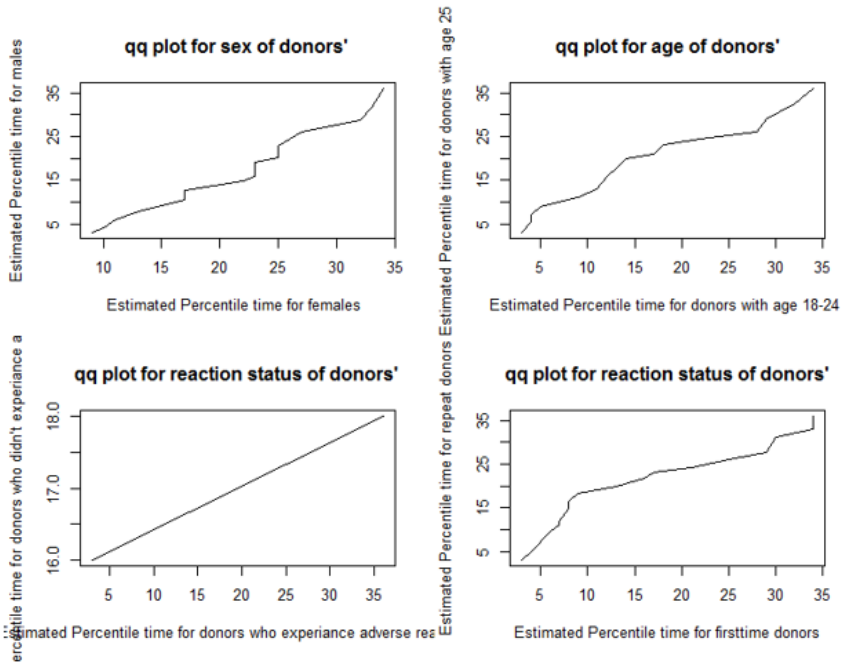

Figure 4.4: Q-Q plots to check the adequacy of the accelerated failure time model

4.5 Discussion

The main goal of this study was modeling time to return of volunteer blood donor's: in National Blood Bank, Addis Ababa, Ethiopian. From the total of 6,019 voluntary blood donors, only 2811 (46.7%) were returned to donate blood two or more times during the study period. This result is small as compared with findings of Kasraian & Tavassoli, et al., (2012), return rate to be 51.7% during 3 years after the first donation. Similarly, it is lower than study findings in Netherland, return rate within a time interval of nearly one year to be 82% without adverse reaction(Wiersum-osselton et al., 2014a). On the other hand, it is much better than study findings by (Fantahun et al., 2017) found that the return rate among first time blood donors for subsequent donation to be 37.7% per two years.

Lognormal- Gamma shared frailty model having minimum AIC value selected as best fit the donors' data set. Lognormal really shines for skewed distributions, large variances (i.e, data with a large standard deviation), and all-positive values. Additionally, if we were to take the natural log of each random variable and its result is a normal distribution, then the Lognormal is the best fit.

The findings of this study revealed that being male donor, donation experience (repeat donor)

and increasing weight (in Kg) significantly shorten/minimize the time-to-return of blood donors, while being in the age group (45-65), being a student and experiencing donors' reaction prolongs the time to- return of volunteer blood.

Results of our findings showed that gender of donors significantly influenced time to return of volunteer blood donors. Donors with gender male had shorter return time for the next donation by a factor of 0.481 than the reference group (female donors). The result was similar with studies conducted by (Wevers & Baaren, 2014) and (Notari *et al.*, 2009). They suggested that, male donors were more likely return to donate after an index donation.This result also similar with studies done in Iran and Germany that more men tend to be repeat donors than women's (Kasraian & Tavassoli, 2012).

Opposing this finding, a study done in china found that being a female donor to be linked with subsequent return for blood donation (Guo *et al.*, 2012). Even though gender influences donor returns in many literatures including our result, study by (Tilburg, 2019) and (Lourençon, 2011) declared that the effect of gender on donor return is not statistically significant.

Our findings showed that age (in years) of donors significantly influenced time to return of volunteer blood donors. Study by (Tilburg, 2019) also supports this fact. Acceleration factor of $theta$ =1.283, indicates that, donors of age (45-65) had prolonged time to return of volunteer blood donors as compared to donors of age (18-24). This study is also consistent with (Kheiri & Alibeigi, 2015; Wevers & Baaren, 2014), age has a negative effect on time intervals between donations: the younger the donor, the higher the chances of further donations, i.e. the time interval between two donations has been shorter. This fact is also agrees with a study done in China, subsequent return to blood donation were found to be higher among younger donors (Guo *et al.*, 2012). This study is also similar with study done by (Flegel & Wagner, 2000), which was stated as, donors aged 18 years were returned for successive donations comparing to older ones.

In contrary to the presented literatures (Brian Custer, 2013) and (Wevers & Baaren, 2014) found that, increasing age was positively associated with subsequent visits. The observed return rates dropped to the lowest among the age groups from 18-24 years and then gradually rose with age (Brian Custer, 2013).

Results of this study had shown that body weight had an effect on return to donation. In this

study, as the weight of the donors' increases, their time to return for donation were decreases. This finding agrees with study conducted by (Kheiri & Alibeigi, 2015). The higher the weight, the more the chances of return to donation and the shorter the time interval between donations. This study is also consistent with a research done on Six REDS-II blood centers in the United States, repeat donors were heavier than first-time donors (Brian Custer, 2013).

The results of this study suggested that experiencing donor reaction severely impacted donors return. For donors experiencing donor reaction, the acceleration factor was much greater than 1(ϕ =5.215). This shows, donors who experience adverse reaction took longer time to donate blood secondly as compared to the reference group (didn't experience adverse reaction). This finding is consistent with a study done by (Custer et al., 2012; Eder & Edward, 2014; Wiersum-osselton *et al*., 2014a). They reported from their finding that, complications reduce donor return or took longer time to return for donation. In a study of 69,289 whole blood donors, Eder et al showed that 27% of donors who had any type of reaction returned to donate compared to 35% of those with no reaction. Conversely to our result, a study done in Australia found that, experiencing donor reactions may not universally deter donors from continuing to donate (Masser *et al*., 2016).

