STATISTICAL ANALYSIS OF FACTORS ASSOCIATED WITH DURATION OF EXCLUSIVE BREAST FEEDING IN ETHIOPIA: A COMPARISON OF COX PH AND SHARED FRAILTY MODELS

By**:** Anware Mohammed

A Thesis Submitted to the Department of Statistics, School of Graduate Studies, College of Natural Science, Jimma University in Partial Fulfillment for the Requirements of Masters of Science (MSc) Degree in Biostatistics

October, 2014

Jimma, Ethiopia

STATISTICAL ANALYSIS OF FACTORS ASSOCIATED WITH DURATION OF EXCLUSIVE BREAST FEEDING IN ETHIOPIA: A COMPARISON OF COX PH AND SHARED FRAILTY MODELS

MSc Thesis

By: Anware Mohammed

Advisor: Yehenew Getachew (PhD)

Co-advisor: Akalu Banbeta (MSc)

October, 2014

Jimma, Ethiopia

JIMMA UNIVERSITY

SCHOOL OF GRADUATE STUDIES, DEPARTMENT OF STATISTICS

As thesis research advisors, we here by certify that we have read the thesis prepared by **ANWARE MOHAMMED ALI** under our guidance, which is entitled "Statistical Analysis of Factors associated with Duration of Exclusive Breastfeeding in Ethiopia: a Comparison of Cox PH and Shared Frailty Models", in its final form and have 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.

Yehenew Getachew (PhD)	_____	_____
Advisor	Signature	Date
Akalu Banbeta (MSc)	_____	_____
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.

_____	_____	_____
Name of chairman	Signature	Date
_____	_____	_____
Name of Major Advisor	Signature	Date
_____	_____	_____
Name of Co-advisor	Signature	Date
_____	_____	_____
Name of internal Examiner	Signature	Date
_____	_____	_____
Name of External Examiner	Signature	Date

STATEMENT OF AUTHOR

I declare that this thesis is a result of my genuine work and all sources of materials used, for writing it, have been duly acknowledged. I have submitted this thesis to Jimma University in the partial fulfillment for the requirements of 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.

Anware Mohammed

Date_____

Signature _____

Jimma University, Jimma

ACKNOWLEDGMENT

This thesis could not have been accomplished without assistance of a number of persons and I would like to take these few lines to thank them warmly for their help.

First of all, I owe my deep gratitude to my advisor, Dr. Yehenew Getachew without whom accomplishment of this work would not have been made possible. His continuous supportive comments and constant guidance throughout the course of my thesis work get me out of my frustration. It is a real pleasure working with him. I can only hope that his cooperation will keep on going in the future.

Secondly, I gratefully acknowledge my co-advisor Mr. Akalu Banbeta for his continuous and passionate support in every aspect I needed. His consistent concern and help made me concentrate on my thesis.

Thirdly, I always thank and thought you for everything you made throughout my life to my mother Asnaku Abejew and my father Mohammed Ali even if you are not with us now.

My completion of this thesis could not have been accomplished without the support of my classmates and my friends. Thank you for allowing me time away from you to research and write.

Finally, to my caring, loving, and supportive wife, Nefisa Mohammed: my deepest gratitude. Your encouragement when the times got rough are much appreciated and duly noted.

TABLE OF CONTENTS

LISTS OF TABLES

LISTS OF FIGURES

ABSTRACT

Background: Exclusive breastfeeding during the early months of life reduce infant morbidity and mortality. Current recommendation in Ethiopia is to continue exclusive breastfeeding up to six months of age.

Objective: The main objective of this paper was to model duration of exclusive breastfeeding by using Cox PH and shared frailty models.

Methods: The data was obtained from the EDHS, 2011.The study sample (n = 1371) was based on infants (0–6 months old) during the survey period; extracted from the women data base. The analysis was done by using ordinary Cox's PH model, Cox PH with gamma frailty and Gaussian frailty models with the help of R statistical package. These models were compared by using AIC.

Results: Two-thirds (63.96%) of children were introduced with liquid and solid food supplements and breast feeding becomes partial before and at six months of age. The estimated median time for infants to introduce breast milk substitute is five months with 95% CI (4, 5.3) months. The variance of the frailty term for Cox PH with gamma frailty model was 0.281 which was significant.

Conclusion: The cox proportional hazard with gamma frailty model provides a suitable choice for modeling duration of exclusive breastfeeding. Place of residence, economic status (wealth index), contraceptive use, education level have significant influence on duration exclusive breastfeeding. There is regional variation in duration of exclusive breast feeding.

Key words*: Breast feeding, Cox PH, EBF, hazard ratio, frailty, Time-to-event analysis.*

ACRONYMS

AIC	Akaike Information Criteria
AIDS	Acquired Immune Deficiency Syndrome
CSA	Central Statistical Agency
DHS	Demographic and Health Survey
EA	Enumeration Area
EBF	Exclusive Breast Feeding
EDHS	Ethiopia Demographic Health Survey
HIV	Human Immune deficiency Virus
HR	Hazard Ratio
ML	Maximum Likelihood
MoH	Ministry of health
MDG	Millennium Development Goal
PDF	Probability Density Function
PH	Proportional Hazard
UNICEF	United Nations Children's Fund
USAID	United States Agency for International Development
WHO	World Health Organization

1. INTRODUCTION

1.1. Background

Exclusive breastfeeding defined by WHO as practice of feeding only breast milk and allows the baby to receive vitamins, minerals or medicines and water, breast milk substitutes, other liquids and solid foods are excluded. All women should be enabled to practice EBF and all infants should be fed exclusively on breast milk from birth to 4 to 6 months of age and thereafter, children should continue to be breastfed while receiving appropriate and adequate complementary foods for up to 2 years of age or beyond (WHO, 1995). The previous recommendation of EBF for four months was changed to 6 months based on accumulative evidence based research by WHO (WHO, 2001; WHO, 2002). World Health Assembly of WHO in 2001 made resolution that exclusive breastfeeding for the first six months is the most appropriate infant feeding practice (Owen *et al*., 2008). Recent developments suggest full breastfeeding should continue to six months (Santos *et al*., 2008). There is good evidence these two more months of EBF from fourth to six months provides infants with additional protection against gastrointestinal and acute respiratory infections during that two months period (Santos and Victoria, 2008).

The benefits of prolonged breastfeeding for mother and infant health are documented in a vast scientific literature. Extensive research in various countries has also provided evidence that breastfeeding has clear health benefits for infants as well as the mother. Infants who have been breastfed optimally have reduced risk of common childhood illnesses such as gastrointestinal and respiratory infections, otitis media, atopic eczema, and allergy during childhood (Yoon, 1996; Foo, 2005).

Malnutrition has been responsible, directly or indirectly, for 60% of the 10.9 million deaths annually among children under five worldwide. Over two-thirds of these deaths, which are often associated with inappropriate feeding practices, occur during the first year of life. Globally, no more than 35% of infants are exclusively breastfed during the first four months of life (WHO, 2003). Global risk assessment of suboptimal breastfeeding indicates that 96% of all infant deaths in developing countries are attributable to inappropriate feeding occurring during the first six months of life (Lauer et al.,2006)

In low income countries like Ethiopia, it has been estimated that practicing EBF can reduce under five mortality by 13% (Jones, et al., 2003). In order to achieve the Millennium Development Goal of reduction of child mortality, infant breastfeeding has been identified as one of the major intervention areas both globally and nationally (Aarts and Foo, 2005). But still too early introduction of breast milk substitutes and too late introduction of semi-solid complementary feeds are common and are responsible for rapid increase in the prevalence of under nutrition between 6-24 months (Lambert*iet al*., 2011). To strengthen the effort in reducing child mortality, the Ethiopian Ministry of Health had targeted an increase in the proportion of exclusively breastfed infants under age 6 months to 70 percent by 2015 as one strategy to improve child health (MoH, 2010).The 2011 Ethiopian Demographic and Health Survey (EDHS) showed the proportion of infants under six months who received EBF as 52% (CSA, 2012) which improved slightly (only 3%) compared to 2005 EDHS (CSA, 2006). Therefore, assessing factors associated with exclusive breastfeeding is crucial to implement interventions that speed up the government efforts and decrease the rates and burden of infant morbidity and mortality.

Research into demographic and socioeconomic variables and factors related to healthcare and to the habits of mothers and babies on a population level can be of great utility to identifying factors related to duration of exclusive breastfeeding or complementary feeding. They can, therefore, be useful tools in the quest to increase breastfeeding rates in our country. Nevertheless, regional differences in breastfeeding practices underscore the need for focused diagnoses that can guide the taking of intervention measures aimed at promoting, supporting and protecting breastfeeding. The objective of this study was to identify variables associated with duration of exclusive breastfeeding in Ethiopia and choose the appropriate model to model these variables.

1.2. Statements of the Problem

Progress EBF rates has been made since the early 1990s, although the rates remain too low across the developing world and poor continuation of breastfeeding with inadequate complementary feeding practices is still widespread. Only about one-third (36%) of newborns are exclusively breastfed for the first six months of life (UNICEF). Current breastfeeding patterns are still far from the recommended level and considerable variation exists across regions

It is worth noting that each year under nutrition is implicated in about 40% of the 11 million deaths of children under five in developing countries, and lack of immediate and EBF in infancy causes an additional 1.5 million of these deaths (www.unicef.org/nutrition/index.html). Based on several studies done in Ethiopia, breastfeeding is nearly universal but the proportion of EBF children under 6 months of age is less than the optimal recommendations (Girma, 2002; CSA, 2006).

In our data set children are clustered within a locality (regions). It is recognized that individuals in the same region are more similar than individuals in different region because they shared similar (possibly unmeasured) environmental exposures. In traditional survival data analysis independence among observations is a standard but important assumption. However; in many epidemiological studies this assumption may be violated since survival times are clustered into groups such as families or geographical units: some unmeasured characteristics shared by the members of that cluster, such as genetic information or common environmental exposures could influence time to the studied event, but still Survival times from different clusters are assumed to be independent. Thus, a further extension of the Cox model should be considered by taking in to account the hierarchical (clustered) structure of the data, i.e., the nesting of children with in localities. This clustering can be taken in to account by adding a random effect as extra term. The locality is taken to be a random effect rather than a fixed effect because the individual locality is not of interest by itself; interest is rather in the heterogeneity between localities. Furthermore, introducing many fixed effects in a model might lead to convergence problems, especially if there is little variation in the covariates between localities (McGilchrist and Aisbett, 1991).

In the thesis we have applied the shared frailty model on 2011 Ethiopia Demographic Health Survey (EDHS 2011) data. The focus is on the effect of important biological and social factors short duration of EBF in Ethiopia. The data clustered in to nine regional state and two administrative cities, with large variation in term of economic performance and standard of living. To have the correct estimates of our parameter we have used the shared frailty model and we compare this model with the cox proportional hazard model. Therefore, this thesis would like to address the following interesting research questions:

- Is there a significant heterogeneity among regions with respect to duration of EBF?
- What are the covariates which influence the duration of EBF in Ethiopia?

- ❖ Which type of survival models, Cox-PH, shared gamma or shared inverse Gaussian frailty models, predicts well the duration of EBF in Ethiopia?

1.3. Objectives of the Study

1.3.1. General Objective

The main objective of this paper was modeling duration of EBF by using Cox PH and shared frailty models based on the data obtained from EDHS (2011) in Ethiopia.

1.3.2. Specific Objectives

Specific issues that this paper addressed include:

- ✓ Comparing adjusted and unadjusted Cox hazard models for the clustering.
- ✓ Comparing semi-parametric gamma and inverse Gaussian frailty models for the clustering.
- ✓ Identifying risk factors to early stop to EBF.
- ✓ Estimating the median duration of EBF in children under six months in Ethiopia.

1.4. Significance of the Study

This study will be useful to understand how important to consider the shared structure of this survival data. It will be useful for policy makers to show how status of EBF is now in Ethiopia in order to achieve the MDGs of reduction of child mortality because infant breastfeeding has been identified as one of the major intervention areas both globally and nationally. The study outcome can also guide the taking of intervention measures aimed at promoting, supporting and protecting breastfeeding. The outcome of this study will encourage health-care providers to recommend exclusive breastfeeding and will reinforce health strategies for promoting prolonged breastfeeding. It is useful for future researchers who want to conduct their research on breast feeding related aspect as a base line. It will also helpful for those who are interested to know the socioeconomic and demographic factors to too early introduction of breast milk substitutes in children under six months of old in Ethiopia.

2. LITRATURE REVIEW

2.1. Breast Feeding

Breastfeeding is a practice that has been around for thousands of years and which offers well-known nutritional, immunological, cognitive, economic and social benefits. These benefits are only taken advantage to the full when breastfeeding continues for at least 2 years, being the sole means of feeding infants for their first six months of life (WHO, 1995). WHO estimates that worldwide only 35% of children between birth and their fifth month are breastfed exclusively (WHO, 1995).

World Health Assembly of WHO in 2001 made resolution that exclusive breastfeeding for the first six months is the most appropriate infant feeding practice (Owen et al., 2008). Recent developments suggest full breastfeeding should continue to six months (Santos et al., 2008). EBF has been defined by WHO as the situation where 'the infant has received only breast milk from his/her mother or a wet nurse, or expressed breast milk and no other liquids, or solids, with the exception of drops or syrups consisting of vitamins, minerals supplements, or medicines (WHO, 1991). EBF is adequate in quality as well as quantity in terms of energy, protein, nutrients, water etc., for an infant's need under six months of age (WHO, 2002). Despite its demonstrated benefits, EBF prevalence and duration in many countries including Ethiopia are lower than the international recommendation of exclusive breastfeeding for the first six months of life (Foo, 2005; Haider, 2006). Based on several studies done in Ethiopia, breastfeeding is nearly universal but the proportion of exclusively breastfed children up to 6 months is less than the optimal recommendations (Girma, 2002; CSA, 2006). The low prevalence of EBF in most developing countries including Ethiopia is attributed to various maternal and child factors such as place of residence, sex and age of the child, mother working outside home, maternal age and educational level, access to mass media and economic status by several researchers (Simopoulos,1984;Ssenyonga,2000;Morisky,2002; Shirima,2004; Haider,2006).

2.2. Benefits of Breast Feeding

In resource poor countries, where the negative impact of HIV/AIDS is high, EBF for the first six months has greater benefit than mixed feeding or formula feeding for the prevention of mother to

child transmission of HIV (Ssenyonga et al., 2004). Breast milk provides all of the nutrients, vitamins and minerals an infant needs for growth for the first six months, and no other liquids or food are needed. Breast milk carries antibodies from the mother that help combat disease, which breast milk substitutes cannot contain. In addition, breast milk contains digestive enzymes which breast milk substitutes do not contain, and therefore the infant easily digests and efficiently uses the breast milk (Lawrence, 1994). From the age of six months, breast milk is no longer sufficient by itself, but it continues to be an important source of energy, high quality nutrients and anti-infective factors beyond six months of age.

Optimal breastfeeding practices in the first two years of life, especially EBF for the first six months of life, can have the single largest impact on child survival of all preventive interventions, with the potential to prevent 12-13% of all under-5 deaths in the developing world, or 1.4 million lives, according to the 2008 Lancet Nutrition Series (Robert, 2008). There is growing evidence of the significant impact of early initiation of breastfeeding (within first hour and first day after birth) on reducing overall neonatal mortality. A recent study from rural Ghana (Edmond, 2006) shows that early initiation within the first hour could prevent 22% of neonatal deaths and initiation of breastfeeding within the first day could prevent 16% of neo-natal deaths. A study in Nepal (Mullany et al,2008) found that approximately 19.1% and 7.7% of all neo-natal deaths could be avoided with universal initiation of breastfeeding within the first hour and first day of life respectively. Breastfed children have at least 14 times greater chance of survival in the early months than non-breastfed children (Black et al., 2008). In the first 6 months of life, non-breastfed infants were more than 14 times more likely than to die from all causes, 10 times more likely to die from diarrhea and 15 times more likely to die from acute respiratory infection two major child killers. During the first six months, the rates of infections are lower for exclusively breastfed than for partially breastfed infants (Arifeen et al., 2001).

Breastfeeding protects infants against diarrhea through two mechanisms: 1) reduced risk of bacteria from contaminated formula, other liquids and complementary foods, and 2) the transfer of maternal antibodies through breast milk (Long et al, 1994). In three large studies conducted in Africa, exclusive breastfeeding for up to six months was associated with a three to four fold decreased risk of transmission of HIV compared to mixed feeding (Becquet et al., 2005; Iliff et al, 2005; Coovadia et al., 2007). Weighed against the low but ongoing risk of transmission

through breast milk (BHITS Group, 2004), breastfeeding substantially reduces the risk of infant mortality from other infectious diseases and malnutrition on average by 4–6 fold in the first six months and close to two fold in the second six months of life (WHO, 2000). Since breastfeeding as commonly practiced carries a cumulative risk of HIV transmission of around 1% for every additional month of breastfeeding, feeding the infant with breast milk for a shorter period than usual reduces this risk. However, any benefits of shortening the exclusive breastfeeding period in terms of HIV transmission are unlikely to overcome the elevated risks of morbidity, mortality and malnutrition from early cessation before six months. Breastfeeding of HIV-infected infants beyond six months was associated with improved survival compared to stopping breastfeeding (WHO, 2006).

Due to its large impact on reduction of infectious diseases, breastfeeding plays a role in reduction of stunting, as infectious diseases are important determinants of stunting (Scrimshaw et al., 1968). However, breastfed children will still become stunted if they do not receive an adequate quantity and quality of complementary foods from the age of six months onwards (Jones et al., 2008). Breastfeeding also protects against weight loss due to diarrhea (Bøhler et al., 1995), and in some studies children exclusively breastfed were shown to be less likely to be stunted (Engebretsen et al., 1968).

Breastfeeding contributes to maternal health in the immediate postpartum period by helping the uterus to contract rapidly, thereby reducing blood loss. In the short term breastfeeding delays a woman's return to fertility (WHO, 2004), and in the long term it reduces the risk of cancers of the breast and ovary (Harder et al., 2006). In many poor countries, feeding a child on breast milk eliminates expenditures on infant formula or other substitutes, which can be substantial.

2.3. Literatures Related with the Variable Used in the Study

Most studies reported that the socio demographic factors such as mother’s education level, place of residence, wealth index, contraceptive use, religion, marital status, age of mothers are known to be associated with duration of EBF.

Analysis models for variables associated with breastfeeding duration using Cox regression model by Edson et al. (2013) revealed that younger mothers are more likely to expose to early cessation of breastfeeding. Lower family income remained as a protective factor of breastfeeding.

Akter and Rahman (2010) use Cox PH model to examine the socioeconomic and demographic determinants of breastfeeding in Bangladesh. Their studies result revealed that mothers who were married at an early age had a longer duration of breastfeeding than those who were married in older age. The urban mothers breastfed their children for a relatively-shorter duration than did the rural mothers. Women who use contraceptive had a lower risk of stopping breastfeeding compared to women who hadn't use contraceptive. The Muslim mothers had higher risk of stopping breastfeeding than their non-Muslim peers. The risk of cessation of breastfeeding increased with increasing maternal education. Increasing maternal age can lead to breastfeeding of a shorter duration.

Gudina et al. (2013) investigate the predictors of non-exclusive breastfeeding using the multivariable logistic regression. According to their studies result Non–exclusive breastfeeding was more common among mothers with no marital relationships, poor access to health facilities, and inadequate knowledge about infant and young child feeding practices.

Chadasama et al. (2007) uses Life table and Kaplan Meier method in order to study factors associated with duration of exclusive breast feeding. The result revealed that the duration of EBF was shorter in urban, educated women, women from high income group and those who had delivered in health facility.

Perera et al. (2012) used analysis of variance in order to compare exclusive breast feeding between different groups. According to their studies result mothers starting to work and concerns regarding adequacy of breast milk were the major reasons to cease EBF.

