



Kandidatuppsats

Statistiska institutionen

Bachelor thesis, Department of Statistics

Nr 2021:2

Multilevel Cox Regression of Transition to Parenthood among Ethiopian Women

**Flernivå-coxregression av kvinnors övergång till
föräldraskap i Etiopien**

Amanda Akinyi Lagehäll och Elelta Yemane

Självständigt arbete 15 högskolepoäng inom Statistik III, VT2021

Handledare: Gebrenegus Ghilagaber

ABSTRACT

The birth of the first child is a special event for a mother whose life can change dramatically. In Ethiopia women's timing to enter motherhood vary between the regions. This paper is therefore focusing on how birth cohort, education and residence affect the rate of entering motherhood for Ethiopian women in the different regions and the entire country. The dataset is extracted from the 2016 Ethiopia Demographic and Health Survey (EDHS) and contains 15,019 women from 487 different households. For more accurate estimations and results, the correlation within households is taken into consideration with multilevel survival analysis. The methods used are the Cox proportional hazard model and two frailty models. The results of the paper show that women residing in rural areas have an increased rate of entering motherhood compared to those residing in urban areas, every age group older than those born 1997 to 2001 have a higher intensity to enter parenthood and those with education have a decreased intensity ratio compared to the women with no education. It also shows that there is a regional difference in the effect of the estimated ratios of the covariates. Performing the multilevel analysis only changes the estimated effects of the covariates in the cities and one region. It is concluded that the estimated intensity ratio of multilevel survival analysis only varies from the standard Cox regression when the region is heterogeneous.

PREFACE

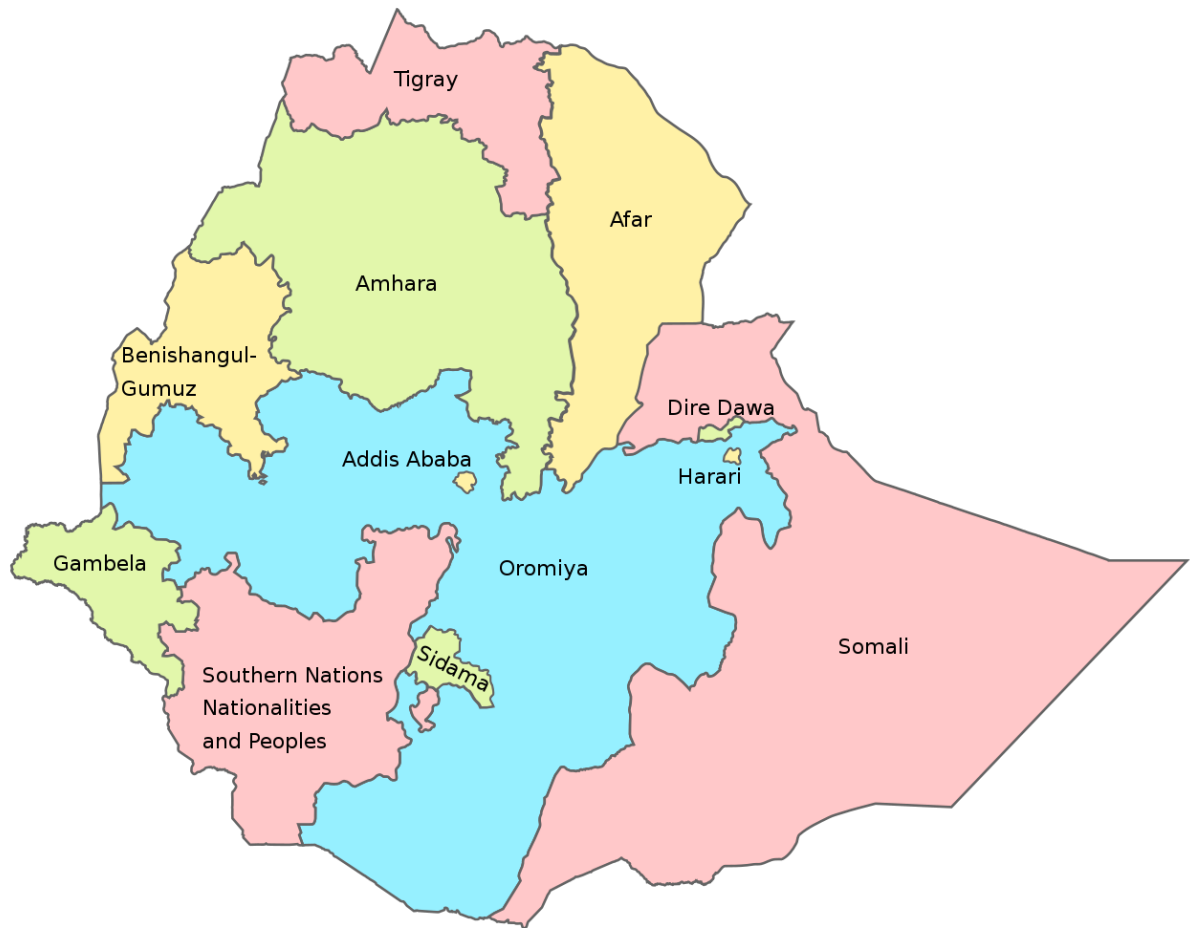
This paper would not have been possible without our supervisor Gebrenegus Ghilagaber who has generously given out his time to help us with this paper since the beginning. His guidance through the process of finding good eligible data, cleaning up data and writing an empirical study has been insightful and joyful. We cannot thank you enough.