The results of this study suggested that donation experience was significantly affect donors return status. For repeat donors, the acceleration factor was less than 1 (ϕ=0.405). This shows returning time for repeat donors were shorter than the reference group (1st time donors). The finding of the current study agrees with study conducted in Malaysia found that, donors who practice donation has a higher aspiration to do it again (Zainie *et al*., 2013). This study is also consistent with study done in Netherlands, those who had higher previous return rate seemed to return more often (Wevers & Baaren, 2014). Similarly the finding agrees with a study done by (Guo *et al*., 2012), odds of returning to donate increase in relation to previous donation frequency. Even donors with one previous donation had an odds of future return 3.7 times that of first time donors (Guo *et al*., 2012). A study done in Brazil also revealed that, those who donated five or more times, 43.8% returned for the next donation at an interval of less than six months (de Araújo *et al*., 2010).

Occupation of the donor was identified as a significant factor for time to return of volunteer blood donors. For those whose occupation was a student, the acceleration factor was greater

than 1(ϕ=1.471). This shows returning time for donors who were a student were longer (less likely to return) than the reference group (civil servant). This finding was consistent with a study conducted in Madawalabu University Students to assessed the practice of voluntary blood donation and associated factors. The study revealed that, even though the university students are in the range of age of huge pool for blood donation, proportion of students had ever donated blood were small (Darega *et al*., 2015). On the contrary to our finding, study conducted by Kheiri *et al*., (2015) found that, job classes of university students have had a higher chance of donation compared with others like housekeepers.

This study also showed that there was a clustering (frailty) effect on modeling time to return of volunteer blood donors which is due to the heterogeneity with in donation sites where they donate blood.

Assuming donors who donate blood either in fixed or mobile site share similar risk factors, indicating that considering the clustering effect in modeling the hazard function was important. The heterogeneity in the donation sites was significant and estimated to be θ=1.1, and the dependence within clusters is about τ = 0.523(52.3%). Those values were the maximum among the variance of the random effects and the Kendall's tau of all the candidate models. This result consolidates the idea that larger values of *theta*, indicates that there is a higher degree of heterogeneity among groups and strong association within groups (Abdulkarimova, 2013).

5 Conclusion and Recommendation

5.1 Conclusion

The following are the major concluding remarks of this study mainly based on the results obtained in the analysis of this thesis work:

The Lognormal-gamma frailty model is the model that best described time to return of the donors' dataset. Analysis based on Lognormal- Gamma frailty model shows that being male donor, donation experience (repeat donor) and increasing weight (in Kg) significantly shorten/minimize the time-to-return of blood donors, while being in the age group (45-65), being a student and experiencing donors' reaction prolongs the time to- return of volunteer blood donors.

According to the study, the median return time of the volunteer blood donors was high (median =26). There is a frailty (clustering) effect on the time-to-return of donors' that arises due to heterogeneity in between the donation sites.

5.2 Recommendation

Based on the study finding the following recommendations were forwarded: -

- In order to increase the return rate, Ethiopian national blood bank service should increase number of fixed donation sites.
- Since occurrences of donors' adverse reaction severely decreases donors return rate, special attention must be paid to the impact of it and the concerned body as well researchers should do further studies to identify factors that leads to adverse reaction to minimize its occurrence.
- Since the likely hood of return for donors with small weight, who are female donors and for those whose occupations were student is much less, policy makers and human resource managers should make interventions, and design appropriate policy, programs and donor motivational strategies to improve their return time.

5.3 Limitations of the study

Our analysis suffers from some limitations. Donor without reaction during initial donation might have developed delayed reactions (after leaving the donation site) and took longer time to return or failed to return at all, but there wasn't a system developed to record post donation reactions. Another limitation was inability to classify reaction into different levels (minor and major) to see the effect of each level separately to give more attention to the level that most affected return rate. In addition, categorization of variables like weight and age are alit bit deviates from others researchers categorized them. This study also did not include variables like educational status, knowledge and attitude of donors. These factors have been found to be significantly affect time to return of donors in different studies. They were, however, not included in donors' data set.

Reference

Abderrahman, B. H., & Saleh, M. Y. N. (2014). Investigating Knowledge and Attitudes of Blood Donors and Barriers Concerning Blood Donation in Jordan. Procedia - Social and Behavioral Sciences, 116(2000), 2146–2154. https://doi.org/10.1016/j.sbspro.2014.01.535

Agnihotri, N., Marwaha, N., & Sharma, R. R. (2012). Analysis of adverse events and predisposing factors in voluntary and replacement whole blood donors: A study from north India. Asian Journal of Transfusion Science, 6(2), 155–160. https://doi.org/10.4103/0973-6247.98922

Arage, G., Ibrahim, S., & Adimasu, E. (2020). Blood donation practice and its associated factors among health professionals of University of Gondar Hospital, Northwest Ethiopia: a cross sectional study. BMC Research Notes, 2017, 1–6. https://doi.org/10.1186/s13104-017-2618-5

Brian Custer, D. J. W. (2013). Adverse reactions and other factors that impact subsequent blood donation visits. 52(1), 118–126. https://doi.org/10.1111/j.1537-2995.2011.03216.x.Adverse

Custer, B., Rios, J. A., Schlumpf, K., Kakaiya, R. M., & Gottschall, J. L. (2012). Blood donors and blood collection. 52(January), 118–126. https://doi.org/10.1111/j.1537-2995.2011.03216.x

Darega,B., Dida, N., Tesfaye,T., & Lencha, B. (2015). Blood Disorders & Transfusion Voluntary Blood Donation Practices and Associated Factors among Regular Undergraduate Madawalabu University Students, Southeast Ethiopia: A Facility- based Cross Sectional Study. 450 ml, 1–7. https://doi.org/10.4172/2155-9864.1000S5-005

Dejen, a. (2015). knowledge , attitude , practice and associated factors of blood donation among health care workers in tikur ambaye dejen (bsc) college of health sciences school of medicine:

Depue, M. (2017). The Benefits of Donating Blood. October, 0–9.

Domaine. (2010). Donor management manual.

Dongen, A. Van. (2015). Easy come , easy go . Retention of blood donors AND NUMBERS. September. https://doi.org/10.1111/tme.12249 Eder, A. F., & Edward, P. (2009). Blood donors and blood collection. https://doi.org/10.1111/j.1537-2995.2012.03666.x

Eder, A. F., & Edward, P. (2014). Blood donors and blood collection. https://doi.org/10.1111/j.1537-2995.2012.03666.x

Eluke, B. C., Okonji, C. U., Ukaejiofo, E., Eluke, C. C., Uchenduk, I., & Ibeagha, I. (2017). Prevalence of Adverse Reaction to Whole Blood Donation Among Voluntary Donors in Asaba, Nigeria. September, 16–21.