Tampah-Naah and Kumi-Kyereme (2013) used logistic regression model to examine the association between selected covariates on duration of exclusive breast feeding. According to their study results marital status, region and place of delivery were found to be associated with the practice of exclusive breastfeeding. The study found that mothers who delivered at a government health facility had a higher probability to practice exclusive breastfeeding compared to mothers who delivered at home, or a private health facility.

Multivariable logistic regression analysis was used to identify independent predictors of exclusive breastfeeding by Setegn (2012). Their study results show that Employed mothers were less likely to practice exclusive breastfeeding.

Exclusive BF duration was associated with higher parental income and the prenatal decision to breastfeed. The duration of any breastfeeding was associated with the mothers' age of ≥30 years

and whether they were exclusively breastfeeding at discharge from the maternity unit according to the study done by Robert et al (2014).They used the cox proportional hazard model for data analysis. They also used Log-Rank and Breslow tests to assess the equality of the survival curves.

Alemayehu et al. (2009) report that maternal education, marital status, wealth index and age of the child were closely associated with EBF practices. They applied logistic regression models to fit the model for the data. Women who were no currently married were more likely to breastfeed their child exclusively than those married. Likewise women in the wealth index ranking middle and above were more likely to EBF than the poor.

2.4. Review of Survival Models

The origin of survival analysis goes back to the time when life tables were introduced. Life tables are one of the oldest statistical techniques and are extensively used by medical statisticians and by actuaries. Yet relatively little has been written about their formal statistical theory. Kaplan and Meier (1958) gave a comprehensive review of earlier work and many new results. Cox (1972) was largely concerned with the extension of the results of Kaplan and Meier to the comparison of life tables and more generally to the incorporation of regression like arguments into life table analysis.

Survival models have the capability of handling censored data. Cox (1972) and Cox and Oakes (1984) used survival analysis in modeling human lifetimes. Fergusson et al. (1984) used hazard functions to study the time to marital breakdown after the birth of child. Hazard functions had been also used in studies of time to shift in attentions in classroom Felmlee et al. (1983); in study of relapse of mental illness (Lavori et al., 1984), marital dissolutions Morgan et al. (1988) and human lifetimes Gross et al. (1975).

Proportional hazards modeling is the most frequently use type of the survival analysis modeling in many research areas, having been applied to topics such as smoking relapse (Stevens and Hollis, 1989), affective disorders childhood family breakdown interruptions in conversation (Dress, 1986), and employee turnover (Morita, et al., 1989), and in medical areas for identification of important covariates that have as significant impact on the response of the interested variables.

Cox (1972) introduced a semi parametric survival model. This model is based on the assumption that the survival times of distinct individuals are independent of each other. This assumption holds in many experimental settings and widely applicable. However; there are instances in which this assumption may be violated. For example, in many epidemiological studies, survival times are clustered into groups such as families or geographical units: some unmeasured /immeasurable characteristics shared by the members of that cluster, such as genetic information or common environmental exposures could influence time to the studied event. To account these factors, we should include the random effect terms in the standard Cox model (Clayton, 1978; Klein et al. 1992: Nielsen et al., 1992, Hastie and Tibshirani, 1993).

Frailty models are an extension of the Cox model that allows such dependence to be captured in an intuitive way. Frailty models are extensions of the PHs model which is best known as the Cox model (Cox, 1972), the most popular model in survival analysis. Frailty models are substantially promoted by its applications to multivariate survival data in a seminar paper by Clayton (1978) without using the notion frailty. Hougaard (1986) used several distributions for frailty including gamma, inverse Gaussian, positive stable distributions and claimed that these two distributions are relevant and mathematically tractable as a frailty distribution for heterogeneous populations. Flinn and Heckman (1982) used a lognormal distribution for frailty, whereas Vaupel et al. (1979) assumed that frailty is distributed across individuals as a gamma distribution. Recent research has addressed the problem of heterogeneity. Hougaard (1986) suggested the power variance function (PVF) distribution which includes gamma, inverse Gaussian, positive stable distributions as frailty model. Hedeker et al. (1996) discussed a frailty regression model for the analysis of correlated grouped time survival data. Frailty models have been applied to the analysis of event history data, including the study of age at time of death for individuals in terms of population (Zelterman, 1992), unemployment duration (McCall, 1994), pregnancy in women (Aalen, 1987) and migration (Lindstorm,1996).

Roberto et al. (2007) use proportional hazards model to examine Factors Associated with Duration of Breast Feeding. Edson et al. (2013) also use Cox PH model for the analyses of Analysis models for variables associated with breastfeeding duration. And Robert et al. (2012) used proportional hazard model to examine Breastfeeding Duration: A Survival Analysis: Data from a Regional Immunization Survey. Abiyot (2013) used Cox PH and shared gamma frailty

model to examine Modeling Time-to-Recovery of Adult Diabetic Patients: A Comparison of Cox-PH and Shared Frailty Models. Belay (2011) also used Cox PH and shared gamma frailty model to examine Modeling Time to Malaria: A Comparison of Cox PH and Shared Gamma Frailty Models.

3. DATA AND METHODOLOGY

3.1. Data Source

Secondary data for the analysis was obtained from the Ethiopia Demographic and Health Survey (EDHS) 2011. The survey was conducted by the Central Statistical Agency (CSA) under the auspices of the Ministry of Health. This was the third Demographic and Health Survey (DHS) conducted in Ethiopia, under the worldwide MEASURE DHS project, a USAID-funded project providing support and technical assistance in the implementation of population and health surveys in countries worldwide. The data set that we have used in this thesis was obtained from MEASURE DHS project.

3.2. Study Population

The 2011 EDHS sample was designed to provide estimates for the health and demographic variables of interest for the following domains: Ethiopia as a whole; urban and rural areas of Ethiopia (each as a separate domain); and 11 geographic areas (9 regions and 2 city administrations), namely: Tigray, Affar, Amhara, Oromiya, Somali, Benishangul-Gumuz; Southern Nations, Nationalities and Peoples (SNNP), Gambela, Harari, Addis Ababa and Dire Dawa. In general, a DHS sample is stratified, clustered and selected in two stages. In the 2011 EDHS a representative sample of approximately 14,500 households from 540 clusters was selected. The sample was selected in two stages. In the first stage, 540 clusters (145 urban and 395 rural) were selected from the list of enumeration areas (EA) from the 2007 Population and Housing Census sample frame (EDHS, 2011). Households comprised the second stage of sampling. A complete listing of households was carried out in each of the 624 selected EAs. A representative sample of 14,500 households was selected for the 2011 EDHS.

The 2011 EDHS used three questionnaires: the Household Questionnaire, the Woman's Questionnaire, and the Man's Questionnaire. The Woman's Questionnaire was used to collect information from all women aged 15-49 from the selected households. The data used for duration of EBF estimation were collected in the birth history section of the Woman's Questionnaire from 16,515 women aged 15-49. The background characteristics of the 14,070 women aged 15-49 was fully obtained in the 2011 EDHS. The study has used the birth history

data of the respondents (mothers) from Ethiopian DHS 2011 and the data are reported retrospectively. This study was only includes mothers whose last birth was not more than six months this enables to minimize recall bias. Because we are only interested in duration of EBF under six months of age infants, we only considered the data of the child age up to six month which includes 1371 children's information.

3.3. Variable Description

3.3.1. Dependent Variables

The duration of EBF for the last child of the respondent was considered as the dependent variable. The mother was asked about the introduction of any breast food or breast milk substitute of any nature in the child's diet: smashed fruits, cow milk, fortified food, or even industrialized foods. If yes, records were made of the month in which the child had started this eating habit and literally what food had been used during the previous 24 hours. Those women who were continuing EBF on the date of interview were considered censored cases and their duration of EBF was recorded and treated as censored data.

3.3.2. Independent Variables

For this thesis, socioeconomic and demographic variables, which contribute to early introduction of breast milk substitute or termination of EBF, were included as independent variable. The socioeconomic variables included are place of residence, work status of mothers, mothers' education level, economic status and place of delivery. The demographic variables included contraceptive-use, religion, marital status and age of mother. For this study the choice of variables was guided by different literatures sagest as the determinants of duration of EBF practices. Description for explanatory variables is given below.

Table 1: Description of explanatory variables

Variable	Explanation	Categories
Mother's age at birth (Age)	The age of the mothers at the time of the most recent birth. Coding is done in three cohorts: 15-19 years, 20-39 years and 40-49 years; representing adolescents/teenagers, young adults and older	0=15-19 1=20-39 2=40-49
Mother's education level	Educational status refers to the highest educational level the woman attained.	0=no education 1=primary 2=secondary & higher
Work status	In the survey, this was defined as if the woman has been currently working in any field other than household work. This was classified as working or not working.	1= working 2=not working
Religion	Classification of this variable was developed according to previous literature as, Catholic, orthodox, protestant, Muslim and others like traditional religion.	0= Muslim 1= Orthodox 2=Protestant 3=others
Residence	This is the type of place of residence where the respondent was lived as either urban or rural.	1=Rural 0=Urban
place of delivery	This refers to whether the infants deliver in health institutions or traditionally at home. Yes (deliver at health institution),No (delivered at home traditionally).	1=Yes 0=No
Household wealth index	This is the measures the standard of living of the family that the woman belongs. It is based on characteristics related to the socio-economic status of a household. We categorize poorest and poor as poor and rich and richest as rich, and middle as middle.	0=rich ,1=middle 2=poor
Marital Status	This is classified as married, never in union, others (separated, divorced and widowed) women based on the response from women's questionnaire.	Married =0 never in union=1 Others=2
Contraceptive use	This is classified as contraceptive user or not based on the response of respondents.	1=Not use 0=Use

3.4. Methods of Survival Analysis

3.4.1. Survival Analysis

The term "survival analysis" pertains to a statistical approach designed to take into account the amount of time an experimental unit contributes to a study. In other words, survival analysis is an important statistical technique used to describe and model time to event data. The purpose of survival analysis is to model the underlying distribution of the failure time variable and to assess the dependence of the failure time variable on covariates. The term survival analysis suggests that the event is death, but that is not necessarily so. Events could also denote success, such as recovery from therapy. Survival time then describes the time from a certain origin to the occurrence of an event.

Survival data are different from other types of continuous data because over the period of study the endpoint of interest is not necessarily observed in all subjects. This may occur because:

(a) Some patients are lost to follow-up, that is, they are not followed to the end of the study and, when last seen, have not experienced the event of interest, or

(b) The event has not occurred in some patients by the time the study ends for analysis. Such data are referred to as censored survival times and are different from missing data in that they provide a lower bound for the actual non-observed survival times. Any analysis carried out on survival data should use statistical methods that do not disregard censored data and, indeed, make the fullest possible use of it to avoid loss of information.

There are three common forms of censoring:

a. **Right Censoring:** The most common form of incomplete data is right censoring. A survival time is said to be right censored if it is recorded from its beginning until a well-defined time before its end time. It means a subject's follow-up time terminates before the outcome of interest is observed. For instance, if an HIV-1 patient is followed until he has a viral load high than 1000 copies/μl and is followed without experiencing this scenario until the end of the observation period. In other words, a survival time is said to be right censored if it begins at time t = 0 and terminates before the outcome of interest is observed.

b. **Left Censoring**: A survival time is said to be left censored if an individual developed the event of interest prior to the beginning of the study. This situation is less common in survival studies and is often not a focus.

c. **Interval Censoring:** A survival time is categorized as interval censored if it is only known that the event of interest occurs within an interval of time without the knowledge of when exactly it occurs. Interval censoring occurs in clinical trials where patients have periodic follow-ups and in industrial experiments where equipment items are inspected periodically, etc.

3.4.1.1. Survival Functions

For most statistical application it is usual to describe models for probability distribution in terms of either the probability density function f(x) or the distribution function F(x). For survival analysis it is usually more appropriate to work with other functions which characterize the probability distribution. Let T be a positive random variable from a homogeneous population, representing the time until the relevant event occurs. In order to characterize the distribution of T one of the most often used functions is survivor function. The survivor function, S (t), is defined for both discrete and continuous distribution as the probability that an individual survives beyond time t i.e., for continuous random variable T, the density function, f (t), is given by

$$S(t) = p(T > t) = 1 - \int_0^t f(u)\partial u , \qquad 0 < t < \infty \ldots\ldots\ldots\ldots\ldots\ldots (4)$$

$$S(t) = \int_t^{\infty} f(u)\partial u$$

Here $0 < s(t) < 1$ since S (0) =1 and $\lim_{t\to\infty} s(t) = 0$

The survivor function, S (t), is defined for discrete random variable T, the density function, this is unconditional probability of the events occurring at time t, f (t), is given by

$$S(t) = p(T > t) = \sum_{t_{j>t}} f(t_j)$$

Many types of survival curves can be shown but the important point to note is that they all have the same basic properties. They are monotone, non-increasing function equal to one at zero and zero as the time approaches infinity. Their rate of decline, of course, varies according the risk of experiencing the event at time t but it is difficult to determine the essence of a failure pattern by simply looking at survival curve. A basic quantity fundamental in survival analysis, which is more informative about the underlying mechanisms of failure than survival function is the hazard function.

3.4.1.2. Hazard Function

The hazard function describes the concept of the risk of an outcome (e.g., death, failure, hospitalization) 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 The hazard function seems to be more intuitive to use in survival analysis than the PDF because it quantifies the instantaneous risk that an event will take place at time t given that the subject survived to time t. The hazard function h(t), is defined as the instantaneous potential of failure at time t given survival to t i.e.,

$$h(t) = \lim_{\Delta t \to \infty} \frac{p(t \le T < t + \Delta t / T \ge t)}{t} \qquad (5)$$

This is a positive measure an sometimes referred to as the conditional failure rate. Following the fundamental theorem of calculus, it can be seen that (4) can be written as

$$f(t) = \frac{\partial F(t)}{dt} = \lim_{\Delta t \to \infty} \frac{F(t + \Delta t) - F(t)}{\Delta t} = \lim_{\Delta t \to \infty} \frac{p(T < t + \Delta t) - p(T < t)}{\Delta t}$$

$$= \lim_{\Delta t \to \infty} \frac{p(t < T < t + \Delta t)}{\Delta t} \qquad (6)$$

Using 4.4 and the definition of conditional probability, the hazard 5 can be written as

$$h(t) = \frac{f(t)}{s(t)} \qquad (7)$$

And from (3.2) it follows that

$$h(t) = -\left[\frac{\frac{dS(t)}{dt}}{S(t)}\right]$$

$$S(t) = \exp(-\int_0^t h(u)du) \qquad (8)$$

The quantity $H(t) = \int_t^\infty h(u)du$ is known as the cumulative hazard function.it also given for discrete time as $H(t) = \sum_{t=0}^{\infty} h(t))$.

From equation (4) one can see that h(t)Δt may be viewed as the approximate probability of an individual of age t experiencing the event in the next instant. It follows that h (t) is a positive function, not necessarily increasing or decreasing. This function is particularly useful in determining the appropriate failure distribution utilizing qualitative information about the mechanism of failure and for describing the way in which the chance of experiencing the event changes with times.

There are many general shapes for the hazard rate, the only restriction on h(t) is that it be non-negative. The hazard rate for the occurrence of a particular event can be increasing, decreasing, constant, bathtub-shaped, hump-shaped, or possessing some other characteristic which describe the failure mechanism. The hazard function, h(t), is defined as the instantaneous potential of failure at time t given survival to t .

If T is a discrete random variable different techniques are required. Suppose that T can take on values t_j, j= 1, 2, 3 . . . with probability mass function (p.m.f):

$f(t_j) = p(T = t_j)$, j=1, 2….where $t_1 < t_2 < .....$

The survival function for a discrete random variable T is given by

$$S(t) = p(T > t) = \sum_{t_{j>t}} f(t_j) \qquad (9)$$

Likewise, the hazard function is given by

$$h(t_j) = p(T = t_j / T \geq t_j) = \frac{f(t_j)}{S(t_{j-1})}, j = 1,2 \quad (10)$$

where $S(t_0) = 1$ because $S(t_{j-1}) - S(t_j)$, in conjunction with (4)

$$h(t_j) = 1 - \frac{S(t_j)}{S(t_{j-1})} \quad (11)$$

It is straight forward to define $p(T = t)$ and $p(T \geq t)$ in terms of the hazard function by considering that $1 - h(t_j)$ is the conditional probability of survival at t_j given survival time to t_j .So for ordered survival time $t_1 < t_2 < .....t_n$,

$$h(t_j) = h(t_i)\prod_{j=1}^{i-1}(1 - h(t_j))$$

And (12)

$$P(T \geq t) = \prod_{j \backslash t_j \leq t}(1 - h(t_j))$$

All of the above basic quantities for a univariate case can be easily generalized to multivariate case. The multivariate survival function of the life times (T_1, . . . , T_k)

$$S(t_1, t_2 ...., t_k) = p(T_1 > t_1, ..., T_k > t_{k)} = \exp(H(t_1, t_2 ..., t_k)) \quad (13)$$

3.4.2. Non-parametric Methods

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. The estimation of the survival distribution provides estimates of descriptive statistics such as the median survival time. These methods are said to be non-parametric methods since they require no assumptions about the distribution of

survival time. In order to compare the survival distribution of two or more groups, log-rank tests can be used.

3.4.2.1. The Kaplan-Meier (KM) Estimator

The Kaplan-Meier (KM) estimator is the standard non parametric estimator of the survival function S (t), proposed by Kaplan and Meier (1958). It is also called the Product-Limit estimator. KM estimator incorporates information from all of the observations available, both censored and uncensored, by considering any point in time as a series of steps defined by the observed survival and censored times. When there is no censoring, the estimator is simply the sample proportion of observations with event times greater than t. The technique becomes a little more complicated but still manageable when censored times are included. Let rank-ordered survival times are given by $0 \le t_1 < t_2 < ... < t_j \le \infty$, then

$$\hat{S} = \begin{cases} 1 & if\ t < t_1 \\ \prod_{j:t(j)\le t}\left[1-\frac{d_j}{r_j}\right] & \text{if } t \ge t_1 \end{cases} \qquad (14)$$

Where d_j is the observed number of events at time $t(j)$ and r_j is the number at risk at time $t(j)$ (i.e., the number of individuals who are alive at time $t(j)$ or experience the event of interest at a time $t(j)$.

The Kaplan-Meier estimator $\hat{S}(t)$ is a step function with jumps at the observed event times. The size of the jump at a certain event time $t(j)$ depends on the number of events observed at $t(j)$, as well as on the pattern of the censored event times before $t(j)$. The variance of the Product-Limit estimator is estimated by Greenwood's formula (Greenwood, 1926), and is given by;

$$\hat{V}(\hat{s}(t)) = [\hat{s}(t))]^2 \sum_{j:t_{(j)}\le t} \frac{d_j}{r_j(r_j - d_j)}; j = 1,2,...,r \qquad (15)$$

Since the distribution of survival time tends to be positively skewed, the median is preferred for a summary measure. The median survival time is the time beyond which 50% of the individuals

under study are expected to survive, i.e., the value of $t(50)$ at $\hat{S}(t(50)) = 0.5$ .The estimated median survival time is given by $\hat{t}(50) = \min\{t_i \mid \hat{s}(t_i) < 0.5)$ where t_i is the observed survival time for the i^{th} individual, i =1, 2... n. In general, the estimate of the p^{th} percentile is:-

$$\hat{t}(p) = \min\{t_i \mid \hat{s}(t_i) < 1 - \frac{p}{100}) \tag{16}$$

The variance of the estimator of the p[th] percentile is

$$\text{var}(\hat{s}(t)) = [\frac{\partial \hat{s}(t(p))}{\partial(t(p))}]^2 \text{var}(t(p)) = (-f(t(p)))^2 v(t(p) \text{var}(t(p))$$

The standard error of $\hat{t}(p)$ is therefore given by

$$SE(\hat{t}(p)) = \frac{1}{\hat{f}(t(p))} SE[\hat{S}(t(p))] \tag{17}$$

The standard error of $\hat{S}(t(p))$ can be obtained by using Greenwood's formula, given in equation (15). An estimate of the probability density function at the p[th] percentile b (p) is used by many software packages

$$[\hat{f}(\hat{t}(p))] = \frac{\hat{S}(\hat{u}(p)) - \hat{S}(\hat{l}(p))}{\hat{l}(p) - \hat{u}(p)}$$

Where

$$\hat{u}(p) = \max\{S(t_j) \geq 1 - \frac{p}{100} + \vartheta\}$$

$$\hat{l}(p) = \min\{S(t_j) \leq 1 - \frac{p}{100} - \vartheta\}$$

Where t(j) is j[th] ordered death time, j = 1, 2….r: $\vartheta = 0:05$ is typically used by a number of statistical packages. Therefore, for median survival time, $\hat{l}(50)$ is the largest observed survival

time from the K-M curve for which $\hat{S}(t(p)) \geq 0.55$ and $\hat{l}(50)$ is the smallest observed survival time from the K-M curve for which $\hat{S}(t(p)) \leq 0.45$.