We would also like to take the opportunity to thank Carren Melinder for tirelessly reading through numerous drafts, helping us through the process. Her help has been invaluable.

Last but not least a big thanks to our families and friends who have supported us.



Picture: A map of Africa



Picture: A map of Ethiopia's regions.

Contents

1. Introduction	1
1.1 Background	1
1.2 Previous studies	2
1.3 Current study	3
2. Data	4
2.1 Data source: The 2016 Ethiopian Demographic and Health Survey.....	4
2.2 Response variable	5
2.3 Explanatory variables	5
3. Method.....	8
3.1 The survival and intensity functions	8
3.2 Cox proportional hazard model and testing the assumption.....	9
3.3 Accounting for clustering of women in households	11
3.3.1 The Frailty model	11
3.3.2 Frailty model and the different distributions	12
4. Results.....	13
4.1 Results of testing the proportional hazards assumption	13
4.2 Results from standard Cox model for the entire country	13
4.3 Results from Cox model with frailty for the entire country	15
4.4 Results from standard Cox model for each region	15
4.4.1 Tigray Region.....	17
4.4.2 Afar Region	17
4.4.3 Amhara Region	17
4.4.4 Oromia Region	18
4.4.5 Somali Region.....	18
4.4.6 Benishangul Region	19
4.4.7 Southern Nations & Nationalities People's Region (SNNPR)	19
4.4.8 Gambela Region	20
4.4.9 Harari Region.....	20
4.4.10 Addis Ababa City Administration	21
4.4.11 Dire Dawa City Administration.....	21
4.5 Results from Cox model with frailty for each region and city administration	21
4.5.1 Results from Cox model with gamma distributed frailty	22
4.5.2 Results from Cox model with log-normal distributed frailty	22
5. Detailed discussion of results	23

5.1 Validity of the proportional hazards assumption.....	23
5.2 Regional similarities and differences in intensity of transition to parenthood	24
5.3 Effect of clustering of women within households.....	25
5.4 Choice of distribution for the frailty term.....	26
6. Summary and conclusions	27
References	29
Appendix A: Tables.....	32
Appendix B: R-codes	35

1. Introduction

1.1 Background

The birth of the first child is a special event for a mother whose life can change dramatically. Priorities of career or education may have to be dropped or put aside since the transition into motherhood comes with responsibilities. The norm and expectations of women throughout history have been - giving birth and raising children. However, the norm around family building has changed. With increasing gender equality, women taking up more space in the work environment and the stigma around women prioritising careers and education slowly dissolves, family building has evolved with it. The mean age of first-time mothers has increased in the last 40 years in developed countries in the Organisation for Economic Co-operation Development, OECD countries (OECD, 2019).