Fantahun, D., Submitted, T., & Health, P. (2017a). Evaluation of return rate and return determinants among first time voluntary blood donors in Ethiopian national blood bank service Addis Ababa University College of Health Sciences School of Public Health Evaluation of return rate and return determinants among first time voluntary blood donors in Ethiopian national blood bank service. June. June.

Federal Ministry of Health Ethiopia(FMOH). (2017). Health sector transformation plan -I. 2009.

Flegel, W. A., & Wagner, F. F. (2000). Predicting a donor's likelihood of donating within a preselected time interval. 1999.

Gadisa Kebede. (2016). a study of pre-and- post donation glucose level among blood donors.

Global Database on Blood Safety. (2011). june, 1–9.

Global Status Report on Blood Safety and Availability. (2016). 2 0 1 6.

Guo,N.,Wang,J.,Ness,P.,Yao,F.,Dong,X.,Bi,X.,Li,J.,He,W.,Lu,Y.,Ma,H., Wen, X., Huang, M., David, J., King,M., High,P.,Nelson, K.,& Shan,H. (2012). NIH Public Access.51(3),523–530. https://doi.org/10.1111/j.1537-2995.2010.02876.x.Analysis

Gutierrez, R. G. (2002). Parametric frailty and shared frailty survival models. 1, 22–44. Hosmer and Lemeshow. (1999). APPLIED SURVIVAL ANALYSIS.

Hougaard,P.(2001). BOOK REVIEW Editor: Niels Keiding. chapter 14, 2533–2534. https://doi.org/10.1002/sim.938

Jemberu, Y. A., Esmael, A., & Ahmed, K. Y. (2016). Knowledge , attitude and practice towards blood donation and associated factors among adults in Debre Markos town , Northwest

Ethiopia. BMC Hematology, 1–8. https://doi.org/10.1186/s12878-016-0062-8

John, C. A., Theodora, U. E., Gloria, A. N., & Chika, E. A. (2017). Adverse reactions to blood donation: A descriptive study of 3520 blood donors in a Nigerian tertiary hospital. 72 to 76. https://doi.org/10.4103/0975-2870.197894

Jorge A. Rios, M. D. (2013). NIH Public Access. 52(1), 118–126. https://doi.org/10.1111/j.1537-2995.2011.03216.x.Adverse

Kasraian,L.(2010). Causes of discontinuity of blood donation among donors in Shiraz,Iran: cross-sectional study. 128(5), 272–275.

Kasraian, L., & Tavassoli, A. (2012). Relationship between first-year blood donation , return rate for subsequent donation and demographic characteristics. 448–452. https://doi.org/10.2450/2012.0097-11

Kheiri, S., & Alibeigi, Z. (2015). An analysis of first-time blood donors return behaviour using. March. https://doi.org/10.1111/tme.12177

Landsteiner, K.,& Wiener, A. S. (1937). Rh blood group system.

Lourençon, A. D. F. (2011). Evaluation of the return rate of volunteer blood donors. 190–194. https://doi.org/10.5581/1516-8484.20110052

Masser, B. M., Bove, L. L., White, K. M., & Bagot, K. L. (2016). role of asking for something different. 56(March), 605–613. https://doi.org/10.1111/trf.13390

Mauka, W. I., Mahande, M. J., Msuya, S. E., & Philemon, R. N. (2015). Factors Associated with Repeat Blood Donation at the Northern Zone Blood Transfusion Centre in Tanzania. 2015.

Mirutse,G.,Fisseha,G.,Abebe,L.,& Birhanu,Z.(2014).Intention to donate blood among the eligible population in Mekelle City,Northern Ethiopia: Using the theory of planned behavior Intention to Donate Intention to donate blood among the eligible population in Mekelle City, Northern Ethiopia: Using the.September.https://doi.org/10.11648/j.ajhr.20140204.19

Notari, Zou, S., Fang, C. T., Eder, A. F., Benjamin, R. J., & Dodd, R. Y. (2009). Blood donors

and blood collection. 49(October), 2229–2236. https://doi.org/10.1111/j.1537-2995.2009.02288.x

Nureye, D., & Tekalign, E. (2019). Opportunities and Challenges of Blood Donation and Blood Therapy in Ethiopia. 122–127.

Nwogoh, B., Aigberadion, U., & Nwannadi, A. I. (2013). Knowledge , Attitude , and Practice of Voluntary Blood Donation among Healthcare Workers at the University of Benin Teaching Hospital , Benin City , Nigeria. 2013. O Ogbu, C. U. (2008). Hepatitis B Virus And Blood Transfusion Safety In Sub- Saharan Africa. 7(2), 1–7.

Omura, M. C., Haimowitz, M. D., & Ahn, S. S. (2014). Transfusion complications. 2–5.

Raina,T. R.(2015). Study of Adverse Whole Blood Donor Reactions in Normal Healthy Blood Donors: Experience of Tertiary Health Care Centre in Jammu Region. 31(1), 142–145. https://doi.org/10.1007/s12288-014-0396-y

Tapko, J. B., Toure, B., & Sambo, L. G. (2010). STATUS OF BLOOD SAFETY IN THE Report of the 2010 Survey.

Tilburg,I.Van.(2019). Predicting the effect of complications on donor return: a data driven approach. June.

Veldhuizen, I., Atsma, F., Dongen, A. Van, & Kort, W. De. (2012). Blood donors and blood collection. 52(September), 1871–1879. https://doi.org/10.1111/j.1537-2995.2011.03551.x

Virginia, K., Feliciano, D. O., Ferreira, M., & Mendes, D. M. (2008). First time blood donors and the pattern of return visits to the public blood bank of Recife. 55 81. WBDD. (2014).

Wevers, A., & Baaren. (2014). Characteristics of donors who do or do not return to give blood and barriers to their return. https://doi.org/10.2450/2013.0210-12

WHO. (2010). Towards 100% Voluntary Blood Donation A Global Framework for Action.

WHO. (2011). WHO.

WHO. (2017). 2 0 1 6.