The 95% confidence interval for the p[th] percentile $\hat{t}(p)$ has limits of $\hat{t}(p) \pm 1.96 * SE[\hat{t}(p)]$.

3.4.2.2. Non parametric Comparison of Survival Distributions

The simplest way of comparing the survival times obtained from two or more groups is to plot the Kaplan-Meier curves for these groups on the same graph. However, this graph does not allow us to say, with any confidence, whether or not there is a real difference between the groups. The observed difference may be a true difference, but equally, it could also be due merely to chance variation. Assessing whether or not there is a real difference between groups can only be done, with any degree of confidence, by utilizing statistical tests. Among the various non-parametric tests one can find in the statistical literature, the Mantel-Haenzel (1959) test, currently called the "log-rank" is the one commonly used non-parametric tests for comparison of two or more survival distributions. The log rank test statistic for comparing two groups is given by:

$$Q = \frac{[\sum_{i=1}^{m} wi(d_{1i} - \hat{e}_{1i})]^2}{\sum_{i=1}^{m} w_i^2 \hat{v}_{1i}} \sim \chi^2_{k-1} \qquad (18)$$

$\hat{e}_{1i} = \frac{n_{1i} d_i}{n_i}$ and $\hat{v}_{1i} = \frac{n_{1i} n_{0i} d_i (n_{1i} - d_i)}{n_i^2 (n_i - 1)}$

Where,

n_{0i} is the number at risk at observed survival time $t_{(i)}$ in group 0

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

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

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

d_i is the total number of event at $t_{(i)}$

K is number of groups in each category.

3.4.3. Cox PH Model

The non-parametric method does not control for covariates and it requires categorical predictors. If the groups are similar, except for the treatment under study, then the nonparametric methods can be used directly. When we have several predictor variables, we must use multivariable approaches. But we cannot use multiple linear regression or logistic regression because they cannot deal with censored observations. We need another method to model survival data with the presence of censoring. One very popular model in survival data is the Cox PHs model, which is proposed by Cox (1972).

Let T denote the time to some event. Our data, based on a sample of size n, consists of the triple $(T_j, \sigma_j, Z_j(t)), j = 1,2,...,n$ where T_j is the time on study for the j^{th} subject σ_j is the event indicator for the j^{th} subject (σ_j =1 if the event has occurred and σ_j =0 if the lifetime is censored) and $Z_j(t) = [z_{j1}(t),..., z_{jp}(t)]^t$ is the vector of covariates or risk factors for the j^{th} individual at time t which may affect the survival distribution of T. Here the $Z_{ji}(t)$'s, k=1. . . P, may be time-dependent covariates whose value changes over time, or they may be constant (or fixed) values known at time 0. Then, the Cox PHs model is given by

$$h(t \mid Z) = h_0(t \mid Z)\exp(\beta^t Z) \qquad (19)$$

where $h_0(t)$ is called the baseline hazard function, which is the hazard function for an individual for whom all the variables included in the model are zero, Z is the values of the vector of explanatory variables for a particular individual, and β is a vector of regression coefficient $S(t, Z, \beta) = [S_0(t(p)]^{(\beta^t z)}$ where, $S_0(t)$ the baseline survival function.

This model, also known as the Cox regression model, makes no assumptions about the form of $h_0(t)$ (non -parametric part of model) but assumes parametric form for the effect of the predictors on the hazard (parametric part of model). The model is therefore referred to as a semi-parametric model. However, Cox regression model is valid under the assumption of proportional hazards. Cox observed that if proportional hazards assumption holds, then it is possible to estimate the effect parameter(s) without any consideration of the hazard function. The proportional hazards

assumption refers to the fact that the hazard functions are multiplicatively related. That is, their ratio is assumed constant over survival time. In other words, the Cox proportional hazards model assumes that changes in the hazard of any subject over time will always be proportional to changes in the hazard of any other subject and to changes in the underlying hazard over time.

The beauty of the Cox approach is that this vagueness creates no problems for estimation. Even though the baseline hazard is not specified, we can still get a good estimate for regression coefficients (β), hazard ratio, and adjusted hazard curves.

From the representation in equation (19) one can notice a couple of features. First, if $Z_j = 0$ then the hazards function for the j^{th} individual is the baseline hazard function. It's the hazard function in the absence of covariates or when all of the coefficients of the covariates are assumed to be zero. Second, if we divide both sides by $h_0(t)$, we get equation (20) which shows where the term proportional comes from. Since for each individual $\exp(\beta^t Z)$ is constant across time, equation (16) shows that at every value of t, the j^{th} individual's log hazard function is constant proportion of the baseline hazard. Very loosely speaking, this implies that each individual's hazard function is "parallel" to the $h_0(t)$.

$$\frac{h(t,\beta^t Z)}{h_0(t)} = \frac{h_0(t)\exp(\beta^t Z)}{h_0(t)} = \exp(\beta^t Z) \qquad (20)$$

This is called a semi-parametric model because a parametric form is assumed only for the covariate effect and the baseline hazard rate is treated non-parametrically. If we look at two individuals with covariate values Z and Z^*, the ratio of their hazard rates is:-

$$\frac{h(t \mid Z)}{h(t \mid Z^*)} = \frac{h_0(t)\exp[\sum_{k=1}^{p}(\beta_k Z_k)]}{h_0(t)\exp[\sum_{k=1}^{p}(\beta_k Z_k^*)} = \exp[\sum_{k=1}^{p}\beta_k (Z_k - Z_k^*)] \qquad (21)$$

This is constant with respect to time. So, the hazard rates are proportional. The quantity given in equation (21) is called the relative risk (hazard ratio) of an individual with risk factor Z having the event as compared to an individual with risk factor Z*.

The Cox proportional hazards model can equally be regarded as linear model, as a linear combination of the covariates for the logarithm transformation of the hazard ratio given by:

$$\log\{\frac{h(t,Z)}{h_o(t)}\} = \beta^t Z$$

Where $Z = (z_1, z_2 ...., z_p)^t$ is the value of the vector of explanatory variables for a particular individual and $\beta = (\beta_1, \beta_{2,....,} \beta_p)^t$ is a vector of coefficients.

Again the cumulative hazard function is given by $H(t) = H_0(t)\exp(\beta^t Z)$; the corresponding survival functions for fitting the Cox Proportional Hazards Model related as $S(t, Z, \beta) = \{S_0(t)\}\exp(\beta^t Z)$.

3.4.3.1. Fitting the Cox PH Model

The data in survival analysis based on the sample size n are denoted by the triplet (t_i, δ_i, x_i) , i=1, 2… n, where t_i is the time at which the i^{th} individual dies from the disease of interest, δ_i is the event indicator $\delta_i = 1$ if the event has occurred and $\delta_i = 0$ if it is censored (the lifetime may be right, left or interval censored) ,and x_i is the vector of covariates or the risk factors for the i^{th} individual.

The Cox model will be fitted by estimating the unknown regression coefficients through the maximum likelihood method. The actual likelihood function is constructed by considering the contribution of the probability that a subject with covariate value x dies from the disease of interest at time t (i.e., f (t, β, x)), and the probability that a subject with covariate value x survives at least t time units (i .e, S (t, β, x)). That is, under the assumption of independent observations, the full likelihood function is obtained by multiplying the respective contributions of the observed triplets, a value of f (t, β, x) for a non-censored observation and a value of S (t, β, x) for censored observations.

Thus, the contribution of each triplet to the likelihood is the expression

$$[f(t,\beta,x)]^{\delta_i} \times [S(t,\beta,x)]^{1-\delta_i} \quad (22)$$

Since the observations are assumed to be independent, the likelihood function is the product of the expression in (22) over the entire sample and is formulated as:

$$l(\beta) = \prod_{i=1}^{n} \{[f(t_i, x_i, \beta)]^{\delta_i} \times [S(t_i, x_i, \beta]^{1-\delta_i}\} \quad (23)$$

It can be further simplified as

$$l(\beta) = \prod_{i=1}^{n} \{([h(t_i, x_i, \beta)) \times (S(t_i, x_i, \beta))]^{\delta_i} \times [S(t_i, x_i, \beta]^{1-\delta_i}\}$$

$$= \prod_{i=1}^{n} \{[h(t_i, x_i, \beta)]^{\delta_i} \times [S(t_i, x_i, \beta]\} \quad (24)$$

Cox (1972) proposed using an expression he called a partial likelihood function due to the fact that the likelihood formula considers probabilities only for those subjects who fail, and does not explicitly consider probabilities for those subjects who are censored. In other words, the likelihood for the Cox model does not consider probabilities for all subjects. Let us consider a sample of n subjects and suppose a total of m failures occur, with m smaller than n, due to the presence of censoring. Let $t_1 < t_2 < ... < t_m$ be the m distinct ordered failure times observed and let $R(t_i)$ be the set of individuals at i^{th} failure time, which consists of all subjects with survival or censored times greater than or equal to the specified time (Hosmer and Lemeshow, 1999). Thus the partial likelihood is given by the expression:

$$l_p(\beta) = \prod_{i=1}^{n} \left[\frac{e^{x_i \beta}}{\sum_{j \subseteq R(t_i)} e^{x_j \beta}} \right]^{\delta_i} \quad (25)$$

The expression assumes that there are no tied times, and designed in such a way that it excluded terms when $\delta_i = 0$ .As a result the equation in (25) becomes,

$$l_p(\beta) = \prod_{i=1}^{n} \left[\frac{e^{x_i \beta}}{\sum_{j \subseteq R(t_i)} e^{x_j \beta}} \right] \tag{26}$$

To obtain the maximized likelihood with respect to the parameters of interest, β, we maximize the log partial likelihood function as

$$L_p(\beta) = \sum_{j=1}^{m} \left\{ x_i \beta - \ln \left[\sum_{j \subseteq R(t_j)} e^{x_j \beta} \right] \right\} \tag{27}$$

We obtain the maximum partial likelihood estimator by differentiating the right hand side of (27) with respect to β, setting the derivatives equal to zero and solving for the unknown parameters. This is known as the Newton-Raphson iterative method. That is, for each derivative

$$U(\beta) = \frac{L_p(\beta)}{\partial \beta} = \sum_{j=1}^{m} \left\{ \beta - \ln \left[\sum_{j \subseteq R(t_j)} e^{x_j \beta} \right] \right\} = \sum_{j=1}^{m} \left\{ x_{(i)} - \sum_{j \subseteq R(t_{(j)})} w_{ij}(\beta) x_j \right\} = \sum_{1}^{m} \left\{ x_{(i)} - \bar{x}_{wi} \right\} = 0 \tag{28}$$

Where $w_{ij}(\beta) = \dfrac{e^{x_j \beta}}{\sum_{j \subseteq R(t_{(j)})} e^{x_j \beta}}$ and $\bar{x}_{wi} = \sum_{j \subseteq R(t_{(j)})} w_{ij}(\beta) x_j$

U (β) is called the score or gradient vector. The solution to the equation (29) is denoted by $\hat{\beta}$ .The estimator of the variance of the estimator of the coefficient is obtained in the same manner as variance estimators are obtained in most maximum likelihood estimation applications. The estimator is the inverse of the negative of the second derivative of the log partial likelihood at the value of the estimator. Derivation of the expression in (30), will result in

$$\frac{L_p(\beta)}{\partial \beta^2} = -\sum_{j=1}^{m} \left\{ \left[\sum_{j \subseteq R(t_j)} e^{x_j \beta} \right] \left[\sum_{j \subseteq R(t_{(j)})} x_j^2 e^{x_j \beta} \right] - \left[\sum_{j \subseteq R(t_{(j)})} x_j e^{x_j \beta} \right] \times \frac{1}{\left[\sum_{j \subseteq R(t_{(j)})} e^{x_j \beta} \right]^2} \right\} \tag{31}$$

The expression in (31) shall be simplified using $w_{ij}(\beta)$ in equation (24) above.

That is,

$$\frac{\partial^2 L_p(\beta)}{\partial \beta^2} = -\sum_{i=1}^{m} \sum_{j \subseteq R(t_{(j)})} w_{ij}(x_j - x_{wi})^2 \quad (32)$$

The negative of the 2nd derivative of the log partial likelihood in either (31) or (32) is known as the observed information and denoted by

$$I(\beta) = -\frac{\partial^2 L_p(\beta)}{\partial \beta^2} \quad (33)$$

If we consider models that contain more than one covariate, the result in (33) becomes $I(\beta) = -\frac{\partial^2 L_p(\beta)}{\partial \beta \ \partial \beta^t}$ which is known as the observed information matrix (Hessian matrix).

According to the Newton-Raphson procedure an estimate of β at the (j+1)th of the iterative procedure, $\hat{\beta}_{j+1}$, is $\hat{\beta}_{j+1} = \hat{\beta}_j + I^{-1}(\hat{\beta}_j)U(\hat{\beta}_j), j = 0,1,2,...$ (34)

As a result, the estimator of the variance of the estimated coefficient is the inverse of (32) evaluated at $\hat{\beta}$ and is

$$\hat{Var}(\hat{\beta}) = \hat{\beta}_j + I^{-1}(\hat{\beta}_j) \quad (35)$$

After fitting the regression model, we go for assessing the significance of the coefficient and the construction of the confidence interval as well. The three different tests used to assess the significance of the coefficient are explained below (Hosmer and Lemeshow, 1999).

i) The partial likelihood ratio test

The partial likelihood ratio test, G, is computed as twice the difference between the log partial likelihood of the model containing covariates and the log partial likelihood of the model not containing the covariates. Mathematically,

$$G = 2\{L_p(\hat{\beta}) - L_p(0)\} \quad (36)$$

Where ` $L_p(0) = -\sum_{i=1}^{m} \ln(n_i)$ (37)

And n_i denotes the number of subjects in the risk set at observed survival time t(i)

Under the null hypothesis, that the coefficients are equal to zero, that is $H_0 : \beta = \overline{0} = (0,0,....,0)^t$, the statistic, G, follows a chi -square distribution with one degree-of-freedom for "sufficiently" large sample size.

ii) The Wald test

The Wald statistic $(Z = \frac{\hat{\beta}}{\hat{SE}(\hat{\beta})})$ is defined as the ratio of the estimated coefficient to its estimated standard error. Under the null that is a single parameter β_i, $H_0 : \beta_i = 0$, the Wald statistic follows a standard normal distribution (i.e., N (0, 1) ~ Z). Obviously, the square of Wald statistic follows a chi -square distribution with one degree-of-freedom.

Similarly, a 100(1-α) % Wald-statistic-based confidence interval for β will be $\hat{\beta} \pm Z_{\alpha/2}\hat{SE}(\hat{\beta})$, $Z_{\alpha/2}$ is the upper α/2 percentile point of the standard normal distribution.

iii) The score test

The score test which is obtained by computing the ratio of the derivative of the log partial likelihood in (24) to the square root of the observed information in (27), all evaluated at β=0.

Thus, the equation will be:

$$z^* = \left. \frac{\partial L_p / \partial \beta}{\sqrt{I(\beta)}} \right|_{\beta=0} \quad (38)$$

Under the null hypothesis that each parameter is equal to zero, $H_0 : \beta_i = 0$, this statistic follows a standard normal distribution. The score test may be reported as the squared by some statistical packages that will follow a chi-square distribution with 1 degree-of-freedom under the null

hypothesis. When there is a disagreement among the three tests of the significance of the coefficient, the partial likelihood ratio test will prevail.

3.4.4. Model Development

The methods of selecting a subset of covariates in a PHs regression model are essentially similar to those used in any other regression models. The most common methods are purposeful selection, step-wise (forward selection and backward elimination) and best sub-set selections. Survival analysis using Cox regression method begins with a thorough univariable analysis of the association between survival time and all important covariates (Hosmer and Lemeshow, 1999).

Recommendable procedure in selecting variables in the study

According to Hosmer and Lemeshow (1999) and Collett (2003) it is recommended to follow the steps given below.

1. Include all variables that are significant in the univariable analysis at the 20 to 25 percent level and also any other variables which are presumed to be clinically important to fit the initial multivariable model.

2. The variables that appear to be important from step one are then fitted together in a model. In the presence of certain variables others may cease to be important. As a result, backward elimination is used to omit non-significant variables (i.e., those variables that do not significantly increase the value of $-2\log\hat{L}$) from the model. Once a variable has been dropped, the effect of omitting each of the remaining variables in turn should be examined. Here, $\hat{L}$ denotes the maximized likelihood under an assumed model and computed from the partial likelihood equation (20) by replacing the β's by their maximum likelihood estimates under the model.

3. Variables, that were not important on their own, and so were not under consideration in step 2, may become important in the presence of others. These variables are therefore added to the model from step 2, with forward selection method (i.e., any that reduce $-2\log\hat{L}$ significantly are retained in the model). This process may result in terms in the model determined at step 2 ceasing to be significant.

4. A final check is made to ensure that no term in the model can be omitted without significantly increasing the value of $-2\log\hat{L}$, and that no term not included significantly reduces $-2\log\hat{L}$.

3.4.4.1. Model Selection

To select the model that can predict the survival time of duration of EBF, we used Akaikie information criterion (AIC). Akaikie (1974) proposed an informative criterion (AIC) statistic to compare different models and/or models with different numbers. For each model the value is computed as:

AIC = -2 x log(maximum likelihood) + kx p;

where p is the number of parameters in each model under consideration and k a predetermined constant. This statistic is called Akaike's (1974) information criterion (AIC); the smaller the value of this statistic, the better the model.

We can rewrite the AIC to address parametric regression models considered in the text. For the parametric models discussed, the AIC is given by

AIC = -2 x log(maximum likelihood) + 2 x (a + b)

where a is the number of parameters in the specific model and b the number of one-dimensional covariates. For example, a = 1 for the exponential model, a = 2 for the gamma, inverse Gaussian, and log-normal models.

3.4.5. Assessment of Model Adequacy

Model-based inferences depend completely on the fitted statistical model. For these inferences to be valid in any sense of the word, the fitted model must provide an adequate summary of the data upon which it is based. Some of the methods for the assessment of a fitted proportional hazards model can be equally used for parametric regression models with the exception that assessing the adequacy of survival models have to cope with the occurrence of censored survival times.

3.4.5.1. Residual analysis

Many model checking procedures are based on quantities known as residuals. A residual is the difference between the observed value of the outcome variable and that value predicted by the

model. The two key assumptions in the definition of a residual are the value of the outcome is known and the fitted model provides an estimate of the mean of the dependent variable or systematic component of the model. However, the two assumptions are not valid when using partial likelihood to fit the proportional hazards model to censored survival data. The absence of an obvious residual has led to the development of several different residuals, each of which plays an important role in examining some aspect of the fit of the proportional hazard model. These include the Cox-Snell, martingale and Schoenfeld residuals (Collett, 2003 and Hosmer&Lemeshow, 1999).

Cox-Snell residuals (rc_i) are residuals most widely used in the analysis of survival data. The Cox-Snell residual for the i^{th} subject is given by

$$rc_i = \hat{H}_i(t) = -\hat{S}_i(t) \tag{39}$$

Where $\hat{H}_i(t)$ and $\hat{S}_i(t)$ are the estimated values of the cumulative hazard and survivor functions of the i^{th} subject at time t, respectively.