The World Health Organisation (WHO) reports that with the increased age of first-time mothers, the estimated adolescent birth rate is decreasing. However, the actual number of childbirths has not decreased due to the large and in some parts, growing population of young women in the 15-19 age group (WHO, 2020). The United Nations Children's Fund (UNICEF) provides statistics that show how East and Southern Africa follows the same tendency of decreasing birth rates but has one of the highest adolescent birth rates in the world at 92.1 out of 1000 girls aged 15 to 19 as of 2015-2020 (UNICEF, 2019). In lower developed areas, at least 39% marry before the age of 18 and young women choose to become pregnant due to limited education and employment prospects. In such areas, motherhood is valued and marriage and childbearing are the best alternatives (WHO, 2020).

Adolescent motherhood has negative health aspects as adolescent mothers face a higher risk of eclampsia, puerperal endometritis, and systemic infections. Their infants face a higher risk of premature birth, low birth weight, and severe neonatal conditions. Teenage pregnancies and childbirth can also have a negative socio-economic impact. As teen mothers are more likely to drop out of school (Central Statistical Agency (CSA) [Ethiopia] and ICF, 2016), the consequences of the unfinished education may be low or zero qualifications which may lead to low wages or unemployment.

Ethiopia is located in East Africa and is the second-largest country in Africa. Its bordering countries are South Sudan and Sudan in the west, Djibouti in the east, Eritrea in the north, Kenya and Somalia in the south. The country consists of 10 regions and two administrative cities. The northern regions are Tigray, Afar and Amhara, the western regions are Benishangul-Gumuz and Gambela, to the south are Southern Nations Nationalities and Peoples (SNNPR) and formed June 2020 the Sidama region, to the east the Somali and Harari regions are located. Oromia is a large region stretching from the south to the central and the western part of the country, with the administrative city of Addis Ababa in the centre of it. The last administrative city, Dire Dawa is located to the central east. The major religions practised within the country

are Christian orthodox, Islam, and Christian protestant. The percentage of women practising each respective religion is approximately 43%, 31% and 23%, while the remaining 3% practices other religions (CSA [Ethiopia] and ICF, 2016). Ethiopia is considered the fastest growing economy in the region (Worldbank, 2021). At the same time, Ethiopia is considered a poor country. Approximately 78% of the population are located in rural areas (CSA [Ethiopia] and ICF, 2016). Urbanization is essential for economic growth, improving quality of life, and reducing poverty. Different projects are in motion to make this possible and creating job opportunities in urban areas is fundamental (Alemayehu, 2019).

Another aspect of developing the country is education. From the year 2000 to 2016 the enrollment to primary school tripled. On the contrary, several children do not complete their education. 15% will not make it past 5th grade and 46% will not make it past 8th grade (UNICEF, 2018). Education is a way out of poverty and the importance of it cannot be emphasized enough. The increase in primary school indicates that the country is moving in the right direction.

The population in Ethiopia is considered young with 41% of the population under the age of 15. More than 28% are between the ages of 15-29 (USAID, 2017). The Ethiopian Demographic and Health Survey (EDHS) from 2016 reported that one in ten teenage girls has already started childbearing in the country, indicating that young pregnancies and adolescent birth are a predominant issue. The median age for women entering parenthood in Ethiopia for women aged 25-49 is 19.2, i.e., 50% of the women in the sample gave birth before turning 20 years of age (CSA [Ethiopia] and ICF, 2016). The percentage of young women aged 15-19 who have begun childbearing vary from 3% in the capital region to 23% in the Afar region located to the northeast (WHO, 2020). The overall fertility rate varies between different regions of the country as well, the total fertility rate in the capital city Addis Ababa is 1.8 compared to 7.2 in the Somali region in the eastern part of the country.