Wiersum-osselton, J. C., Kreek, T. M. Der, Brand, A., Veldhuizen, I., Bom, J. G. Van Der, &

Kort, W. De. (2014a). Risk factors for complications in donors at first and repeat whole blood donation: a cohort study with assessment of the impact on donor return. https://doi.org/10.2450/2013.0262-12

Zainie,N.,Hamid,A.,Basiruddin,R.,& Hassan, N. (2013). The Intention to Donate Blood: An Analysis of Socio-Demographic Determinants.3(6). https://doi.org/10.7763/IJSSH.2013.V3.292

Annex 1: Univariable Analysis Using Parametric Shared Frailty Models for Donors' Dataset

A. Exponential- Gamma and Weibull- Gamma Univariable Analysis *Source: National Blood*

Variable	Category	Exponential - Gamma				Weibull - Gamma			
		Coef	SE	ϕ	p-value	Coef	SE	ϕ	p-value
Gender	Female	Ref		1		Ref.		1	
	Male	-0.40	0.0411	0.670	***	-0.493	0.054	-0.493	***
Age (in years)	18-24	Ref		1		Ref.		1	
	25-44	-0.473	0.0387	0.623	***	-0.580	0.051	0.559	***
	45-65	-0.409	0.0967	0.664	***	-0.504	0.126	0.604	***
Weight (in Kg)	45-49	Ref		1		Ref.		1	
	50-59	-0.306	0.123	0.736	0.013	-0.378	0.158	0.685	0.017
	60-69	-0.724	0.122	2.062	***	-0.889	0.157	0.411	***
	70-79	-0.970	0.123	0.379	***	-1.187	0.159	0.305	***
	$\geq$ 80	-1.310	0.125	0.269	***	-1.588	0.162	4.893	***
Blood group	A	Ref		1		Ref.		1	
	B	0.134	0.0523	1.143	0.01	0.165	0.0686	1.179	0.016
	AB	0.228	0.0867	1.256	0.0085	0.283	0.1138	1.327	0.013
	O	0.084	0.0453	1.087	0.063	0.104	0.0594	1.109	0.08
Rh	Negative	Ref		1		Ref.		1	
	Positive	-0.0006	0.0727	0.999	0.99	-0.00137	0.0954	0.998	0.99
Occupation	Civil servant	Ref		1		Ref.		1	
	Private/ NGO worker	0.460	0.0790	1.584	***	0.543	0.102	1.721	***
	Student	1.078	0.0849	2.938	***	1.301	0.111	3.672	***
	Unemployed	-0.711	0.1947	0.491	0.00026	-0.804	0.251	0.447	0.0014
	Others**	0.434	0.2101	1.543	0.039	0.507	0.271	1.660	0.061
Volume of blood donated(ml)	350	Ref		1		Ref.		1	
	450	-0.724	0.0395	0.484	***	-0.717	0.0526	0.488	***
Donation experience	1st time donor	Ref		1		Ref.		1	
	Repeat donor	0.0877	0.0412	1.091	0.033	0.0989	0.0539	1.103	0.066
Donors reaction	No donor reaction	Ref		1		Ref.		1	
	Donor reaction	-1.10	0.0406	0.332	***	-1.16	0.0486	0.313	***

Bank, Addis Ababa, A.A, Ethiopia; donates blood from September 06, 2012 to September 11, 2013 and whether he/she returned or not would followed until September 2015, Coef=coefficients of the model, SE=Standard error, ϕ =Acceleration factor

B. Lognormal with Gamma and Exponential With Inverse-Gaussian Univariable Analysis

Variable	Category	Lognormal - Gamma				Exponential - Inverse - Gaussian			
		Coef	SE	ϕ	p-value	Coef	SE	ϕ	p-valu
Gender	Female	Ref		1		Ref.		1	
	Male	-0.416	0.0518	0.659	***	-0.399	0.041	0.670	***
Age (in years)	18-24	Ref		1		Ref.		1	
	25-44	-0.501	0.0500	0.605	***	-0.472	0.0387	0.623	***
	45-65	-0.436	0.1292	0.646	0.007	-0.409	0.0967	0.664	***
Weight (in Kg)	45-49	Ref		1		Ref.		1	
	50-59	-0.277	0.135	0.758	0.04	-0.306	0.123	0.736	0.013
	60-69	-0.732	0.135	0.480	***	-0.723	0.122	0.4852	***
	70-79	-0.976	0.138	0.376	***	-0.970	0.123	0.379	***
	$\geq$ 80	-1.317	0.142	0.267	***	-1.309	0.125	0.270	***
Blood group	A	Ref		1		Ref.		1	
	B	0.155	0.0683	1.167	0.023	0.1339	0.052	0.1339	0.01
	AB	0.274	0.1104	1.315	0.013	0.2282	0.0867	1.256	0.008
	O	0.105	0.0595	1.110	0.076	0.0842	0.0453	1.087	0.063
Rh	Negative	Ref		1		Ref.		1	
	Positive	-0.003	0.094	0.996	0.97	-0.0008	0.0727	0.999	0.99
Occupation	Civil servant	Ref		1		Ref.		1	
	Private/ NGO worker	0.414	0.111	1.512	***	0.459	0.0790	1.582	***
	Student	1.101	0.116	3.00	***	1.076	0.084	2.932	***
	Unemployed	-0.681	0.304	0.506	0.025	-0.711	0.1947	0.491	***
	Others**	0.414	0.280	1.512	0.014	0.434	0.2101	1.543	0.039
Volume of blood donated(ml)	350	Ref		1		Ref.		1	
	450	-0.717	0.052	0.488	***	-0.723	0.0395	0.485	***
Donation experience	1st time donor	Ref		1		Ref.		1	
	Repeat donor	0.099	0.0539	1.103	0.066	0.0873	0.0412	1.091	0.034
Donors rxn	No donor reaction	Ref		1		Ref.		1	
	Donor reaction	-1.16	0.048	0.313	***	-1.10	0.040	0.332	***

Source: National Blood Bank, Addis Ababa, A.A, Ethiopia; donates blood from September 06, 2012 to September 11, 2013 and whether he/she returned or not would followed until September 2015, Coef=coefficients of the model, SE=Standard error, ϕ =Acceleration factor