Schoenfeld residuals (rs_{ik})

Schoenfeld (1982) proposed residuals for use with a fitted proportional hazards model and packages providing them refer to as the "Schoenfeld residuals", which are based on the individual contributions to the derivative of the log partial likelihood. It is obtained by taking the first derivative of the log of the partial likelihood function for the k^{th} covariate as follows:

$$\frac{L_p(\beta)}{\partial \beta_k} = \sum_{i=1}^{n} \delta_i \left\{ x_{(ik)} - \frac{\sum x_{jk} e^{x_j^t \beta}}{\sum_{j \subseteq R(t_{(i)})} e^{x_j^t \beta}} \right\} = \sum_{i=1}^{n} \left\{ x_{(ik)} - \bar{x}_{w_i k} \right\} = 0 \tag{40}$$

Where $\bar{x}_{w_i k} = \dfrac{\sum x_{jk} e^{x_j^t \beta}}{\sum_{j \subseteq R(t_i)} e^{x_t^t} \beta}$ (41)

The estimator of the Schoenfeld residual for the i[th] individual on the k[th] covariate is obtained from (41) by substituting the partial likelihood estimator of the coefficient, $\hat{\beta}$, and is

$$\hat{r}s_{ik} = \delta_i (x_{ik} - \bar{\hat{x}}_{w_i k}) \qquad (42)$$

Wheres $\bar{\hat{x}}_{w_i k} = \dfrac{\sum_{j \subseteq R(t_{(i)})} x_{jk} e^{x_j^t \beta}}{\sum_{j \subseteq R(t_i)} e^{x_t^t \beta}}$ is the estimator of the risk set conditional mean of the covariate.

Since the partial likelihood estimator of the coefficient, $\hat{\beta}$, is the solution to the equations obtained by setting (40) equal to zero, the sum of the Schoenfeld residuals is zero.

It is suggested that (Grambsch and Therneau, 1994) scaling the Schoenfeld residuals by an estimator of its variance yields a residual with greater diagnostic power than the un scaled one.

Let the vector of p Schoenfeld residuals for the i[th] subject be denoted as

$$\hat{r}s_i^t = \hat{r}s_{i1}, \hat{r}s_{i2}, ..., \hat{r}s_{ip}$$

Where $\hat{r}s_{ik}$ is the estimator in (43), with the convention that $\hat{r}s_{ik}$ =missing if $\delta_i = 0$.

Thus, the vector of scaled Schoenfeld residuals is given as the product of the inverse of the covariance matrix and the vector of residuals;

$$\hat{r}_i^* = [\hat{V}ar(\hat{r}_i)]^{-1} \hat{r}_i \qquad (43)$$

Where $\hat{V}ar(\hat{r}_i)$ is the estimator of the p×p covariance matrix of the vector of residuals for the i[th] subject.

However, Grambsch and Therneau (1994) suggest based on their experience that the matrix, $\hat{V}ar(\hat{r}_i)$, tends to be fairly constant, the use of an easily computed approximation for the scaled Schoenfeld residuals. If this matrix is constant, its inverse may be approximated by multiplying the estimator of the covariance matrix of the estimated coefficients by the number of events (in our study numbers of mothers who introduce breast milk substitutes for their babies' m).That is:

$$[\hat{V}ar(\hat{r}_i)]^{-1} = m\hat{V}ar(\hat{\beta})$$

Consequently, the approximate scaled Schoenfeld residuals are obtained by substitution as

$$\hat{r}_i^* = m\hat{V}ar(\hat{\beta})\hat{r}_i \quad (44)$$

Each of these residuals provides a useful tool for examining one or more aspects of model adequacy.

A. Identification of influential subjects

Another important aspect of model evaluation is a thorough examination of the regression diagnostic statistics to identify which, if any, subjects have an unusual configuration of covariates, exert an illegitimate influence on the estimates of the parameters or have an undue influence on the fit of the model. Such observations may be termed as influential (aberrant) observations and the data from such individuals will need to be the subject of further scrutiny.

Conclusions from survival analyses are often framed in terms of estimates of the relative hazard, which depends on the estimated values of the coefficients in the Cox regression model. For that reason, it has particular importance to examine the influence of each observation on these estimates (Hosmer and Lemeshow, 1999).

In many occasions, the influence that each observation has on the estimated hazard function will be of interest, and it will then be important to identify observations that influence the complete set of parameter estimates in the model. In other words, it may happen that the structure of the fitted model is particularly sensitive to one or more observations in the data set. Such observations can be analyzed through diagnostics that are designed to highlight observations that influence the complete set of parameter estimates in the linear predictor. This could be done by fitting the model to all n observations in the data set, and then fitting the same model to the sets of n-1 observations obtained by omitting each of the n observations in turn. To achieve this purpose, to examine influence in the proportional hazards setting, we need to use statistics analogous to Cook's distance in linear regression. This is denoted as

$$\Delta\hat{\beta}_{ki} \approx \hat{\beta}_k - \hat{\beta}_{k(-i)} \quad (45)$$

Where $\hat{\beta}_k$ denotes the partial likelihood estimator of the coefficient computed using the entire sample of size n and $\hat{\beta}_{k(-i)}$ denotes the value of the estimator if the ith subject is removed. Cain and Lange (1984) show that an approximate estimator of (45) is the kth element of the vector of coefficient changes.

$$\Delta\hat{\beta}_i \approx (\hat{\beta} - \hat{\beta}_{(-i)}) = \hat{Var}(\hat{\beta})\hat{L}_i \qquad (46)$$

Where $\hat{L}_i$ is the vector of score residuals and $\hat{Var}(\hat{\beta})$ is the estimator of the covariance matrix of the estimated coefficients.

These are commonly referred to as the scaled score residuals and their values may be obtained from some software packages, for instance, SAS (Collett, 2003 and Hosmer and Lemeshow, 1999).

B. Methods for Assessing the PHs Assumption

The proportional hazards assumption is vital to the interpretation and use of a fitted proportional hazards model. However, there are various grounds for which the model may not have proportional hazards (or constant hazard ratio over time). If hazards are not proportional, this means that the linear component of the fitted model varies with time in some manner. As a result, we need to plot the logarithm of the Kaplan-Meier cumulative hazards function based on different factors so that it helps in assessing the proportional hazards assumption before fitting a Cox model (Collett, 2003 and Hosmer and Lemeshow, 1999). Obviously, if the assumption of proportional hazards is met, the plots should be parallel (or the two curves are equidistant over time).

There are a number of ways in which the proportional hazards model can be changed to non-proportional hazards functions or log-hazard functions that are not equidistant. For instance, if the j^{th} time-independent variable is denoted as, x_j then we can define the j^{th} product term as $x_j \times g_j(t)$ where $g_j(t)$ is some function of time for the j^{th} variable. Likewise, Grambsch and Therneau (1994) also considered a specific form of time-varying coefficient:

$$\beta_j(t) = \beta_j + \gamma_j g_j(t) \quad (47)$$

Where $g_j(t)$ is a specific function of time and γ_j is the coefficient of the same.

Thus, the extended Cox model that simultaneously considers all time-independent variables of interest can be formulated as:

$$h(t, x, \beta) = h_0(t)\exp(\sum_{1}^{p}\beta_j x_j + \sum \gamma_j x_j g_j(t)) \quad (48)$$

In order to check the proportional hazards assumption, we consider the null hypothesis that all the γ terms are equal to zero so that the model reduces to the proportional hazards model. The hypothesis all γ_j's are zero $(H_0 : \gamma_j = 0)$ is tested via the partial likelihood ratio test, score test or Wald test. Furthermore, the plot of scaled Schoenfeld residuals of each covariate versus the logarithm of analysis time (i.e., the time variable in survival analysis) used to confirm whether there is some departure from proportional hazards or not.

C. Overall Goodness of Fit

A number of plots based on residuals can be used in the graphical assessment of the adequacy of a fitted model. For instance, if the fitted model is correct, the Cox-Snell residuals were shown to have an exponential distribution with unit mean. Putting it in other words plots of these residuals against the survival times, the rank order of the survival times, or explanatory variables may indicate whether there are particular survival times, or values of the explanatory variables, where the model does not fit well (Collett, 2003). In addition, it is possible to use some measure analogous to standard R as the case of all regression analysis as a measure of model performance. Suppose we have G groups, the score test in proportional hazards model requires an introduction of G-1 dummy (design) variables, and then fitting proportional hazards model including the G-1 dummy variables. The log of partial likelihood for the fitted model with p covariates is identified as L_p and the log partial likelihood for model zero, the model with no covariates, as L_0 and n number of subjects. Thus, the measure of goodness of fit based on partial likelihood is given by:-

$$R_p^2 = 1 - \left\{ \exp\left[\frac{2}{n}\left(L_0 - L_p\right) \right] \right\}, \tag{49}$$

3.4.6. Shared Frailty Models

3.4.6.1. Definition of Shared Frailty Model

The shared frailty concept is one of the important approaches in multivariate survival modeling and is relevant to event times of related individuals, similar organs and repeated measurements. Individuals in a cluster are assumed to share the same frailty U that is why this model is known as shared frailty model. It was introduced by Clayton (1978) and extensively studied in Hougaard (2000), Therneau and Grambsch (2000), Duchateau et al. (2002), (2003), Duchateau and Janssen (2004).

A shared frailty model in survival analysis is defined as follows. Suppose there are n clusters and that the i^{th} cluster has n_i individuals and associates with an unobserved frailty U_i (1<i<n). Let $t_{ij}' = (t_{i1}, t_{i2}, ...t_{in_i})$ represent the survival times for the j^{th} individuals in the i^{th} cluster. Conditional on frailties, U_i, the survival times are assumed to be independent and their hazard functions to be of the form:

$$h_{ij}(t \mid U_i, x_{ij}) = U_i h_0(t) \exp(\beta^t x_{ij}) \tag{50}$$

With $h_0(t)$ the baseline hazard function and β is a vector of fixed effect parameters to be estimated. The frailties U_i are assumed to be identically and independently distributed random variables with common density function $f(u,\theta)$, where θ is the parameter of the frailty distribution. A semi-parametric shared frailty model is a frailty model with a non-parametric baseline hazard function $h_0(t)$. The variability of U_i determines the degree of heterogeneity among the groups. In empirical applications, the observed survival data are used to estimate the parameters of the distribution of frailty $f(u,\theta)$ and to actually predict the individual frailties.

For reasons of convenience, analysts often choose parametric representations of frailty models that are mathematically tractable. Hougaard (1986) used several distributions for frailty including gamma, inverse Gaussian, positive stable distributions and claimed that these two distributions are relevant and mathematically tractable as a frailty distribution for heterogeneous populations. Flinn and Heckman (1982) used a lognormal distribution for frailty, whereas Vaupel et al. (1979) assumed that frailty is distributed across individuals as a gamma distribution. In this study we used the gamma distribution which is the main frailty distributions widely used in the literature because of its simplicity and mathematical tractability.

Frailty models can be expressed in terms of Laplace transform. Once the Laplace transform of frailty distribution is obtained, it is easy to obtain the estimates the parameters of frailty models.

Let $S_{ij}^{*}(t_{ij})$ be the baseline survivor function of the j^{th} member in cluster i at time $T_{ij} = t_{ij}$. In frailty models, conditionally on a random effect of a cluster, say U_i, the survivor function takes a form of the Lehmann family of alternatives, and the failure times for cluster i, $T_{i1}, T_{i2}, ..., T_{in_i}$ are assumed to be dependent. Then the joint survivor function of $t_{i1}, t_{i2}, ..., t_{in_i}$ conditioning on X_i, say $S_i(t_{i1}, t_{i2}, ..., t_{in_i} | U_i)$ is

$$S_i(t_{i1}, t_{i2}, ..., t_{in_i} | U_i) = \prod_{j=1}^{n} S_{ij}(t_{ij} | U_{ij}) = \prod_{j=1}^{n} S_{ij}^{*}(t_{ij})^{u_i}$$

$$= \exp\left(-U_i \sum_{j=1}^{n} H_0(t_{ij}) \exp(\beta^t x_{ij}) \right) \qquad (51)$$

Additionally, U_i is assumed to be distributed as $f(u_i)$ with this two assumptions, the joint survivor function is

$$S_i(t_{i1}, t_{i2}, ..., t_{in_i}) = p_r(T_{i1} > t_{i1}, T_{i2} > t_{i2}, ..., T_{in} > t_{in})$$

$$= \int \prod_{j=1}^{n} S_{ij}(t_{ij})^{u_i} \, f(x_i) dU_i$$

$$= \int \exp\left(-U_i \sum_{j=1}^{n_i} H_0(t_{ij}) \exp(\beta^t x_{ij}) \right) f(U_i) dU_i$$

$$= \mathrm{L}\left(\sum_{j=1}^{n_i} H_0(t_{ij}) \exp(\beta^t x_{ij}) \right) \quad (52)$$

Where L is the Laplace transform of the density function and $H_0(t) = \int_0^t h_0(v)dv$ .And the marginal survivor function is

$$S_i(t_{ij}) = \int S_{ij}^*(t_{ij})^{u_i} f(x_i) dU_i \quad (53)$$

$S_i(t_{ij})$ can be interpreted as the average survivor function for the population.

From the above it is clear that the joint survivor function for one group is the Laplace transform of the frailty density function f(u) with parameter $\left(\sum_{j=1}^{n} H_0(t_{ij}) \exp(\beta^t x_{ij}) \right)$. In this thesis we have considered only the gamma distribution and Gaussian frailty models. For other distribution see Hougaard (2000), and Ohman and Eberly (2001).

Gamma distributions have been used for many years to generate mixtures in exponential and Poisson models. From a computational point of view, gamma models fit very well into survival models, because it is easy to derive the formulas for any number of events. This is due to simplicity of the derivatives of the Laplace transform. This is also the reason why this distribution has been applied in most of the applications published until now.

3.4.6.2. Frailty Distributions

The frailty denoted by U_i is an unobservable realization of a random variable U with probability density function $f(u)$, the frailty distribution. Since U_i multiplies the hazard function, U has to be non-negative. Another constraint is further needed for identifiability reasons, more specifically; the mean of U is typically restricted to unity in order to separate the baseline hazard from the overall level of the random frailties. The main difference between multivariate

and univariate frailty models is the assumption of how frailty is distributed in the data. Shared (multivariate) frailty models assume that similar observations share frailty i.e. the frailty distribution variability is related to a measure of dependence between clustered subjects, whereas it is rather interpreted as a measure of over dispersion which is caused either by misspecification or omitted covariates in the univariate case. In this research, frailty distributions namely the gamma and the inverse Gaussian were used. In both cases, as a single heterogeneity parameter (denoted by θ) indexes the degree of independence.

3.4.6.2.1. The Gamma Frailty Distribution

The gamma distribution has been widely applied as a mixture distribution (for example, Greenwood and Yule, 1920, Vaupel et al., 1979, Congdon, 1995, Santos et al., 1995 and Hougaard, 2000. 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). The density of a gamma distributed random variable with parameter is given by

$$f_u(u) = \frac{u_i^{\frac{1}{\theta}-1}\exp(-u_i/\theta)}{\theta^{\frac{1}{\theta}}\Gamma(\frac{1}{\theta})}, \theta > 0 \tag{54}$$

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

$$L(u) = (1+u/\theta)^{-\theta} \tag{55}$$

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

$$S_\theta(t) = [1-\theta\ln(S(t))]^{-1/\theta}, \tag{56}$$

And the conditional hazard function is given by:

$$h_\theta(t) = h(t)[1-\theta\ln(S(t))]^{-1} \tag{57}$$

Where S(t) and h(t) are the survival and the hazard functions of the baseline distributions. For the Gamma distribution, the Kendall's Tau (Hougaard 2000), which measures any two event times from the same cluster in the multivariate case, can be compute by: $\tau = \frac{\theta}{\theta+2} \in (0,1)$

3.4.6.2.2. The Inverse Gaussian Frailty Distribution

The inverse Gaussian (inverse normal) distribution was introduced as a frailty distribution as an alternative to the gamma distribution by (Hougaard, 1984). The gamma distribution is the most commonly used frailty distribution because of its mathematical convenience. However, it has drawbacks (see Kheiri et al.,2007) for example it may weaken the effect of covariates. Alternative to the gamma distribution Hougaard (1984) introduced inverse Gaussian as a frailty distribution. The inverse Gaussian distribution has many similarities to standard Gaussian distribution (Chikkara and Folks, 1986). Furthermore, it provides much flexibility in modeling, when early occurrences of failures are dominant in a life time distribution and its failure rate is expected to be non-monotonic. In such situations the inverse Gaussian distribution might provide a suitable choice for the lifetime model. Also inverse Gaussian is almost an increasing failure rate distribution when it is slightly skewed and hence is also applicable to describe lifetime distribution which is not dominated by early failures. Secondly, for the in-verse Gaussian distribution the surviving population becomes more homogeneous with respect to time, whereas for gamma distribution the relative heterogeneity is constant. The inverse Gaussian distribution has unimodal density and is the member of exponential family, while its shape resembles the other skewed density functions, such as lognormal and gamma. These properties of inverse Gaussian distribution motivate us to use inverse Gaussian as frailty distribution.

Let a continuous random variable U follows inverse Gaussian distribution with parameters μ and θ then the density function of U is,

$$f(u)=\left[\frac{\theta}{2\pi}\right]^{1/2} u^{-3/2} e^{-\frac{\theta(u-\mu)^2}{2u\mu^2}}, u>0, \theta>0, \mu>0 \qquad (58)$$

The mean and variance of frailty models are $E(U)=\mu$ and $Var(Z)=\frac{\mu^3}{\theta}$.

The mean and the variance are 1 and θ, respectively with Laplace transform

$$L(s) = L_Z(s) = \exp\left(\frac{\theta}{\mu} - \sqrt{\frac{\theta^2}{\mu^2} + 2\theta s}\right), \; s \geq 0 \tag{59}$$

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

$$S_\theta(t) = \exp\left\{\frac{1}{\theta}\left(1 - [1 - 2\theta \ln\{S(t)\}]^{1/2}\right)\right\} \tag{60}$$

And the conditional hazard function is given by:

$$h_\theta(t) = h(t)[1 - 2\theta \ln(S(t))]^{-1/2} \tag{61}$$

where S(t) and h(t) are the survival and the hazard functions of the baseline distributions

With multivariate data, an Inverse Gaussian distributed frailty yields a Kendall's Tau given by:

$$\tau = \frac{1}{2} - \frac{1}{\theta} + 2\frac{\exp(2/\theta)}{\theta^2}\int_{2/\theta}^{\infty} \frac{\exp(-u)}{u} du \in (0, 1/2) \tag{62}$$

3.4.6.3. Penalized Partial Likelihood for Shared Frailty Models

Hastie and Tibshirani (1993) proposed a general model with time varying coefficients and suggested estimation through penalized partial likelihood. For the vectors of baseline hazard functions, if conditionally on b the censoring is independent and non-informative also of b, then the likelihood for model (50) in terms of the parameters $(h_0(t, \beta, \theta)$ is:

$$\begin{aligned} l(h_0(t), \beta, \theta) &= \int \prod_{i=1}^{n} h_i(t/b)^{\delta_i} S(t/b)\, p(b; D/\theta) \partial b \\ &= \int \prod_{i=1}^{n} h_0(t/b) \;\; \exp(X_i^t \beta + U_i b)^{\delta_i} \;\; [\exp\{-H_0(t)\exp(X_i^t \beta + U_i b \;)\}\, p(b; D/\theta) \partial b \end{aligned} \tag{63}$$

Where $H_0(t) = \int_0^t h_0(u) \partial u$ and the observed frailties are integrated out.