C. Weibull with Inverse-Gaussian and Lognormal with Inverse-Gaussian Univariable Analysis

Variable	Category	Weibull- Inverse -Gaussian				Lognormal- Inverse -Gaussian			
		Coef	SE	ϕ	p-value	Coef	SE	ϕ	p-value
Gender	Female	Ref		1		Ref.		1	
	Male	-0.493	0.054	0.610	***	-0.446	0.0517	0.640	***
Age (in years)	18-24	Ref		1		Ref.		1	
	25-44	-0.579	0.051	0.560	***	-0.524	0.0499	0.592	***
	45-65	-0.504	0.126	0.604	***	-0.445	0.128	0.640	***
Weight (in Kg)	45-49	Ref		1		Ref.		1	
	50-59	-0.378	0.158	0.685	0.017	-0.277	0.134	0.758	0.039
	60-69	-0.889	0.157	0.411	***	-0.748	0.134	0.473	***
	70-79	-1.187	0.159	0.305	***	-1.009	0.137	0.364	***
	$\geq$ 80	-1.587	0.162	0.204	***	-1.376	0.141	0.252	***
Blood group	A	Ref		1		Ref.		1	
	B	0.165	0.0686	1.179	0.016	0.1586	0.067	1.171	0.02
	AB	0.283	0.1138	1.327	0.013	0.2700	0.110	1.309	0.014
	O	0.104	0.059	1.109	0.08	0.0996	0.059	1.104	0.093
Rh	Negative	Ref		1		Ref.		1	
	Positive	-0.0016	0.095	0.998	0.99	-0.0037	0.094	0.996	0.97
Occupation	Civil servant	Ref		1		Ref.		1	
	Private/ NGO worker	0.543	0.102	1.721	***	0.462	0.111	1.587	***
	Student	1.300	0.110	3.669	***	1.166	0.116	3.209	***
	Unemployed	-0.804	0.251	0.447	0.0014	-0.673	0.302	0.510	0.026
	Others**	0.507	0.271	1.660	0.061	0.411	0.279	1.508	0.14
Volume of blood donated(ml)	350	Ref		1		Ref.		1	
	450	-0.872	0.051	0.418	***	-0.784	0.0529	0.456	***
Donation experience	1st time donor	Ref		1		Ref.		1	
	Repeat donor	0.107	0.054	1.112	0.047	0.106	0.0537	1.111	0.049
Donors reaction	No donor reaction	Ref		1		Ref.		1	
	Donor reaction	-1.31	0.053	0.269	***	-1.22	0.0485	0.295	***

Source: National Blood Bank, Addis Ababa, A.A, Ethiopia; donates blood from September 06, 2012 to September 11, 2013 and whether he/she returned or not would followed until September 2015, Coef=coefficients of the model, SE=Standard error, ϕ =Acceleration factor

Annex 2: Multivariable Analysis Using Different Parametric Shared Frailty Models for Donors Dataset

A. Exponential - Gamma multivariable shared Frailty Model

Covariates	Category	Coef	S.E	ϕ	95% CI	p-value
	Female	Ref		1		
Gender	Male	0.114	0.045	0.892	[0.8167, 0.974]	0.011
	18-24	Ref		1		
Age	25-44	5.03	1.033	1.092	[0.996, 1.197]	***
(in years)	45-65	0.191	0.102	1.210	[0.989, 1.480]	0.063
	45-49	Ref		1		
	50-59	-0.176	0.124	0.839	[0.658, 1.069]	0.16
Weight	60-69	-0.402	0.126	0.669	[0.523, 0.856]	0.0014
(in Kg)	70-79	-0.445	0.131	0.641	[0.496 0.829]	***
	≥ 80	-0.593	0.135	0.553	[0.425 0.719]	***
	A	Ref		1		
Blood	B	0.097	0.052	1.101	[0.994 1.221]	0.065
group	AB	0.184	0.087	1.203	[1.0144 1.426]	0.034
	O	0.059	0.045	1.061	[0.970 1.159]	0.19
	Civil servant	Ref		1		
	Private/ NGO worker	0.076	0.080	1.079	[0.922 1.264]	0.34
Occupation	Student	0.313	0.093	1.368	[1.141 1.641]	***
	Unemployed	-0.318	0.197	0.728	[0.495 1.070]	0.11
	Others**	0.365	0.211	1.440	[0.953 2.176]	0.084
Volume of blood don ated(ml)	450	Ref 0.0658	0.046	1 1.068	[0.976 1.169]	0.15
Donation	1st time donor	Ref		1		
experience	Repeat donor	-0.962	0.049	0.382	[0.347 0.421]	***
Donors	No donor reaction	Ref		1		
reaction	Donor reaction	1.415	0.135	4.115	[3.161 5.360]	***

*Source: National Blood Bank, Addis Ababa, A.A, Ethiopia; donates blood from September 06, 2012 to September 11, 2013 and whether he/she returned or not would followed until September 2015, *Statistically significant at 5% level, ϕ=Acceleration factor, θ=Variance of the random effect, τ= Kendall's tau,S.E=Standard error, Ref=Reference*

B. Weibull-Gamma multivariable shared Frailty Model

Covariates	Category	Coef	S.E	ϕ	95% CI	p-value
	Female	Ref		1		
Gender	Male	-0.124	0.054	0.883	[0.795, 0.981]	0.021
	18-24	Ref		1		
Age	25-44	0.100	0.056	1.105	[0.990, 1.233]	0.074
(in years)	45-65	0.220	0.123	1.246	[0.980, 1.584]	0.073
	45-49	Ref		1		
	50-59	-0.205	0.148	0.814	[0.609, 1.089]	0.17
Weight	60-69	-0.463	0.150	0.630	[0.468, 0.845]	0.0021
(in Kg)	70-79	-0.5134	0.157	0.598	[0.440, 0.813]	0.001
	≥ 80	-0.678	0.161	0.507	[0.370, 0.695]	***
	A	Ref		1		
Blood	B	0.108	0.063	1.114	[0.985, 1.259]	0.084
group	AB	0.205	0.104	1.227	[1.002, 1.505]	0.048
	O	0.066	0.054	1.067	[0.960, 1.187]	0.23
	Civil servant	Ref		1		
	Private/ NGO worker	0.083	0.096	1.086	[0.899, 1.311]	0.39
Occupation	Student	0.361	0.111	1.434	[1.154, 1.782]	0.001
	Unemployed	-0.322	0.235	0.724	[0.457, 1.1484]	0.17
	Others**	0.402	0.251	1.494	[0.912, 2.445]	0.11
Volume of	350	Ref		1		
blood don	450	0.072	0.055	1.074	[0.964, 1.197]	0.19
ated(ml)						
Donation	1st time donor	Ref		1		
experience	Repeat donor	-1.069	0.059	0.343v	[0.306, 0.385]	***
Donors	No donor reaction	Ref		1		
reaction	Donor reaction	1.603	0.162	4.967	[3.617, 6.821]	***

*Source: National Blood Bank, Addis Ababa, A.A, Ethiopia; donates blood from September 06, 2012 to September 11, 2013 and whether he/she returned or not would followed until September 2015, *Statistically significant at 5% level, ϕ=Acceleration factor, θ=Variance of the random effect, τ= Kendall's tau,S.E=Standard error, Ref=Reference*

C. Log-Logistic-Gamma multivariable shared Frailty Model

Covariates	Category	Coef	S.E	ϕ	95% CI	p-value
	Female	Ref		1		
Gender	Male	-0.086	0.054	0.917	[0.825, 1.019]	0.11
	18-24	Ref		1		
Age	25-44	0.108	0.057	1.114	[0.995, 1.247]	0.06
(in years)	45-65	0.268	0.126	1.307	[1.020, 1.676]	0.34
	45-49	Ref		1		
	50-59	-0.185	0.138	0.830	[0.633, 1.088]	0.18
Weight	60-69	-0.446	0.141	0.639	[0.485, 0.843]	0.0016
(in Kg)	70-79	0.147	0.603	[0.451, 0.806]	***	
	≥ 80	-0.674	0.153	0.509	[0.377, 0.688]	***
	A	Ref		1		
Blood	B	0.113	0.063	1.120	[0.988, 1.269]	0.076
group	AB	.184	0.104	1.202	[0.978, 1.476]	0.079
	O	0.061	0.055	1.063	[0.953, 1.185]	0.27
	Civil servant	Ref		1		
	Private/ NGO worker	0.082	0.103	1.086	[0.886, 1.331]	0.42
Occupation	Student	0.405	0.115	1.499	[1.195 1.880]	***
	Unemployed	-0.109	0.263	0.896	[0.534, 1.502]	0.68
	Others**	0.359	0.266	1.432	[0.849, 2.414]	0.18
Volume of	350	Ref		1		
blood don	450	0.066	0.057	1.068	[0.954, 1.195]	0.25
ated(ml)						
Donation	1st time donor	Ref		1		
experience	Repeat donor	-0.976	0.062	0.376	[0.333, 0.425]	***
Donors	No donor reaction	Ref		1		
reaction	Donor reaction	1.685	0.149	5.394	[4.021, 7.234]	***

*Source: National Blood Bank, Addis Ababa, A.A, Ethiopia; donates blood from September 06, 2012 to September 11, 2013 and whether he/she returned or not would followed until September 2015, *Statistically significant at 5% level, ϕ=Acceleration factor, θ=Variance of the random effect, τ= Kendall's tau,S.E=Standard error, Ref=Reference*

D. Exponential - Inverse-Gaussian multivariable shared Frailty Model

Covariates	Category	Coef	S.E	ϕ	95% CI	p-value
	Female	Ref		1		
Gender	Male	-0.114	0.045	0.892	[0.816, 0.974]	0.011
	18-24	Ref		1		
Age	25-44	0.087	0.047	1.091	[0.995, 1.197]	0.062
(in years)	45-65	0.190	0.102	1.210	[0.989, 1.480]	0.063
	45-49	Ref		1		
	50-59	-0.176	0.124	0.838	[0.657, 1.069]	0.16
Weight	60-69	-0.402	0.125	0.668	[0.522, 0.856]	0.0014
(in Kg)	70-79	-0.444	0.131	0.641	[0.495, 0.828]	***
	≥ 80	-0.593	0.134	0.552	[0.424, 0.719]	***
	A	Ref		1		
Blood	B	0.096	0.052	1.101	[0.993, 1.220]	0.651
group	AB	0.184	0.086	1.202	[1.014, 1.426]	0.034
	O	0.058	0.045	1.060	[0.970, 1.159]	0.19
	Civil servant	Ref		1		
	Private/ NGO worker	0.076	0.080	1.079	[0.921, 1.263]	0.34
Occupation	Student	0.313	0.092	1.368	[1.140, 1.642]	***
	Unemployed	-0.318	0.196	0.727	[0.494, 1.070]	0.11
	Others**	0.364	0.210	1.440	[0.952, 2.176]	0.084
Volume of	350	Ref		1		
blood don	450	0.065	0.046	1.067	[0.975, 1.168]	0.16
ated(ml)						
Donation	1st time donor	Ref		1		
experience	Repeat donor	-0.961	0.048	0.382	[0.347, 0.420]	***
Donors	No donor reaction	Ref		1		
reaction	Donor reaction	1.414	0.134	4.115	[3.160, 5.359]	¡.001 ***

*Source: National Blood Bank, Addis Ababa, A.A, Ethiopia; donates blood from September 06, 2012 to September 11, 2013 and whether he/she returned or not would followed until September 2015, *Statistically significant at 5% level, ϕ=Acceleration factor, θ=Variance of the random effect, τ= Kendall's tau,S.E=Standard error, Ref=Reference*

E. Weibull - Inverse-Gaussian multivariable shared Frailty Model

Covariates	Category	Coef	S.E	ϕ	95% CI	p-value
Gender	Female	Ref		1		
	Male	-0.123	0.053	0.883	[0.795, 0.9814]	0.021
Age (in years)	18-24	Ref		1		
	25-44	0.100	0.056	1.105	[0.990, 1.233]	0.074
	45-65	0.220	0.122	1.246	[0.980, 1.584]	0.73
Weight (in Kg)	45-49	Ref		1		
	50-59	-0.204	0.148	0.814	[0.609, 1.089]	0.17
	60-69	-0.462	0.150	0.629	[0.468, 0.845]	0.0021
	70-79	-0.513	0.156	0.598	[0.440, 0.813]	0.001
	≥ 80	-0.677	0.1607	0.507	[0.370, 0.695]	***
Blood group	A	Ref		1		
	B	0.108	0.062	1.114	[0.985, 1.259]	0.084
	AB	0.205	0.103	1.227	[1.00, 1.504]	0.048
	O	0.065	0.054	1.067	[0.960, 1.187]	0.23
Occupation	Civil servant	Ref		1		
	Private/ NGO worker	0.082	0.096	1.086	[0.899, 1.311]	0.39
	Student	0.361	0.110	1.434	[1.154, 1.782]	0.0011
	Unemployed	-0.321	0.234	0.724	[0.457, 1.148]	0.17
	Others**	0.401	0.251	1.494	[0.912, 2.445]	0.11
Volume of blood donated(ml)	350	Ref		1		
	450	0.071	0.055	1.074	[0.964, 1.196]	0.19
Donation experience	1st time donor	Ref		1		
	Repeat donor	-1.069	0.059	0.343	[0.305, 0.385]	***
Donors reaction	No donor reaction	Ref		1		
	Donor reaction	1.602	0.161	4.967	[3.617, 6.821]	***