We restrict b to follow a multivariate normal distribution, but the derived likelihood approximations can be easily adapted to other frailty distributions as well. The approximate marginal log likelihood:-

$$L(h_0(t),\beta,\theta) \approx -\frac{1}{2}\log|D(\theta)| - \frac{1}{2}\log|\sum_{i=1}^{n} H_o(t)\exp(X_i^t\beta + U_i b - U_i U_i - D(\theta)^{-1}| +$$

$$\sum_{i=1}^{n}\delta_i[\log(h_0(t) + X_i^t\beta + U_i\beta] - H_0(t)\exp(X_{ij}^t\beta + U_i\beta) - \frac{1}{2}bD(\theta)^{-1}b \qquad (64)$$

If both θ were known and b were considered a fixed effects parameter, then the second line in (64) Would be a penalized log likelihood (Green, 1987), where $-\frac{1}{2}bD(\theta)^{-1}b$ is the penalty term penalizing for extreme values of b. Since the second line is the full likelihood for a Cox model with b as another set of parameters and a penalty term, it turns out that it can be maximized using penalized fixed effects partial likelihood (PPL),

$$\delta_i[X_i^t\beta + U_i b] - \log\sum_{j\in R(t_i)}(\exp + U_i b) - \frac{1}{2}bD(\theta)^{-1}b$$

For given θ, the estimated equation based on the first partial derivatives of the PPL are, for β

$$\sum_{i=1}^{n}\delta_i[X_i - \frac{U_i\exp(X_i^t\beta + U_i b)}{\sum_{j\in R(t_i)}\exp(X_j^t\beta) + U_i b)}] = 0 \qquad (65)$$

The estimated equation for b

$$\sum_{i=1}^{n}\delta_i[U_i - \frac{U_i\exp(X_i^t\beta + U_i b)}{\sum_{j\in R(t_i)}\exp(X_j^t\beta) + U_i b)}] - D(\theta)^{-1} = 0 \qquad (66)$$

And $(\beta(\theta), b(\theta))$ can be found by alternating between solving (65) and (66).

4. STATISTICAL ANALYSIS AND RESULTS

4.1. Baseline Characteristics of the Study

Baseline categorical covariates are illustrated in Table 2. There were 1371 children under six months of age who were under breast feeding. The average mean time of EBF was 3.128 months and the standard error was 1.874 months. The maximum and minimum time of EBF was zero and six months respectively. 588 (40.7%) mothers continue feeding their child breast milk exclusively until six months of infant's age (censored) and 813(59.3%) of mothers introduce additional substitute liquids and solid foods before six months of age based on the women's questionnaire under EDHS, 2011 data (event). .

1155(84.25%) and 216(15.75%) of infants were from rural and urban parts of Ethiopia respectively. The proportion of mothers who introduce foods and breast milk substitute under based on the two residences were 57.23% and 70.37%, respectively. 1245 and 126 of infants mothers' were user and non-user of contraceptive method, respectively. Out of the non-user group 58.63% infants were fed breast milk non-exclusively and 41.36% exclusively.916 of infants' mothers were non-educated, 373 have attended primary education, and, 82 of infants mothers' attended secondary & higher education. Among those non-educated mothers 58.62% have fed their infants non-exclusive breast milk and 41.32% of them fed exclusively breast milk.

434, 240 and 391 of infants were comprised from rich, middle, poor family respectively. Among the rich mothers considered, 62.90% of them were introduced breast milk substitute for their child during the six month period while the rest 42.77% were censored. Out of poor mothers under study 29.92% of mothers ends EBF 69.08% of mothers continue EBF. 90.153% of infant's mothers under study were married and the remained were not married at the interview time. 49.599% of mothers of infant's under study were Muslim, and 2.845% were other religion followers. 58.351% of infants were delivered at home and 41.648% of infants were delivered at health institution. 9.555% of infant's mothers' age under study was between 15 and 19, 85.560% of them are between 20 and 39 and 4.887% of mothers are between 40 and 49 age which were based on young, adult and old categories.

Table 2: Socio-demographic characteristics of respondents (n = 1371).

Socio-demographic variables	Category	Number out of 1371	Status	
			Event	Censored
Residence	Rural	1155(84.245%)	661(57.23%)	494(42.77%)
	Urban	216(15.755%)	152(70.37%)	64(29.63%)
Contraceptive	User	126(9.913%)	83(65.87%)	43(34.13%)
	Non-user	1245(90.810%)	730(58.63%)	515(41.36%)
Education	No education	916 (66.812%)	537(58.62%)	379(41.38%)
	Primary	373(27.206%)	219(58.71%)	154(41.29%)
	Secondary& higher	82(0.060%)	57(69.51%)	25(30.49%)
Wealth	Rich	434(31.656%)	273(62.90%)	161(37.10%)
	Middle	240(17.505%)	117(48.75%)	123(51.25%)
	Poor	391(28.520%)	117(29.92%)	274(69.08%)
Religion	Muslim	680(49.599%)	458(67.35%)	222(32.65%)
	Orthodox	384(28.009)	173(45.03%)	211(54.97%)
	Protestant	268(19.550%)	161(60.07%)	107(30.03%)
	Others	39(2.845%)	18(46.15%)	21(53.85%)
Women's age	15-19	131(9.555%)	70(54.43%)	61(45.57%)
	20-39	1173(85.560%)	706(60.19%)	467(39.81%)
	40-49	67(4.887%)	37(55.22%)	30(44.78%)
Work status	Not-working	1074(78.340%)	637(59.31%)	437(40.69%)
	Working	297(21.663%)	176(59.26%)	121(40.74%)
Place of delivery	Home	800(58.351%)	472(59.00%)	328(41.00%)
	Health Institution	571(41.648%)	341(59.72%)	230(40.28%)
Marital status	Married	1236(90.153%)	738(59.71%)	498(40.29%)
	never in union	75(5.47%)	44(58.67%)	31(41.33%)
	Others	60(4.376%)	31(51.67%)	29(48.33%)

Region	Number out of 1371	Status	
		Event	Censored
Addis Ababa	44(3.21%)	33(75.00%)	11(25.00%)
Affar	136(9.92%)	112(82.35%)	24(17.65%)
Amhara	138(10.06%)	42(30.43%)	96(69.56%)
Benishangul-Gumuz	114(8.31%)	61(53.51%)	53(46.49%)
Dire Dawa	88(6.42%)	55(62.50%)	33(37.50%)
Gambela	108(7.88%)	84(77.78%)	24(22.22%)
Harari	68(4.96%)	46(67.65%)	22 (32.35%)
Oromia	232(16.92)	126(54.31%)	106(45.69%)
Southern Nations and Nationalities	189(13.78%)	87(46.03%)	102(53.97%)
Somali	137(9.99%)	114(83.21%)	23(16.79%)
Tigray	117(8.53%)	53(45.30%)	64(54.70%)

The maximum and minimum average duration of EBF was recorded in Dire Dawa administration and Benshangul-Gumuz region which was 3.51 and 2.77 months respectively. Among Addis Ababa administration 75% of infants were introduced other breast milk substitute in addition to breast milk (non-exclusive breast fed) and 25% of them were exclusively breast fed (censored). In Dire Dawa administration 62.50% of infants introduce other solid and liquid food and water and 37.50% were exclusively used breast milk. 53.51% of infants in Benishangul-Gumuz were non exclusively used breast milk and 46.49% of them were use breast milk exclusively. In Oromia in which the largest sample was used, 54.31% of infants were used breast milk non-exclusively and 45.69% of them used breast milk exclusively.

4.2. Non-parametric Survival Analysis

1371 infants who are under six months of age were considered to study factors affecting early termination of EBF. Out of the total of 1371 infants 813 (59.30%) of infants introduced with other foods in addition to mother breast milk (event) and 558 (40.7%) of infants were exclusively breast feed until the end of the study (censored). 37 (2.70%), 122 (8.90%), 233 (17%), 363 (26.48%), 510 (62.73%), and 642 (37.20%), of infants introduced other additional food within

zero up to five months respectively. The overall median estimated EBF time of infants under the study was 5 (95% CI: 4-5.3) months. In order to get a closer look at the estimate of the survival time we have used the Kaplan-Meier and Nelson-Aalen estimation techniques.

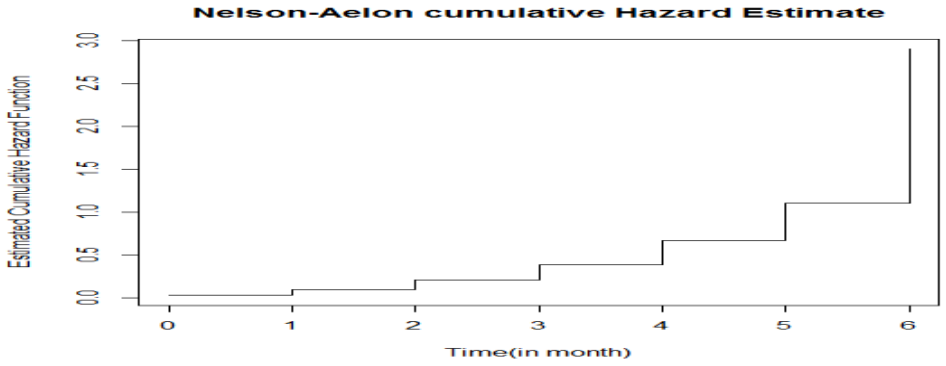


Figure 1: The plot of the overall estimate of Nelson-Aalen cumulative hazard estimates of infants under exclusive breast feeding based on EDHS, 2011 data.

The graph, Figure 2, of the estimate for overall Kaplan-Meier survivor function depicted that, relatively, few number of mothers start to give other substitute food for their child at the earlier months of birth and half of mothers start to feed other substitute food in addition to mother breast milk at the fifth month, the same graph showed the decrement over a follow up period. A separate graph of the estimates of the Kaplan-Meier survivor functions is constructed for different covariates. In so doing it is possible to see the existence of difference in survival experience between the indicated categories of individuals.

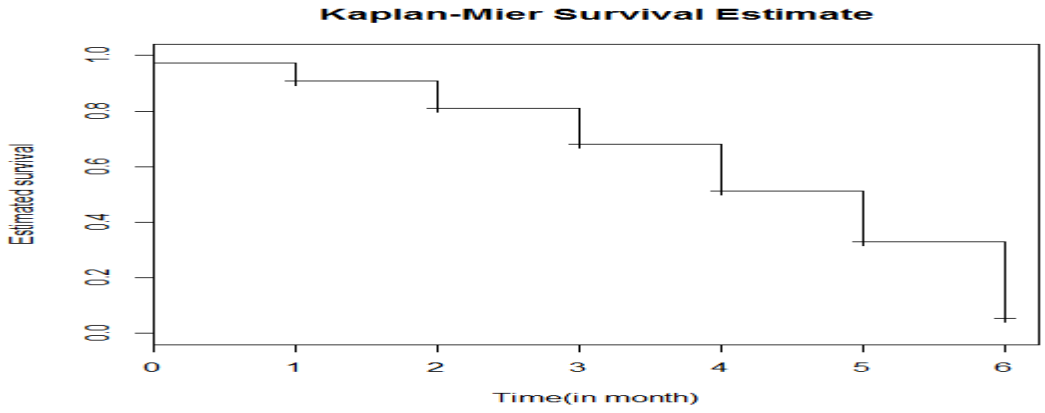


Figure 2: The plot of the overall estimate of Kaplan-Meier survivor function of infants under exclusive breast feeding based on EDHS, 2011 data.

In general, the pattern of one survivorship function lying above another means the group defined by the upper curve had a better survival than the group defined by the lower curve. Some of the graphs did not show clear differences among the intended categories.

Table 3: Kaplan-Meier analyses of survival times for exclusive breast feeding according to important socio-demographic characteristics of infant's mothers, in Ethiopia, EDHs, 2011.

Covariate	Category	Median	SE	95% CI
Residence	Rural	5	0.018	[4 ,5]
	Urban	4	0.038	[4 ,5]
Contraceptive	User	5	0.051	[5 ,6]
	Non-user	5	0.017	[4 ,5]
Education	No education	5	0.020	[4 ,5]
	Primary	5	0.031	[4 ,5]
	Secondary& higher	4	0.064	[4 ,5]
Wealth	Rich	5	0.028	[4 ,5]
	Middle	5	0.044	[5, 5]
	Poor	4	0.022	[4 ,5]
Religion	Muslim	4	0.022	[4 ,5]
	Orthodox	5	0.036	[5 , 6]
	Protestant	4	0.035	[4 , 5]
	Others	5	0.103	[4 ,5]
Womens'age	15-19	5	0.056	[4 , 5]
	20-39	5	0.017	[4 , 5]
	40-49	5	0.076	[4 , 6]
working status	No	4	0.018	[4 , 5]
	Yes	5	0.035	[5 , 5]
Place of delivery	Home	4	0.021	[4 , 5]
	Health stations	5	0.026	[4 ,5]
Marital status	Married	5	0.017	[4 , 5]
	Never in union	5	0.072	[4 ,5]
	Others	5	0.084	[4 , 5]

However, among others, graphs of wealth status, contraceptive use, education status of infant's mothers and religion of mothers relatively shows fair gaps and convey similar information as Table 3.

For instance, infants whose mother live in rural areas, have no education and at primary education level, mothers who are orthodox and other religion followers, mothers who were currently on work at the interview time have longer experience of survival time than who live in urban, secondary and higher education level, Muslim & Protestant religion followers, mothers at work. Those infants who were delivered in health station have also longer experience of survival than those that had delivered at home in cessation of EBF.

To check for significant differences among categories of factors that are shown using the Kaplan-Meier estimates of the survivor functions, we employed a log-rank statistical test. Based on the log-rank test, there was no significant difference in survival experience between the various categories of education level of mothers, marital status, mother's age, and infants born in health station or home. However, the log-rank test showed that the survival experience of feeding breast milk exclusively in different categories of residence of mothers, whether a mother is contraceptive user or not, mothers work status, mother's wealth status and religion differ significantly at 10% level of significance. The results are shown in Table 7 of the appendix. A close examination of Figure 5 (a - i) and Table 3 in the appendix reveal that infants mothers who were not currently working at the interview time, who lived in rural areas, mothers who were rich and have middle income, mothers who were orthodox and other religion followers feed exclusively breast milk for their child relatively for longer median time (had better survival time).

4.3. Results of the Cox PH Model

The aim of model development is to obtain a model that satisfactorily describes the data at hand. For the same purpose, the first step is to select covariates which are important in the study at some relaxed level of significance. In this study, a model that contains all variables that are significant in the univariable analysis (Table 8 in the Appendix) at the 10% percent level of significance is used for selection of important covariates.

In the univariable (table 8 in the Appendix) Cox PH models, the model with the covariates; contraceptive non-user (P-value=0.0171 when user is as a base line), orthodox (P-value=0.000, when Muslim is as reference), rural children (P-value=0.0618 when urban children is as reference), poor (P -value= 0.062 when rich is as a reference) and secondary and higher educated mothers (p-value=0.069 when non-educated mothers taken as a reference) shows statistically significant association with duration of EBF at 10% level of significance. So that these significant variables were included in the multivariable model (table 9 in the Appendix) and the remaining four covariates were not included. Then the full multivariable Cox PH model is fitted including all the potential covariates which are significant at 10% level, at the univariable level variables by using forward selection method. Since the p-value of the added variables in the full multivariable model is larger than the value of entry=0.1 we stopped entry of new variable here. Accordingly the final Cox PH model candidate variables with minimum AIC were; educational level, wealth status, place of residence, contraceptive usage, religion of mothers.

Therefore the potential covariates to be kept in the multivariable Cox PH model are contraceptive, place of residence, wealth index, education level, and religion. These are significant at the 5% level of significance (Table 4). Finally, we take into account the possible interactions among covariates that are significant at multivariable level of analysis (i.e., by taking one covariate at a time to the preliminary model). The likelihood ratio test is employed for the same task and the result verifies that none of the interaction terms were significant at the 5% level (Table 10 in the Appendix). Once again, the result ensures that the preliminary model of the study will contain only the five covariates in Table 4. The parameter estimates and hazard ratios of the covariates are shown in Table 4 below.

Table 4: Estimated values of the coefficients, hazard ratios, 95% CI for the hazard ratio and P-values of the explanatory variables on fitting the proportional hazards model to the data from EDHs 2011 under six months of age old children.

Covariates	Coeff	SE	P-value	HR	95% CI of HR
Contraceptive					
Non-user	0.335	0.129	0.009*	1.398	[1.085 , 1.800]
Religion					
Orthodox	-0.487	0.092	0.000*	0.614	[0.513 , 0.735]
Protestant	-0.137	0.094	0.149	0.872	[0.725 , 1.047]
Others	0.025	0.224	0.911	1.025	[0.660 , 1.591]
Residence					
Rural	-0.413	0.119	0.001*	0.661	[0.524 , 0.836]
Wealth					
Middle	0.109	0.125	0.383	1.115	[0.873 , 1.424]
Poor	0.351	0.099	0.001*	1.420	[1.170 , 1.725]
Education level					
Primary	0.254	0.088	0.001*	1.289	[1.086 , 1.531]
Secondary&higher	0.389	0.161	0.016*	1.475	[1.076 , 2.022]
Likelihood ratio test=62.9 on 9 DF, p-value =000.					

Coef: coefficient for covariate, SE: standard error of coefficient, HR: hazard ratio; p-value: probability value, 95%C.I HR: 95% confidence interval for HR, * Significant at 0.05 level.

From Table 4 we observe that contraceptive non-user (P-value=0.009 when user is as a base line), orthodox (P-value=0.000, when Muslim is as reference), rural children (P-value=0.001when urban children is as reference), poor (P -value= 0.0010 when rich is as a reference) and secondary and higher educated mothers (p-value=0.016 when non-educated mothers taken as a reference) and primary educated mothers(p-value=0.001 when non-educated mothers taken as a reference) shows statistically significant association with duration of EBF at the 5% level of significance. Again also, the likelihood ratio test is also highly significant.

4.3.1. Diagnosis of the Cox PH Model

Since fitting a model is not the end of the story, we need to asses some requirement of the model of the study. In other words, the preliminary final model shall be diagnosed for describing our data optimally or not. In this setting, the requirement is all about the diagnosis for the final proportional hazards model that consists of testing the assumption of PHs, checking for the presence of leverages (influential observations) and measuring the overall goodness of fit of the model.

4.3.1.1. Assessing the PH Assumption

The PHs assumption, which asserts that the hazard ratios are constant overtime, is vital to the interpretation and use of a fitted PHs model. That means the risk of failure must be the same no matter how long subjects have been followed. In order to test the said assumption above, the extended Cox model is employed and graphical display is used to substantiate the same. For that reason, all interactions of covariates with the logarithm of survival times are modeled together with the main effects; and likelihood ratio test is used to test the significance of the interaction.

Table 11 of the Appendix showed that none of the coefficients of the interaction terms are significant at 5% level. Since the interaction effect is found to be non- significant (meaning high p-values), this is a confirmation for the absence of time varying covariates. Put differently, there are no covariates which show a trend/pattern with the time and therefore the HRs will be constant over the study time. Therefore, there is no sufficient evidence to reject the null hypothesis that the coefficients of the time varying variables (interaction terms) are zero. This ascertains the validity of the assumption of the PHs.

Furthermore, plotting the scaled Schoenfeld residuals of each covariate against log time was used to check whether the assumption of PHs is violated or not. Figure 6 of the Appendix shows the plot of scaled Schoenfeld residuals of each covariate for the final model. The graphs depict that each of the nine plots are random, smooth and approximate the horizontal line through zero. Thus, there is no covariate which has interaction with log of time revealing the PHs assumption is met.

4.3.1.2. Checking for Overall Goodness of Fit

The next step in the model assessment is to measure the overall goodness of fit. For this objective we use the Cox-Snell residuals. The plot of the Nelson-Aalen estimate of the cumulative hazard function of the Cox-Snell residual against the Cox-Snell residuals is presented in Figure 5 below.

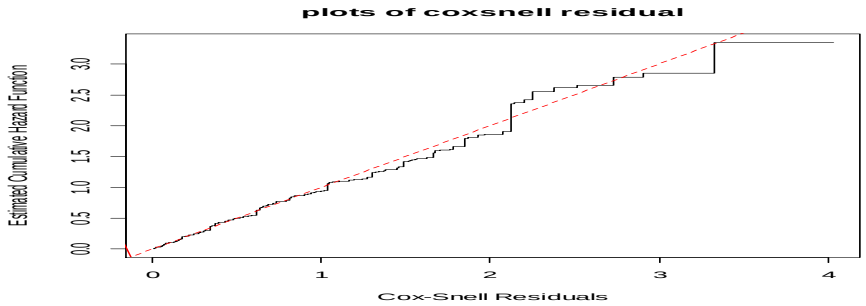


Figure 3: Cumulative hazard plot of the Cox-Snell residuals of the proportional hazards Cox regression model in table 4. The 45^0-striaght line through the origin is drawn for reference

It can be seen that the plot of the residuals in Figure 5 is fairly close to the 45^0 straight lines through the origin. Thus, the plot is evidence that the model fitted to the data is satisfactory. In addition, results of the likelihood ratio, score and Wald tests for model goodness of fit displayed in Table 5 suggest that the model was good fit (i.e. at 5% level of significance). Therefore, the model with estimates as given in Table 5 was the final model of the Cox PH model.

4.4. Results of Cox PH with Shared Gamma Frailty and Inverse Gaussian Frailty Models

In Cox PH with shared gamma frailty models the same to Cox PH done above, first univariable (Table 12 in the Appendix) analysis were done for all variables. Then the full multivariable Cox PH model is fitted including all the potential covariates which are significant at 10% at the univariable level by using forward selection method.

In the univariable shared gamma frailty (Table 12 in the Appendix) model all covariates become statistically insignificant except education level and contraceptive use at 10% level of significance. The Cox PH with gamma frailty model estimates the effect of education level,

religion and wealth on the log hazard was -0.168 (se: 0.0900, p-value: 0.062), 0.452 (se: 0.089 and p-value: <0.000) and 0.145 (Se: 0.078 and p-value: <0.062) respectively. All tests of the likelihood-ratio test of frailty term, for all univariable models is significant (p-value < 0.000), indicating that the variability among regions should have to be considered.

The multivariable Cox shared gamma frailty model was fitted with covariates contraceptive use and education level (Table 13 in the Appendix). Then, we added new variables in the model until the p-value of the added variables in the full multivariable model is larger than the value of entry=0.1. Accordingly the final Cox PH with shared gamma frailty model candidate variables with minimum AIC were; educational level, wealth status, residence, contraceptive usage, marital status with regional random term.

In the same way the in univariatiable Cox PH with shared inverse Gaussian frailty model (Table 14 in the Appendix) all covariates also become statistically insignificant except education level and contraceptive use at 10% level of significance. The likelihood-ratio test of the frailty term, for all univariate models were significant (p-value < 0.000), meaning that the variability among regions should have to be considered.

The multivariable model is fitted with only two covariates contraceptive use and education level (Table 15 in the Appendix). Accordingly the final Cox PH with shared inverse Gaussian frailty model candidate variables with minimum AIC were selected by using forward selection method; educational level, wealth status, residence, contraceptive usage, marital status with regional random effect.

4.5. Comparison of Cox PH versus Shared Frailty Models

In this study, in order to compare the efficiency of the models the AIC (Akaike's Information Criterion), and log likelihood was used. From the Table 5 we can see that the shared gamma frailty model has a minimum AIC, indicating that this model fit the data better than the Cox PH model which did not take in to account the clustering. Accordingly, the Cox PH with gamma shared frailty model explained the dataset better than inverse Gaussian shared frailty model and the traditional Cox PH model.

.

Table 5: Comparison of Cox PH without frailty term and with Gamma and inverse Gaussian shared frailty models.

Models	Log-liklihood (intercept)	Log-liklihood (full model)	DF	AIC
Cox PH	-4971.718	-4940.262	8	9898.523
Cox PH with Gamma frailty	-4971.718	-4907.153	8	9830.305
Cox PH Inverse Gaussian frailty	-4971.718	-4907.734	9	9831.467

DF: degree of freedom

In the multivariable shared gamma and inverse Gaussian frailty model the heterogeneity parameter θ is estimated to be 0.281 (Chisq= 44.810) and be 0.1193 (Chisq =85.36).

Generally, for every combination of the covariates we fitted, the Cox PH with shared gamma frailty model yield a minimum value of Akaike's Information Criteria (AIC=9830.305) suggesting that the semi-parametric shared gamma frailty model may provide a better description of the data set for any combination of variable in the dataset. Hence, we use Cox PH with shared gamma frailty model with a combination of covariates to discuss the effect of covariates on duration of EBF (given in Table 6). Therefore the interpretation of the covariates was based on the shared gamma frailty model with covariates place of residence, economic status (wealth index), contraceptive use, education level and marital status.

4.6. Results and Presentations of the Final Model

The interpretation from the results of the final model that consists of the main effects is based on the conditional hazard ratios. That is, the coefficient of a categorical explanatory variable in the model can be interpreted as the logarithm of the ratio of the hazard of EBF to the baseline (reference group) hazard. In other words, comparison is made with the reference category and between groups for the categorical covariates. Consequently, the interpretation of covariates that are included in the final proportional hazard gamma shared frailty model for infants under six months of age is as follows.

It is observed that the test for the regression parameters is rejected with likelihood ratio test = 129.100 with degree of freedom= 18.290 and probability value is 0.001. Variables that are found

to be significantly associated with duration of EBF in the fitted gamma shared Cox regression model are place of residence, economic status (wealth index), contraceptive use, education level at 5% significance level, on the other hand marital status have no significant influence on duration of EBF. The HR for rural infants given heterogeneity among regions in relation to those who are urban is 0.778 (95% CI: 0.607 0.997). It means the rural infants have about 22.22% lower rate to have other substitute liquids and solid foods in addition to breast milk than infants living in urban. The 95% confidence interval indicates that the hazard rate goes to a maximum of 0.997 and a minimum of 0.607. The duration of EBF for infants from middle economic status

Table 6: Estimated values of the coefficients, hazard ratios, 95% CI for the hazard ratio and P-values of the explanatory variables on fitting the Cox PH with shared gamma frailty model to the data extracted from women's data sheet, EDHs (2011).

Covariates	HR	Coef($\widehat{\beta}$)	SE	P.value	95% CI of HR
Residence					
Rural	0.778	-0.251	0.127	0.047*	[0.607,0.997]
Contraceptive					
Non-user	1.475	0.389	0.133	0.003*	[1.137,1.915]
Wealth					
Middle	1.139	0.130	0.126	0.300	[0.889,1.459]
Poor	1.257	0.229	0.103	0.026*	[1.028,1.538]
Education level					
Primary	1.266	0.234	0.091	0.009*	[1.059,1.513]
Secondary & higher	1.489	0.398	0.162	0.014*	[1.083,2.048]
Marital status					
partner(1)	1.185	0.169	0.161	0.290	[0.864,1.625]
Others(2)	0.726	-0.320	0.188	0.090	[0.502,1.051]
Variance of random effect= 0.281, 1, chisqr=79.06,prob>= chisqr=1.1e-^{12}; Likelihood ratio test of theta = 129.1, degree of freedom=18.29, p-value<0.001.					

Coef: coefficient for covariate, SE: standard error of coefficient, HR: hazard ratio; p-value: probability value, 95%C.I HR: 95% confidence interval for HR, * Significant at 0.05 level.

family has no significant difference with the infants from rich and richest family. But there is a significance difference in duration of EBF between infants from poor family and rich family at 5% significance level. The conditional HR for infants from poor family in relation to those who are from rich family is 1.257 (95% CI: (1.028, 1.538)). The interpretation is that infants from poor family have 25.760% higher risk to stop EBF early or duration of EBF for infants from lower income family is shorter than infants from rich family conditional on frailty. The confidence interval suggests that the conditional HRs can go as low as 1.028 and as high as 1.538.Infants whose mothers didn't use contraceptive have 47.550% shorter duration of EBF than infants whose mothers have used contraceptive (HR=1.475,95% CI: 1.137 - 1.914). With increasing education level of mothers the possibility of mothers to feed their infant's breast milk exclusively for the first six months is lower and lower. Mothers who are at primary education level have 26.600% probability of terminating EBF in relation to non-educated mothers (HR=1.266, 95% CI: 1.059, 1.512). Mothers who are at secondary and higher education level have 48.92% tendency of terminating EBF in relation to non-educated mothers within six months of age of infants (HR=1.489,95% CI: 1.083,2.047).

4.6. 1. Diagnosis of Cox PH with Shared Gamma Frailty Model

The preliminary final model shall be diagnosed for describing our data optimally or not. In this setting, the requirement is all about the diagnosis for the final gamma shared frailty proportional hazards model that consists of testing the assumption of proportional hazards and measuring the overall goodness of fit of the model.

4.6.1.1. Assessing the PHs Assumptions of Cox PH with Shared Gamma Frailty Model

The PHs assumption, which asserts that the HR is constant overtime, is vital to the interpretation and use of a fitted PHs model. That means the risk of failure must be the same no matter how long subjects have been followed. In order to test the said assumption above, the extended Cox model is employed and graphical display is used to substantiate the same. For that reason, all interactions of covariates with the logarithm of survival times are modeled together with the main effects; and likelihood ratio test is used to test the significance of the interaction terms at 5% level of significance.

Table 16 of the Appendix show that none of the coefficients of the interaction terms are at the 5% significant level. The global test is also not statistically significant. Since the interaction effect is found to be non- significant, this is a confirmation for the absence of time varying covariates. In addition, Plots of scaled Schoenfeld residuals against transformed time for each covariate is shown in Figure 8 in the Appendix. It depicts that systematic departure from a horizontal line which is indicative of non-PHs. The assumption of PHs appears to be supported for all covariates (which is, recall, a dummy variable, accounting for the two bands in the graph).

4.6.1.2. Checking for Overall Goodness of Fit

The final step in the model assessment is to measure the overall goodness of fit. For this objective we use the Cox-Snell residuals. The plot of the Nelson-Aalen estimate of the cumulative hazard function of the Cox-Snell residual against the Cox-Snell residuals is presented in Figure 4 below.

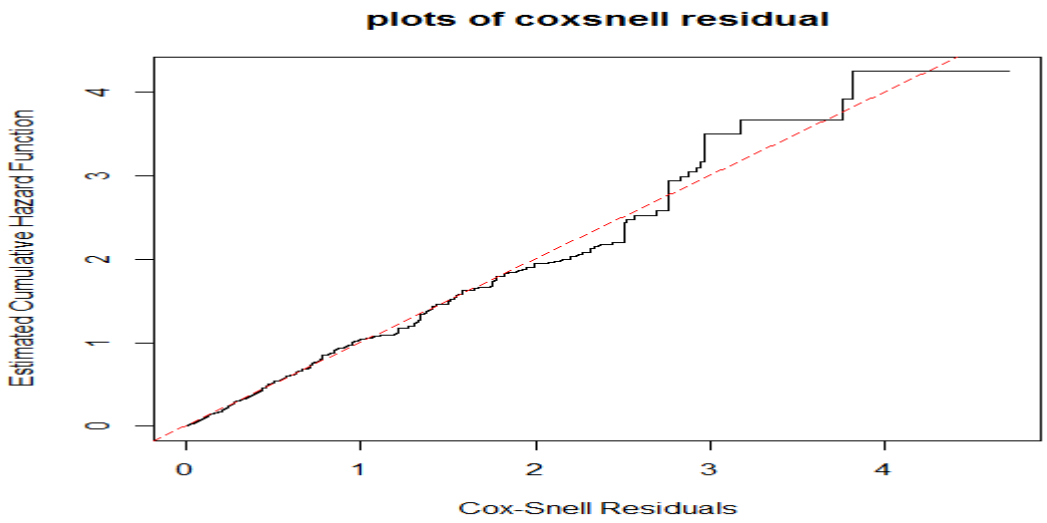


Figure 4: Cumulative hazard plot of the Cox-Snell residuals of the shared gamma frailty proportional hazards Cox regression model in table 6. The 45^0-striaght line through the origin is drawn for reference

It can be seen that the plot of the residuals in Figure 4 is fairly close to the 45^0straight line through the origin. Thus, the plot is evidence that the model fitted to the data is satisfactory. Moreover, results of the likelihood ratio, score and Wald tests for model goodness of fit

displayed in Table 6 suggest that the model was good fit(i.e. at 5% level of significance).Thus, the model with estimates as given in Table 6 was the final model of the Cox PH with shared gamma frailty model.

4.7. Discussions

The superiority of breast milk over any other milk nourishment of the human newborn and infant can hardly be challenged, and over the years it has become more and more apparent that it is the most ideal, safe and complete food that a mother can provide for her newborn. Regrettably, despite the enormous benefits of breast milk, the decline of EBF persists in many developing countries. Efforts made to promote breast milk use in the past few years have been encouraging and noteworthy to see mothers swung to EBF, in developed countries. Paradoxically enough, this unfavorable trend is noticeable in poor countries where the supply of artificial milk is scarce (Foo LL, et al., 2005). According to the UNICEF, only every third child living in the developing world is exclusively breastfed during the first six months of life (www.unicef.org). According to our study and previous studies in Sri Lanka (Sri Lanka Demographic and Health Survey, 2006; Perera,2011), EBF rates in Ethiopia and Sri Lanka are much higher than quoted by UNICEF.

This study tries to estimate and compare the survival time to EBF of infants under six months of age. In the current study, the median duration of EBF was five months which is lower than the WHO recommendations(WHO, 2001) and was attributed to various maternal and child factors. Nonetheless, when compared with previously reported figures for most developing countries like Tanzania (Simopoulos and Grave, 1984), Uganda (Engebretsen et al., 2007), Kenya (Bloss et al, 2004), Brazil (Carvalhaes et al.,2007), and earlier report in Ethiopia (Abate, 1999;Haider, 2006) the observed figure was similar.

In this study by using both the Cox PH without and with frailty model, we found that the factors that significantly affect the duration of EBF were place of residence, economic status (wealth index), contraceptive use, and education level. In the Cox PH model without frailty term, religion was a significantly affect EBF but in Cox PH with frailty model this variable was not selected.

The Cox PH model fitted using complete case analysis found five variables that can serve as predictive factors on the duration of EBF. These were place of residence, economic status (wealth index), contraceptive use, religion, and education level.

Out of the above mentioned predictors, contraceptive use was found to be an important factor that affects duration of EBF. The hazard of stopping feeding breast milk exclusively was lower for contraceptive user mother relative to non-users. This result is in line with the study done in in Bangladesh by Aktor and Rahman (2010). This may be due to the fact that contraceptive user women can protect early pregnancy after birth so that they could have ability to feed and protect their child. Urban mothers breastfed their children for a relatively shorter duration than did the rural mothers. This result is in conformity with the study done in Bangladesh by Aktor and Rahman (2010). In most of rural Ethiopia mothers are less exposed to breast milk substitute foods advertisement and there is also a culture that mother stay at home after birth. This encourages mothers to stay with their child and feed breast milk exclusively. In Urban areas, women are exposed to advertisement of breast milk substitute and mostly bottle feeding through artificial nipple in early age of children is common. In the other way, rich mothers' feed relatively longer duration than mothers whose wealth statuses were middle and poor. This finding agrees with the study done in Ethiopia by Tewodros et al. (2009) and contradicts with the study done in Brasil by Edson Theodoro (2013).

The other interesting finding in this study was that higher maternal education level was found to be associated with lower rate of EBF in Ethiopia. This might be explained by the fact that when women are better educated, the opportunity for employment is eminent and thus the opportunity to stay at home and practice EBF is compromised. This result also coincides with the study done in Ethiopia by Tewodros et al. (2009); SirLanka by Perera et al. (2011); Bangladesh by Aktor and Rahman (2010); Robert et al. (2014). This result contradicts with the study done in Goba district, southern Ethiopia by Tesfaye (2012). Although mothers with higher education are more likely to know benefits of EBF, they are more likely to be employed as well. At the same time, women may be influenced by media advertising milk substitutes. The vast majority of women breast feed their children for a short time, and then cease breast-feeding exclusively too early for the aforementioned reasons. The inverse relationship between maternal level of general education and breastfeeding is also found in other developing countries but is in contra-

distinction to the western countries particularly. For example the study done in the USA by Ryan (2007) and Sweden by Persson (1985) show higher maternal education is related to higher rates of EBF (Ryan, 2007).

The use of prelacteal feeding is another current deterrent impeding the promotion of EBF in many developing countries (Ssenyonga,2004;Engebretsen,2007) including in Ethiopia as is evidenced by the ongoing practice of feeding other than breast milk within the first three days of newborn. The explanations given in the current study was that mothers believe that they need to wait until the milk started flowing, secondly giving liquid will clean the baby's throat; we agree with the statement stated by Engebretsen , 2007. When looked at the given explanation, it appears that the habit harms the newborn and exposes him/her to various morbidities and therefore, the behavior needs to be discouraged.

Factors that interact with the protective effect of breastfeeding include environmental, cultural and economic characteristics. The protective effect of breastfeeding is most important in populations with high infant mortality, high illiteracy, poor sanitation facilities, poor nutritional status, and generally low economic status (Arvalhaeset al., 2007), situation similar to that in Ethiopia. So in Ethiopia, if optimal breastfeeding is going to be practiced, considerable changes could be achieved in the child morbidity and mortality.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusions

In this study we have investigated factors associated with duration of EBF in Ethiopia by using the Cox PH model, Cox PH with shared gamma and inverse Gaussian frailty models. Out of the total of 1371 infants 813 (59.3%) of infants were introduced with breast milk substitutes before six months of age and 558 (40.7%) of infants were exclusively breast feed until the end of the six months of age. The maximum duration of EBF was six and the minimum was zero months. The median duration of EBF was five months. The Cox PH with gamma frailty model with the covariates place of residence, economic status (wealth index), contraceptive use, education level and marital status was a better fit model to explain the data well. Variables that were found to be significantly associated with duration of EBF in the fitted Cox PH with gamma shared model were place of residence, economic status (wealth index), contraceptive use, and education level. The heterogeneity parameter found to be significant, indicating that that the correlation within regions should have to be considered.

5.2. Recommendations

Based on the findings of the study different factors were identified for short duration of EBF. In summary the key recommendations emerging from this study for policy makers, clinicians and the public at large are:

- Since urban infants are more exposed to short duration of EBF in urban special attention is expected from concerned body in order to encourage breast feeding culture. Women should be discouraged during pregnancy from introducing breast-milk substitutes while in the maternity ward.
- As the findings show, when mothers' education level increases, duration of EBF is lower and lower. One factor affecting breastfeeding duration is that many educated mothers are away from their children during the day due to work. Government and private employers should have to give long enough maternity leave to encourage EBF. Support for breastfeeding should be provided in all places, including early care and education facilities. Mothers should aware that to prepare pumped breastfeed at the facility, feeding a mother's pumped breast milk to her baby, thawing and preparing bottles of pumped

milk, and keeping extra breast milk in a freezer. The law also requires employers to grant nursing mothers a "break period" so they can use the lactation stations properly.

- Those contraceptive user mothers fed exclusively breast milk longer duration relative to non-user. Thus, health professionals should have to advise mothers to use comfortable contraceptive method after delivery to avoid early pregnancy at prenatal and post natal follow up time.
- Women, who are poor, are less likely to choose to breast feed. Since in Ethiopia most of the populations are under poverty and most of infants are from poor mothers' strategies required by any government and individuals to campaign, to focus and facilitate actions to protect, promote and support breastfeeding by supporting the poor.
- There was regional variability in duration of EBF. So that support to protect and promote EBF should be based on the magnitude of the problem in the different region. The regions which have discouraging breast feeding culture need more attention.

REFERENCES

Aarts C. Exclusive breastfeeding – Does it make a difference? Uppsala, Sweden: Uppsala University,2005.

Abate G, Kogi-Makau W, Muroki N. Child-feeding practices as predictors of nutritional status of children in a slum area in Addis Ababa, Ethiopia.*Ethiop J Health Dev*. 1999; 13(3):229-38.

Abiyot, N. (2013). Modeling Time-to-Recovery of Adult Diabetic Patients: A Case Study of Jimma University Specialized Hospital.

Agampodi , S., Agampodi, T., Kankanamge U., Piyaseeli D.: Breastfeeding practices in a public health field practice area in Sri Lanka: a survival analysis.*Int Breastfeed J*. 2007; 2:13.