*Source: National Blood Bank, Addis Ababa, A.A, Ethiopia; donates blood from September 06, 2012 to September 11, 2013 and whether he/she returned or not would followed until September 2015, *Statistically significant at 5% level, ϕ=Acceleration factor, θ=Variance of the random effect, τ= Kendall's tau,S.E=Standard error, Ref=Reference*

F. Lognormal - Inverse-Gaussian multivariable shared Frailty Model

Covariates	Category	Coef	S.E	ϕ	95% CI	p-value
Gender	Female	Ref		1		
	Male	-0.073	0.051	0.929	[0.839, 1.029]	0.16
Age (in years)	18-24	Ref		1		
	25-44	0.093	0.055	1.097	[0.983, 1.224]	0.096
	45-65	0.249	0.125	1.283	[1.003, 1.640]	0.047
Weight (in Kg)	45-49	Ref		1		
	50-59	-0.126	0.126	0.881	[0.688, 1.128]	0.32
	60-69	-0.360	0.129	0.697	[0.541, 0.898]	0.0053
	70-79	-0.411	0.136	0.662	[0.507, 0.866]	0.0026
	≥ 80	-0.580	0.142	0.559	[0.423, 0.740]	***
Blood group	A	Ref		1		
	B	0.099	0.062	1.105	[0.978, 1.248]	0.11
	AB	0.148	0.101	1.160	[0.952, 1.413]	0.14
	O	0.052	0.054	1.053	[0.947, 1.171]	0.33
Occupation	Civil servant	Ref		1		
	Private/ NGO worker	0.079	0.104	1.083	[0.881, 1.330]	0.45
	Student	0.385	0.115	1.470	[1.172, 1.844]	***
	Unemployed	-0.173	0.281	0.840	[0.484, 1.459]	0.54
	Others**	0.331	0.260	1.393	[0.836, 2.320]	0.2
Volume of blood donated(ml)	350	Ref		1		
	450	0.049	0.056	1.050	[0.940, 1.174]	0.038
Donation experience	1st time donor	Ref		1		
	Repeat donor	-0.902	0.062	0.405	[0.359, 0.458]	***
Donors reaction	No donor reaction	Ref		1		
	Donor reaction	1.651	0.147	5.214	[3.909, 6.956]	***

*Source: National Blood Bank, Addis Ababa, A.A, Ethiopia; donates blood from September 06, 2012 to September 11, 2013 and whether he/she returned or not would followed until September 2015, *Statistically significant at 5% level, ϕ=Acceleration factor, θ=Variance of the random effect, τ= Kendall's tau,S.E=Standard error, Ref=Reference*

G. Log-Logistic - Inverse-Gaussian multivariable shared Frailty Model

Covariates	Category	Coef	S.E	ϕ	95% CI	p-value
	Female	Ref		1		
Gender	Male	-0.086	0.054	0.917	[0.825, 1.019]	0.11
	18-24	Ref		1		
Age	25-44	0.108	0.057	1.114	[0.995, 1.247]	0.06
(in years)	45-65	0.268	0.126	1.307	[1.020, 1.676]	0.034
	45-49	Ref		1		
	50-59	-0.185	0.138	0.830	[0.633, 1.088]	0.18
Weight	60-69	-0.446	0.141	0.639	[0.485, 0.843]	0.0016
(in Kg)	70-79	-0.504	0.147	0.603	[0.451, 0.806] ***	
	≥ 80	-0.674	0.153	0.509	[0.377, 0.688]	***
	A	Ref		1		
Blood	B	0.113	0.063	1.120	[0.988, 1.269]	0.076
group	AB	0.184	0.104	1.202	[0.978, 1.476]	0.079
	O	0.061	0.055	1.063	[0.953, 1.185]	0.27
	Civil servant	Ref		1		
	Private/ NGO worker	0.082	0.103	1.086	[0.886, 1.331]	0.42
Occupation	Student	0.405	0.115	1.499	[1.195, 1.880]	***
	Unemployed	-0.109	0.263	0.896	[0.534, 1.502]	0.68
	Others**	0.359	0.266	1.432	[0.849, 2.414]	0.18
Volume of	350	Ref		1		
blood don	450	0.066	0.057	1.068	[0.954, 1.196]	0.25
ated(ml)						
Donation	1st time donor	Ref		1		
experience	Repeat donor	-0.976	0.062	0.376	[0.333, 0.425]	***
Donors	No donor reaction	Ref		1		
reaction	Donor reaction	1.685	0.149	5.394	[4.021, 7.234]	***

*Source: National Blood Bank, Addis Ababa, A.A, Ethiopia; donates blood from September 06, 2012 to September 11, 2013 and whether he/she returned or not would followed until September 2015, *Statistically significant at 5% level, ϕ=Acceleration factor, θ=Variance of the random effect, τ= Kendall's tau,S.E=Standard error, Ref=Reference*