Akter and Rahman. Duration of Breastfeeding and Its Correlates in Bangladesh. *J health popul nutr* .2010 Dec;28(6):595-601.

Alemayehu, Haider and Habte. Determinants of exclusive breastfeeding practices in Ethiopia. *Ethiop.J.Health Dev*. 2009;23(1).

Andersen, P., Klein, J., Knudsen, K. and Palacios, R. (1997). Estimation of variance in Cox's reg ression model with shared gamma frailties. Biometrics 53: 1475-1484.

Arifeen S., Black R., Antelman G., Baqui A., Caulfield L., Becker S. Exclusive breastfeeding reduces acute respiratory infection and diarrhoea deaths among infants in Dhaka slums. *Pediatrics.*2001; 108: E67.

Becquet R et al. Acceptability of exclusive breast-feeding with early cessation to prevent HIV transmission through breast milk, ANRS 1201/ 1202 Ditrame Plus, Abidjan Cot e d'Ivoire. *Journal of Acquired Immune Deficiency Syndromes*.2005, 1–9.

Belay, B. (2011).Modelingtime to malaria: a comparison of Cox proportional hazard and Gamma frailty models.

Bloss E, Wainaina F, Bailey RC. Prevalence and predictors of Underweight, Stunting, and Wasting among Children Aged 5 and Under in Western Kenya. *Journal of Tropical Pediatrics*. 2004;50(5):260-70.

Black R. et al. Maternal and child undernutrition: global and regional exposures and health consequences. (Maternal and Child Undernutrition Series 1). *The Lancet*. 2008; 13:16.

Bøhler E., Aalen O., Bergstrøm S., Halvorsen S. Breast feeding and seasonal determinants of child growth in weight in east Bhutan. *ActaPaediatr*. 1995 Sep; 84(9):1029-34.

NIHPA Author Manuscripts. Breastfeeding Patterns, Time to Initiation and Mortality Risk Among Newborns in Southern Nepal. Mullany L etc al. *The Journal of Nutrition*.2008; 138: 599-603.

Brunken, G., Silva , S., França ,G., Escuder, M., Venâncio, S. Risk factors for early interruption of exclusive breastfeeding and late introduction of complementary foods among infants in midwesternBrazil.*Jornal de Pediatria*.2006; 82(6):445-51.

Carvalhaes,M., Parada, C., Costa, M.. Factors associated with exclusive breastfeeding in children under four months old in Botucatu-SP, Brazile. *Rev Latino-am Enfermagem*. 2007; 15(1):62-9.

Coovadia, H. et al. Mother-to-child transmission of HIV-1 infection during exclusive breastfeeding: the first six months of life. *Lancet.*2007; 369:1107-1116.

Cox (1972). Regression models and life tables (with discussion) Journal of the Royal Statistical Society B.

Cox and Snell (1968). A general definition of residuals (with discussion).

Cox, D.and Oakes, D. (1984). Analysis of survival data. London: Chapman and Hal.

CSA, ORC M. *Ethiopia demographic and health survey 2005*. Addis Ababa, Ethiopia and Calverton, Maryland. USA: central statistics agency and ORC Macro; 2006.

CSA [Ethiopia]: *Ethiopia Demographic and Health Survey 2011*. Addis Ababa, Ethiopia, Calverton, Maryland, USA: Central statistical agency and ICF International; 2012.

Department of Census and Statistics in collaboration with Ministry of Healthcare and Nutrition:Sri Lanka Demographic and Health Survey 2006/7.Sri Lanka: Department of Census and Statistics; 2008.

Kloos,H.The Epidemiology and Ecology of Health and Diseases in Ethiopia. 2006. Shama Books: Addis Ababa, Ethiopia.

Duchateau, L. and Janssen, P. (2004). Penalized partial likelihood for frailties and smoothing splines in time to first insemination models for dairy cows. Biometrics interrnational Statistical Review.

Edmond, K et al. Delayed Breastfeeding Initiation Increases Risk of Neonatal Mortality. Pediatrics.2006 Mar;117(3):e380-6.

Edson N., Eliana Z., Adauto E.Analysis models for variables associated with breastfeeding duration.*Rev Paul Pediatr*.2013;31(3):306-14.equences. The Lancet 2008; 371(9608): 243-260.

Egata, Berhane and Worku. Predictors of non-exclusive breastfeeding at 6 months among rural mothers in east Ethiopia: a community-based analytical cross-sectional study. *International Breastfeeding Journal*.2013, 8:8.

Engebretsen, I., Wamani, H., Karamagi,C., Semiyaga, N., Tumwine, J., Tylleskär, T. Low adherence to exclusive breastfeeding in Eastern Uganda: A community-based cross-sectional study comparing dietary.

Engebretsen, I., Tylleskär, T., Wamani, H., Karamagi, C., Tumwine, J. Determinants of infant growth in Eastern Uganda: a community-based cross-sectional study. BMC Public Health. 2008 Dec 22;8:418.

E. Robert, Y. Coppieters, B. Swennen, M. Dramaix. Breastfeeding Duration: A survival analysis data from a regional immunization survey. *BioMed Research International*. 2014 ; Article 8 pages (http://dx.doi.org/10.1155/2014/529790).

Federal MoH. National Strategy for Child Survival in Ethiopia. Addis Ababa, Ethiopia; 2005.

Federal MoH. Health sector development program IV woreda based annual core plan. Addis Ababa, Ethiopia; 2010.

Foo, L, Quek ,S., Lim, M., Deurenberg-yap M. Breastfeeding prevalence and practices among Singaporean Chinese, Malay and Indian mothers. *Health Promotion International*. 2005;20(3):229-37.

Foo, L, Quek, S., Lim MT, Deurenberg-yap M. Breastfeeding prevalence and practices among Singaporean Chinese, Malay and Indian mothers. *Health PromotionInternational*. 2005;20(3):229-37.

Green, P. J. (1987). Penalized likelihood for general semi-parametric regression models.

Grummer-Strawn, L., and Z. Mei. Does Breastfeeding Protect Against Pediatric Overweight: Analysis of longitudinal data from the Centers for Disease Control and Prevention Pediatric Nutrition Surveillance System. *Pediatrics*.Feb. 2004; 113(2): 81–86.

Girma, W., Genebo, T. Determinants of Nutritional Status of Women and Children in Ethiopia. Calverton, Maryland, USA: ORC Macro; 2002.

Haider, J., Kloos, H., Haile Mariam, D. and Demissie T. "Food and Nutrition" In Berhane Y, Haile Mariam.2002.

Harder T, Bergmann R, Kallischnigg G, Plagemann A. Duration of breastfeeding and risk of overweight: a meta-analysis. *Am J Epidemiol*. 2006;163(9):870-2.

Hougaard, P.(1986). Survival models for heterogeneous populations derived from stable distributions. Modeling Time-to-Recovery of Adult Diabetic Patients : A Comparison of Cox-PH and Shared Frailty Models 2014.

Hougaard, P. (1995). Frailty models for survival data. Lifetime Data Analysis. Gutierrez, R. G., S. Carter, and D. M. Drukker. (2001). on boundary-value likelihood-ratio tests.

Iliff P et al. Early exclusive breastfeeding reduces the risk of postnatal HIV-1 transmission and increases HIV-free survival. AIDS, 2005; 19:699-708.

John, F, (2002). Cox Proportional-Hazards Regression for Survival Data Appendix to Rand S-PLUS Companion to Applied Regression.

Joly, P., Commenges, D., Letenneur, L. (1998). A penalized likelihood approach for arbitrarily censored and truncated data.

Jones, M., Swerdlow, A., Gill, L., et al. Pre-natal and early life risk factors for childhood onset diabetes mellitus: A record linkage study. *Int J Epidem*.1998;27:444– 9.

Jones G, Steketee, R., Black, R., Bhutta Z., Morris, S., Group B. How many child deaths can we prevent this year? *Lancet*. 2003;362:65-71.

Kaplan and Meier. (1958). Non-parametric estimation from incomplete observations. Journal of American Statistical Association.

Kalbfleisch, J. , and Prentice, R. (1963). Marginal likelihoods based on Cox's regression and life model.

Lamberti, L., Fischer Walker, C., Noiman, A., Victora C, Black R. Breastfeeding and the risk for diarrhea morbidity and mortality. *BMC Public Health*.2011;11 (Suppl 3):S15.

Lawrence, R. Breastfeeding: A Guide for the Medical Profession, 1994.

Lauer, J, Betrán, A., Barros, A., de Onís, M.: Deaths and years of life lost due to suboptimal breast-feeding among children in the developing world: a global ecological risk assessment**.** *Public Health Nutr* 2006, **9**(6)**:**673-685.

Long, K, Wood, J., Gariby, E., Weiss, K., Mathewson, J., de la Cabada, F., et al. Proportional hazards analysis of diarrhea due to Enterotoxigenic Escherichia coli and breastfeeding in a cohort of urban Mexican children. *Am J Epidem*.1994;139:193– 205.

Luc Duchateau and Paul Janssen. The Frailty Model. SPIN Springer's internal project number, if known. – Monograph –. August 12, 2007. Springer.

McGilchrist, C. (1993). REML estimation for survival models with frailty. Biometrics, 49, 221-25.

Medical eligibility criteria for contraceptive use, Third Edition. Geneva: World Health Organization; 2004.

Mikiel-Kostyra, K., Mazur, J., Wojdan-Godek, E. Factors affecting exclusive breastfeeding in Poland: cross-sectional survey of population-based samples. Soz*Praventivmed*.2005;50:52-9.

Monotoring the situation of childer and women .New York:UNICEF; website:http://www.childinfo.org(accessed on 17 August, 2007).

Morisky, D., Kar, S., Chaudhry, A., Chen, K., Shaheen, M., Chickering, K. Breast Feeding Practices in Pakistan. *Pakistan Journal of Nutrition*. 2002;1(3):137-42.

Mihrshahi, S., Ichikawa, N., Shuaib, M., Oddy, W., Ampon, R., Dibley, M., et al. Prevalence of Exclusive Breastfeeding in Bangladesh and Its Association with Diarrhoea and Acute Respiratory Infection: Results of the Multiple Indicator Cluster Survey. 2003. *J Health PopulNutr*.2007 Jun;25(2):195-204.

Nielsen, G., Gill, R., Andersen, P., Sorensen, T. (1992). A counting process approach to maximum likelihood estimation in frailty models.

Nutrition. New York: UNICEF; web site: http://www.unicef.org/nutrition/index.html (accessed on 17 August 2007).

Nwankwo B., Brieger W. Exclusive breastfeeding is undermined by use of other liquids in rural Southwestern Nigeria. Journal of Tropical Pediatrics. 2002;48:109-12.

O'Sullivan, F. (1988). Fast computation of fully automated log-density and log-hazard estimators.

Owen, C., Whincup, P., Kaye, S., Martin R., Davey Smith, G, Cook, D. Does initial breastfeeding lead to lower blood cholesterol in adult life? A quantitative review of the evidence. *Am J Clin Nutr*.2008;88:305-14.

Perera, P, Fernando M, Warnakulasuria T, Ranathunga N:Feeding practicesamong children attending child welfare clinics in Ragama MOH area: adescriptive cross-sectional study.*Int Breastfeed J*.2011;6:18.

Perera, Ranathunga, Fernando, Sampath and Samaranayake. Actual exclusive breastfeeding rates and determinants among a cohort of children living in Gampaha district Sri Lanka: A prospectiveobservational study. *International Breastfeeding Journal*.2012; 7:21.

ParadaCMGdL, CarvalhaesMAdBL, Jamas MT. Complementary feeding practices to children during their first year of life. *Rev Latino-am Enfermagem*. 2007;15(2):282-9.

Peters E, Wehkamp K-H, Felberbaum, R., ger, D., Linder, R. Breastfeeding duration is determined by only a few factors. *European Journal of Public Health*. 2005;16(2):162-7.

Quantifying the benefits of breastfeeding: a summary of the evidence. Washington, DC: Pan American Health Organization; 2002.

Robert B., Maternal and Child Undernutrition 1: global and regional exposures and health cons, 2007.

Roberto G, Joel, A, Cibele, C. Factors associated with duration of breast feeding.*Jornal de Pediatria* .2007; 241:240-245.

Ryan, A. The resurgence of breastfeeding in United States. *Pediatrics*. 1997;99:12-14.

Santos Neto ET, Oliveira AE, Zandonade E, Molina MD. Pacifier use as a risk factor for reduction in breastfeeding duration: a systematic review. *Rev Bras Saude Mater Infant* 2008;8:377-89.

Scrimshaw, N., Taylor, C, Gordon, J. Interactions of nutrition and infection. Geneva: World Health,Organization, 1968.ecall since birth with 24-hour recall. *BMC Pediatrics*. 2007;7(10).

Setegn ,Tefera, Gerbaba, Deribe, Deribew and Biadgilign. Factors associated with exclusive breastfeeding practices among mothers in Goba district, south east Ethiopia: a cross-sectional study. *International Breastfeeding Journal* .2012;7:17.

Simopoulos AP, Grave GD. Factors Associated with the Choice and Duration of Infant-Feeding Practice. *Pediatrics*.1984; 74:603-14.

Shirima, R., Greiner, T., Kylberg, E., Gebre-Medhin, M. Exclusive breast-feeding is rarely practiced in rural and urban Morogoro, Tanzania. *Public Health Nutrition*. 2000;4(2):147-54.

Shiva, F., Nasiri, M. Study of Feeding Patterns in Young Infants. *Journal of Tropical Pediatrics*. 2003; 49(2):89-92.

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model.

Ssenyonga, R., Muwonge, R., Nankya, I. Towards a Better Understanding of Exclusive Breastfeeding in the Era of HIV/AIDS: A Study of Prevalence and Factors Associated with Exclusive Breastfeeding from Birth, in Rakai, Uganda.*Journal of Tropical Pediatrics*. 2004;50(6):348-53.

The BHITS Group. Late Postnatal Transmission of HIV-1 in breast-fed children: An individual Patient data meta-analyses. *Journal of Infectious Diseases*.2004, 189:2154–2166.17.

Therneau, T.and P.M. Grambsch (2000). Modelling Survival Data: Extending the Cox Model. Springer-Verlag, New York.

Therneau, T. and P.M. Grambsch (2002). Penalized Cox models and frailty. Working manuscript.

Trussel and Richards (1983). Correction for unobserved heterogeneity in hazard models.Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality.Demography.

Tampah-Naah and Kumi-Kyereme. Determinants of exclusive breastfeeding among mothers in Ghana: a cross-sectional study. *International Breastfeeding Journal* .2013; 8:13.

UNICEF:Progress for children, report card on nutrition: number 4.2006, http://www.unicef.org/progressforchildren/2006n4/index_breastfeeding.html.

Victora, C., Matijasevich, A., Santos, I., Barros, A., Horta, B., Barros FC. Breastfeeding and feeding patterns in three birth cohorts in Southern Brazil: trends and differentials. *Cad SaudePublica*.2008;24 (Suppl 3):S409-16.

WHO. Indicators for assessing breast-feeding practices. Report of an informal meeting. Geneva, Seizerland;11-12 ,June1991.

World Health Organization. The World Health Organization's infant feeding recommendation. Geneva: WHO; 1995.

WHO. Nutrient adequacy ofexclusive breastfeeding for the term infant during the first six months of life. Geneva; 2002.

WHO. Global strategy for infant and young child feeding. The optimal duration of exclusive breastfeeding. Geneva; 2001.

WHO. The optimal duration of exclusive breastfeeding: a systematic review. Geneva; 2002.

WHO Collaborative Study Team on the role of breastfeeding on the prevention of infant mortality, effect of breastfeeding on infant and child mortality due to infectious diseases in less developed countries: a pooled analyses. *Lancet.*2000;355:451–455.

WHO: HIV and infant feeding: update based on the technical consultation held on behalf of the Inter-agency Team (IATT) on Prevention of HIV Infections in Pregnant Women, Mothers and their Infants, Geneva, 25-27 October 2006.

Yashin, A., Iachine, A., Alexander, Z., and Vaupel, W. (2000). Hidden Frailty: Myths and Reality.

Yoon, P., Black ,R., Moulton, L., Becker, S. Effect of Not Breastfeeding on the Risk of Diarrheal and Respiratory Mortality in Children under 2 Years of Age in Metro Cebu, the Philippines. *American Journal of Epidemiology.*1996; 143:1142-8.

Zelterman, D. (1992). A statistical distribution with an unbounded hazard function and its application to a theory from demography. Biometrics.

APPENDIX

Table 7: Results of the Log-rank test for the categorical variables of duration of EBFunder six months of age children in Ethiopia, EDHs, 2011.

Covariates	DF	Chi-square	P-value
Residence	1	3.100	0.078*
Contraceptive	1	8.300	0.004*
Education	2	3.500	0.173
Wealth	2	9.100	0.010*
Religion	3	26.400	0.001*
Mothers age	2	0.900	0.624
work status	1	3.500	0.062*
Place of delivery	1	2.50	0.111
Marital Status	2	1.5	0.47

P-value: probability value, * Significant at 0.1 level.

Table 8: Results of univariable Cox PHmodel

Covariates	HR	Coef($\widehat{\beta}$)	SE	P. value	95% CI of HR
Residence					
Rural	0.845	-0.168	0.090	0.062*	[0.709,1.010]
Contraceptive					
Non-user	1.318	0.276	0.116	0.017 *	[1.051,1.655]
Wealth					
Middle	0.905	-0.100	0.111	0.365	[0.728,1.124]
Poor	1.156	0.145	0.078	0.062*	[0.992,1.346]
Women's age					
Between20&35	1.04816	0.047	0.125	0.709	[0.819, 1.342]
Between36&49	1.03096	0.030	0.155	0.844	[0.7605,1.398]
Education level					
Primary	1.088	0.084	0.080	0.293	[0.929,1.273]
Secondary& higher	1.289	0.254	0.139	0.069*	[0.980,1.694]
Place of delivery					
Home	0.929	-0.074	0.071	0.299	[0.807,1.068]
Religion					
Orthodox	0.636	-0.452	0.089	0.001*	[0.534 ,0.758]
Protestant	0.888	-0.118	0.092	0.196	[0.742,1.063]
Others	0.994	-0.006	0.223	0.980	[0.642,1.540]
Work Status					
Not working	0.895	-0.109	0.085	0.197	[0.758,1.059]

Coef: coefficient for covariate, SE: standard error of coefficient, HR: hazard ratio; p-value: probability value, 95%C.I HR: 95% confidence interval for HR, * Significant at 0.1 level.

Table 9: Estimates of multivariable Cox PH model

Covariates	Coeff	SE	P-value	HR	95% CI of HR
Contraceptive					
Non-user	0.335	0.129	0.009*	1.398	[1.085,1.800]
Religion					
Orthodox	-0.487	0.092	0.000*	0.614	[0.513,0.735]
Protestant	-0.137	0.094	0.149	0.872	[0.725,1.047]
Others	0.025	0.224	0.911	1.025	[0.660,1.591]
Residence					
Rural	-0.413	0.119	0.001*	0.661	[0.524,0.836]
Wealth					
Middle	0.109	0.125	0.383	1.115	[0.873,1.424]
Poor	0.351	0.099	0.001*	1.420	[1.170,1.725]
Education level					
Primary	0.254	0.088	0.001*	1.289	[1.086,0.531]
Secondary&higher	0.389	0.161	0.016*	1.475	[1.076,2.022]
Likelihood ratio test=62.9 on 9 DF, p-value =000.					

Coef: coefficient for covariate, SE: standard error of coefficient, HR: hazard ratio; p-value: probability value, 95%C.I HR: 95% confidence interval for HR, * Significant at 0.05 level.

Table 10: Standard error and corresponding p-values of possible interaction terms, added one at a time, to the variables included in the model in Table 5.