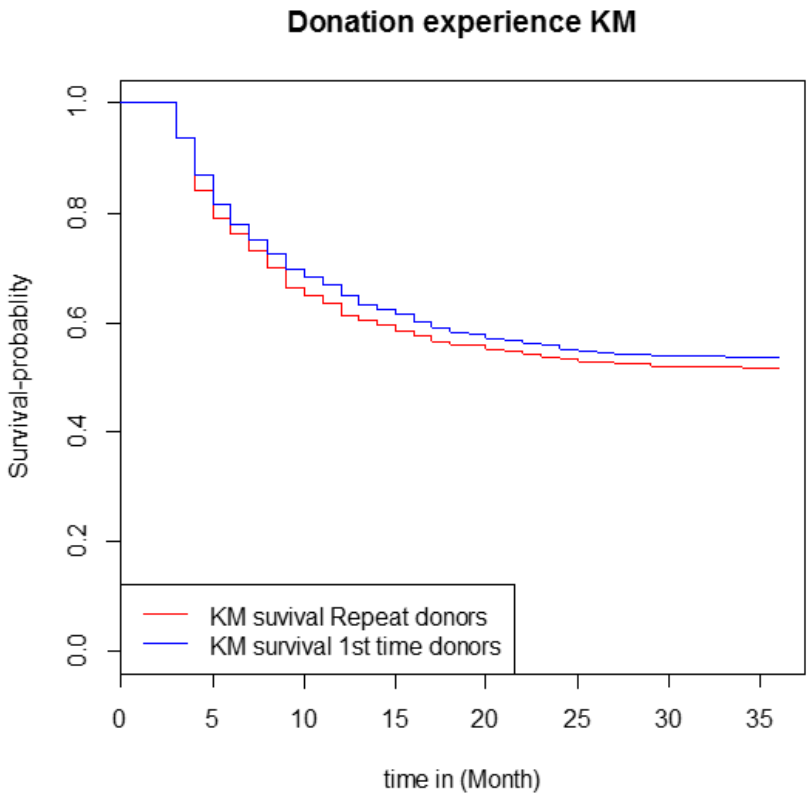

Figure 5.1: The survival functions of donation experiance of blood donors' using the lognormal-gamma frailty model.

Annex 3: Survival curve for significantly different groups

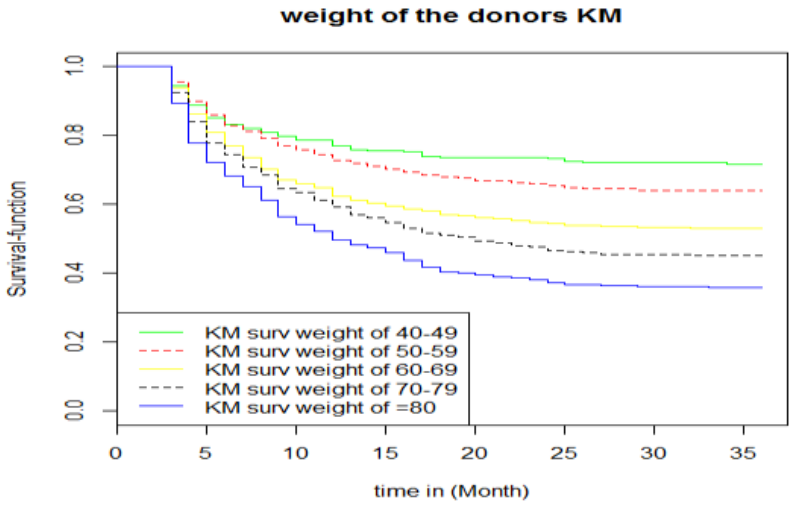

Figure 5.2: The survival functions of Weight of blood donors' using the lognormal- gamma frailty model.

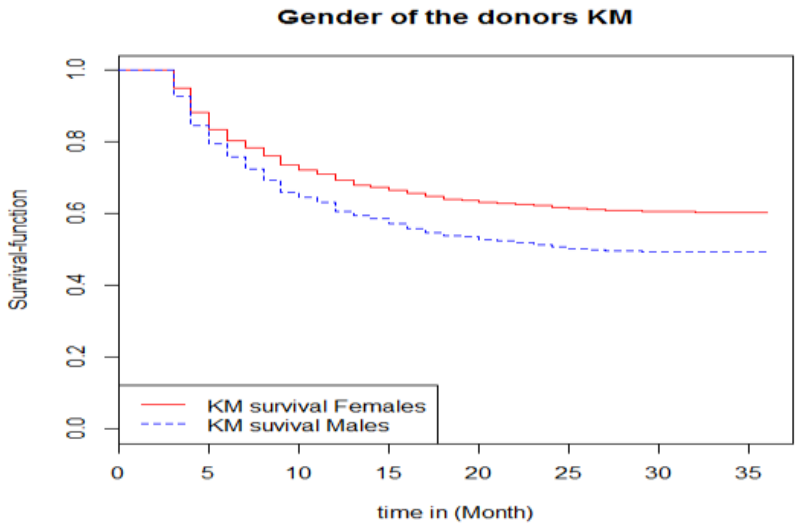

Figure 5.3: The survival functions of gender of donors of blood donors' using the lognormal-gamma frailty model.

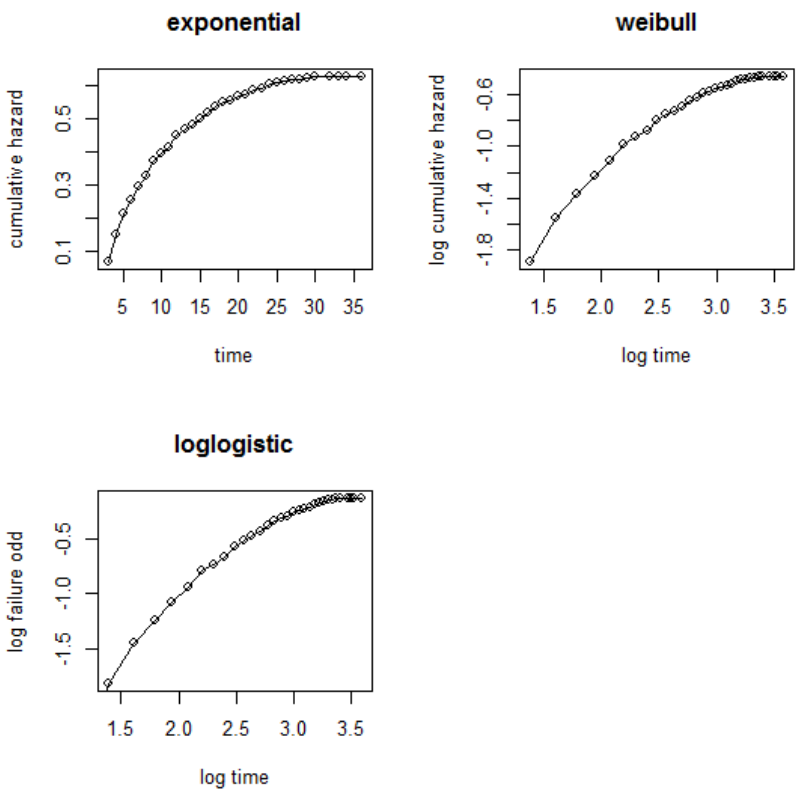

Figure 5.4: Graphical evaluation of the exponential, Weibull, log-logistic and log-normal assumptions

Annex 4: Graphical evaluation of the exponential, Weibull, log-logistic and log-normal assumptions