Interaction between	Covariates/factors	Coef	SE	p-value
Contraceptive	Religion			
	Orthodox	-0.159	0.261	0.540
	Protestant	0.171	0.356	0.630
	Others	0.445	1.045	0.670
	Residence	-0.034	0.249	0.890
	Wealth			
	Middle	0.322	0.418	0.440
	poor	0.103	0.369	0.780
	Education level			
	Primary	0.356	0.277	0.200
	secondary&high	0.147	0.361	0.680
Residence	Religion			
	Orthodox	-0.551	0.205	0.071
	Protestant	0.308	0.326	0.340
	Others	-1.813	1.037	0.080
	Wealth			
	Middle	0.242	0.601	0.690
	poor	0.219	0.383	0.570
	Education level			
	Primary	-0.025	0.217	0.910

	Secondary&higher	0.166	0.353	0.640
Education level	Wealth			
Primary	Middle	-0.002	0.248	0.99
	Poor	-0.012	0.185	0.95
	Wealth			
	Middle	0.025	0.001	0.043
Secondary & higher	Poor	0.974	0.489	0.046

Coef: coefficient for covariate, SE: standard error of coefficient, HR: hazard ratio; p-value: probability value, * Significant at 0.5 level.

Table 11: Results of the multivariable PH Cox regression model containing the variables in Table 4 and their interaction with log time (in months).

Covariates	Rho	Chi-square	P.value
Residence			
Rural	0.057	2.626	0.105
Contraceptive			
Non-user	-0.098	7.880	0.450
Wealth			
Middle	-0.024	0.471	0.492
Poor	-0.055	2.442	0.118
Education level			
Primary	-0.03905	-1.261	0.261
Secondary & higher	0.00724	0.043	0.834
Religion			
Orthodox	0.0189	0.286	0.592
Protestant	-0.032	0.818	0.366
Others	0.031	0.789	0.374
GLOBAL	NA	14.880	0.094

P-value: probability value, * Significant at 0.5 level.

Table 12: Estimates of univariable Cox PH with shared gamma frailty model

Covariates	HR	Coef($\hat{\beta}$)	SE	P.value	95% CI of HR	χ^2	Prob>χ^2
Residence							
Rural	0.865	-0.145	0.104	0.160	[0.706,1.061]		
Theta						88.500	1.8e-14
Contraceptive							
Non-user	1.287	0.253	0.124	0.042*	[1.009,1.643]		
Theta						87.460	2.9e-14
Wealth							
Middle	1.005	0.005	0.115	0.97	[0.802,1.260]		
Poor	1.105	0.099	0.086	0.25	[0.932,1.309]		
Theta						84.840	9.0e-14
Women's age							
Between20&35	1.031	0.031	0.127	0.81	[0.803,1.324]		
Between36&49	0.956	0.045	0.204	0.83	[0.641,1.427]		
Theta		2				89.360	1.2e-14
Education level							
Primary	1.154	0.144	0.086	0.094*	[0.976,1.365]		
Secondary&higher	1.383	0.324	0.147	0.027*	[1.037,1.844]		
Theta			3			92.060	3.7e-15
Religion							
Orthodox	0.8638	-0.1464	0.1275	0.25	[0.673,1.109]		
Protestant	0.9603	-0.0405	0.1518	0.79	[0.713,1.293]		
Others	1.0475	0.0464	0.2478	0.85	[0.644,1.703]		
Theta		1				56.48	1.6e-08
Work							
Yes	0.940	-0.061	0.087	0.480	[0.792,1.116]		
Theta						50	1.7e-14

Coef: coefficient for covariate, SE: standard error of coefficient, HR: hazard ratio; p-value: probability value, 95%C.I HR: 95% confidence interval for HR, * Significant at 0.1 level.

Table 13:Results of multivariable Cox PH with shared gamma frailty model

Covariates	HR	Coef($\hat{\beta}$)	Std.Error	P.value	95% CI of HR
Contraceptive					
Non-user	1.395	0.333	0.133	0.013*	[1.074,1.811]
Education level					
Primary	1.298	0.261	0.093	0.005*	[1.081,1.559]
Secondary&higher	1.555	0.441	0.165	0.007*	[1.125,2.149]]
Likelihood ratio test of Theta = 132: chisqr=43.51 prob>= chisqr=0.000014.					

Coef: coefficient for covariate, SE: standard error of coefficient, HR: hazard ratio; p-value: probability value, 95%C.I HR: 95% confidence interval for HR, * Significant at 0.05 level.

Table 14: Results of univariable Cox PH with inverse Gaussian frailty model

Covariates	HR	Coef($\hat{\beta}$)	SE	P.value	95% CI of HR	χ^2	Prob>= χ^2
Residence							
Rural	0.860	-0.151	0.102	0.140	[0.704,1.050]		
Theta		0.119				85.36	2.7e-14
Contraceptive							
Non-user	1.279	0.246	0.123	0.042*	[1.005,1.629]		
Theta		0.124				83.41	7.1e-14
Wealth							
Middle	0.993	-0.007	0.115	0.950	[0.793,1.243]		
Poor	1.103	0.098	0.085	0.250	[0.933,1.304]		
Theta		0.119				80.91	2.1e-13
Women's age							
20-35	1.034	0.033	0.127	0.800	[0.805,1.326]		
36-49	0.952	-0.049	0.204	0.810	[0.638,1.420]		
Theta		0.121				85.780	2.4e-14
Education level							
Primary	1.154	0.143	0.085	0.093*	[0.976,1.363]		
Sec& higher	1.381	0.328	0.145	0.027*	[1.037,1.838]		
Theta		0.124				88.750	6.2e-15
Placedelivery							
Health instute	0.942	-0.060	0.074	0.420	[0.814,1.090]		
Theta		0.122				85.160	3.2e-14
Religion							
Orthodox	0.835	-0.179	0.122	0.140	[0.657,1.063]		
Protestant	0.954	-0.046	0.145	0.750	[0.717,1.271]		
Others	1.042	0.042	0.245	0.860	[0.645,1.686]		
Theta		0.105				53.690	2.5e-08
Work							
Yes	0.9385	-0.11	0.087	0.470	[0.791,1.113]		
Theta		0.121				85.060	3.3e-14
Marital status							
partner	1.2372	0.213	0.160	0.180	[0.903,1.695]		
Others	0.8254	-0.192	0.185	0.300	[0.573,1.188]		
Theta		0.125				87.930	9.2e-15

Table 15: Estimates of multivariable shared inverse Gaussian frailty model

Covariates	HR	Coef($\hat{\beta}$)	Std.Error	P.value	95% CI of β	
Contraceptive						
Non-user	1.4207	0.35115	0.13367	0.0086*	1.0933	1.8462
Education level						
Primary	1.3136	0.27278	0.09352	0.0035*	1.0936	1.5779
Secondary & higher	1.5518	0.43940	0.16467	0.0076*	1.1237	2.1429

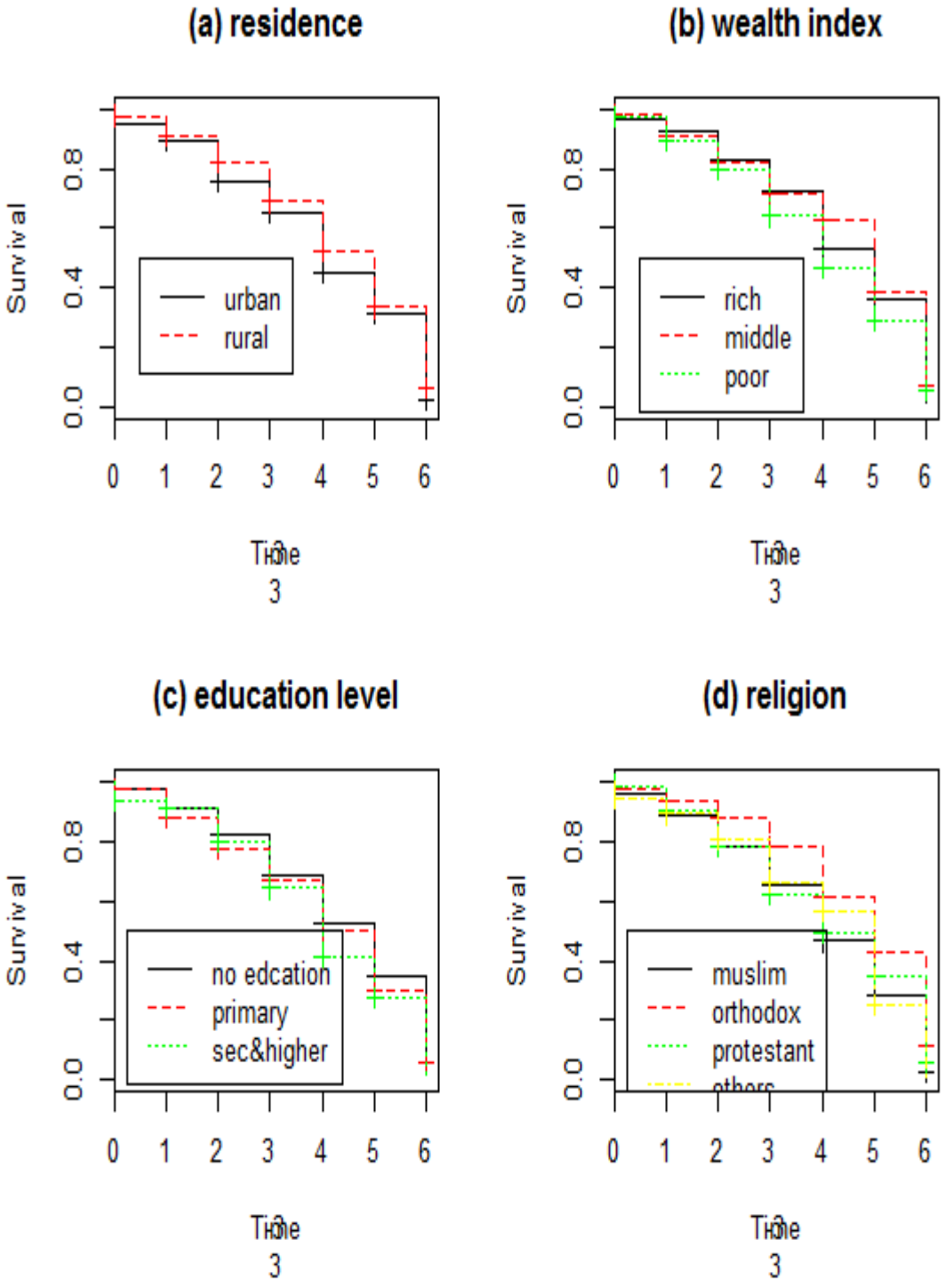
Likelihood ratio test of theta = 130.9: χ^2 =42.97 prob>= χ^2=0. 000018

Coef: coefficient for covariate, SE: standard error of coefficient, HR: hazard ratio; p-value: probability value, 95% C.I HR: 95% confidence interval for HR, * Significant at 0.05 level.

Table 16: Results of the multivariable PH Cox regression model for shared gamma frailty model containing the variables in Table 4 and their interaction with log time (in months)

Covariates	Rho	χ^2	P.value
Residence			
Rural	0.048	2.120	0.14535
Contraceptive			
Non-user	-0.094	7.782	0.40527
Wealth			
Middle	-0.031	0.808	0.36849
Poor	-0.057	2.894	0.08888
Education level			
Primary	-0.038	1.3251	0.24967
Secondary&higher	0.007	0.0456	0.83096
Marital status			
partner	0.0054	0.0255	0.87307
Others	0.013	0.1443	0.70402
GLOBAL	NA	12.4562	0.13198

P-value: probability value, * Significant at 0.05 level.


(a) residence
Survival
0.0
0.4
0.8
urban
rural
0
1
2
3
4
5
6
Time
(b) wealth index
Survival
0.0
0.4
0.8
rich
middle
poor
0
1
2
3
4
5
6
Time
(c) education level
Survival
0.0
0.4
0.8
no edcation
primary
sec&higher
0
1
2
3
4
5
6
Time
(d) religion
Survival
0.0
0.4
0.8
muslim
orthodox
protestant
0
1
2
3
4
5
6
Time

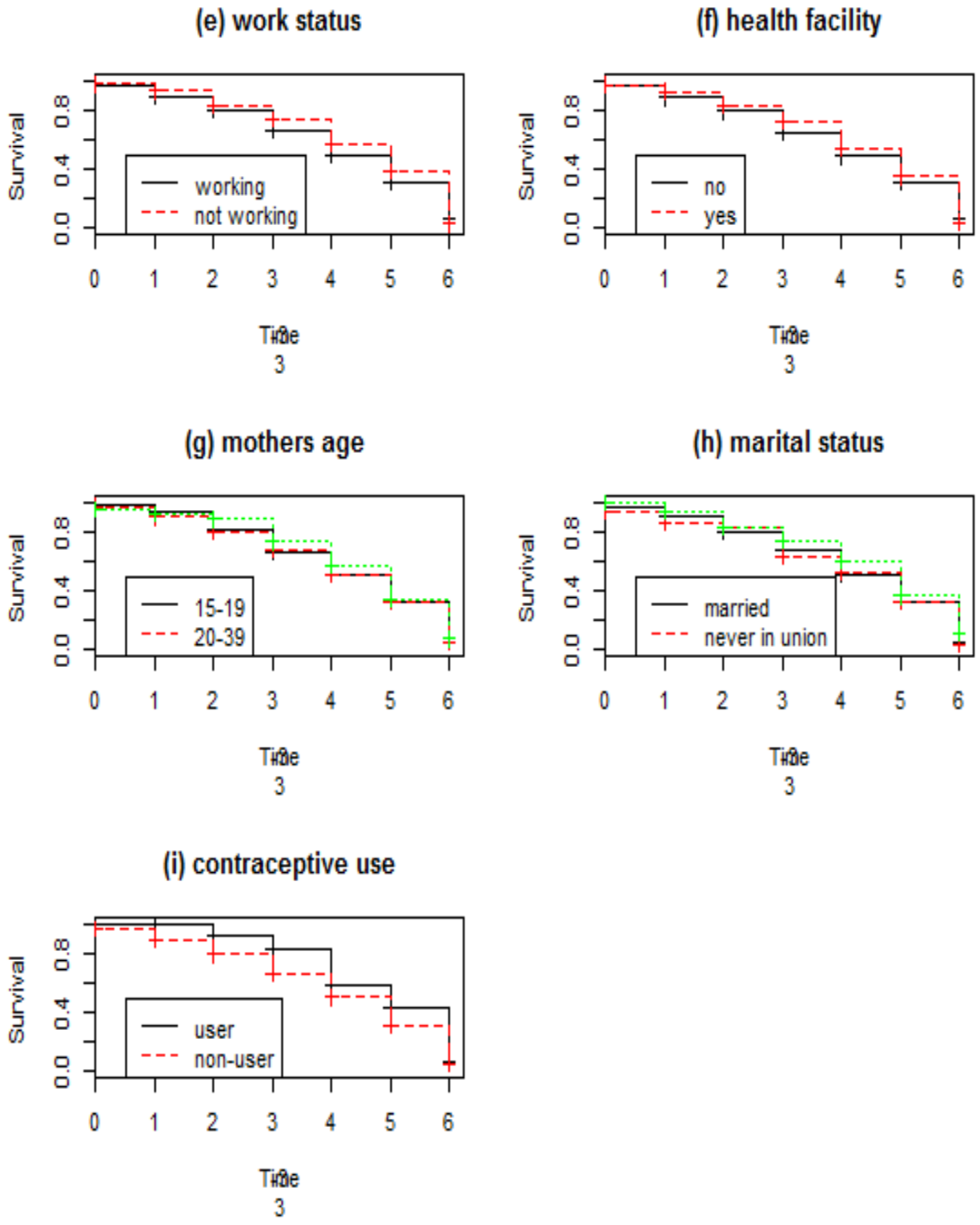


Figure 5 (a – i): Plots of Kaplan-Meier survivor functions based on different categories of covariates, of duration of exclusive breast feeding data taken from EDHS (2011).

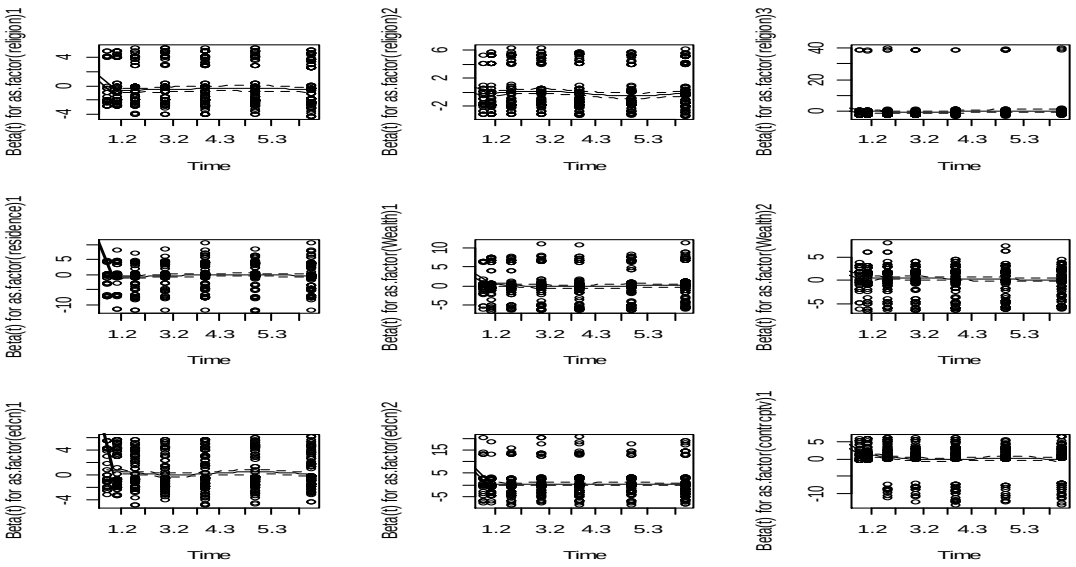


Figure 6: Plots of scaled Schoenfeld residuals against transformed time for each covariate in a model fit to the EDHs breast feeding data. The solid line is a smoothing-spline fit to the plot, with the broken lines representing a±2-standard-error band around the fit.

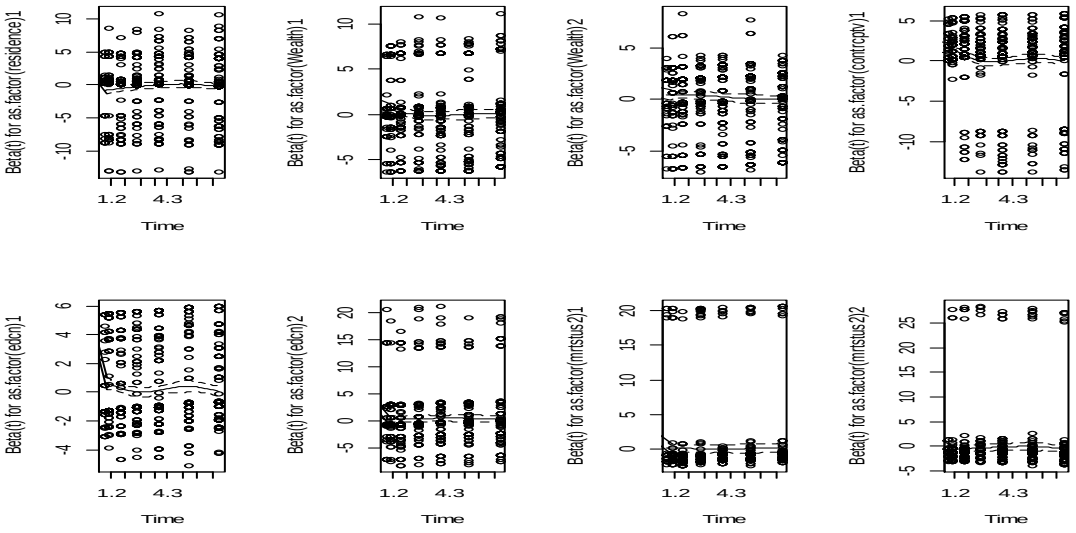


Figure 8: Plots of scaled Schoenfeld residuals against transformed time for each covariate in a model fit to the EDHs breast feeding data in shared gamma frailty model. The solid line is a smoothing-spline fit to the plot, with the broken lines representing a±2-standard-error band around the fit.