It is evident that the intensity of entering motherhood differs between regions in Ethiopia. It would therefore be interesting to analyse how different factors affect the intensity in the country and its regions.

1.2 Previous studies

Starting from birth, humans are naturally affected by their surroundings. Factors such as family, residence area, education and economic conditions etc., shapes a human from an early age, and the impacts will last throughout a person's life.

The circumstances around the event of becoming a mother changed considerably in the 1960s when contraceptive methods were introduced and became available in the Western world. As the choice of delaying parenthood became an option, women chose to undertake education, take on employment, and pursue careers before entering parenthood. Studies show that contraceptives and abortion being legal in the 1970s are associated with postponing parenthood (Mills et al., 2011). The age of women postponing parenthood has continued to increase

throughout the years. Currently, countries such as South Korea, Japan, Switzerland, Greece, Italy and Spain are amongst the countries with the oldest women entering parenthood in the world with an average age of 31 years (Bui and Miller, 2018; Eurostat, 2020). In a journal article entitled “The trend towards delayed parenthood” (Wilkie, 1981) the author reports upon the disparity between women who chose to postpone parenthood and those who did not. The main differences were that women who chose to postpone parenthood had both higher education and higher occupational status, which led them to have more savings and higher income when they eventually decided to have a child compared to the women who did not choose to postpone parenthood. What Wilkie (1981) once considered a “trend” has now become a norm. Higher education and pursuing a career for a stable income is till this day prioritized by women and considered as the main reasons women delay parenthood around the developed world (Forbes, 2020).

Ethiopia is one of the fastest growing economies in the region which will have a positive effect on the whole country (Worldbank, 2021). That being said, Ethiopia is not considered a developed country as it is considered one of the poorest nations in Africa. An article (Chernet et al., 2019) states that women in Ethiopia living in urban areas have an extended time to first birth compared to women living in rural areas. The same applies to employed compared to unemployed women and educated versus uneducated women. Urban areas are usually more developed than rural areas as the population in urban areas have more access to information, and opportunities such as education and/or job opportunities are broad.

As stated previously, young pregnancies and the adolescent birth rate are a predominant issue in Ethiopia. Contrary to the western world, the differences in Ethiopia entering parenthood are greater depending on the socioeconomic status the woman belongs to. A previous report from Indonesia on time to the first child (Hidayat et al., 2014) analysed several variables like age, residence and education. Another article from Nigeria on young females' time to the first child (Kunnuji et al., 2018) used variables such as residence and education amongst other variables.

The variables, education and occupational status, mentioned in the article by Wilkie (1981) are still accurate today which is shown in the articles by Hidayat et al., (2014) and by Kunnuji et al., (2018) where age, education and residence area analysed as reasons for postponing parenthood. For this reason, they are included in this paper. However, none of the mentioned articles takes clustering of the women and the potential correlation within households that the women are clustered into when modelling the women's intensity of entering parenthood.

1.3 Current study

It was previously mentioned in section 1.1 that different statistics like overall fertility rate and percentage of teenage pregnancies differ depending on the region in Ethiopia. It would be interesting to study how different covariates affect women's timing of motherhood in different regions, how they may differ from one another and how accounting for clustering may change the estimated intensity to motherhood. All women are clustered into households and it might affect the timing as Ethiopia is a culturally diverse country with multiple ethnicities and

religions. It would therefore also be interesting to study if taking the household effects into consideration, change the estimated effects of the covariates on the timing of motherhood for women in Ethiopia and the different regions separately. Many studies have been made about how different covariates affect women's timing of giving birth to their first child, however, the household effect is often overlooked. The previous studies mentioned have used the Cox proportional hazard model (Cox PH model) when modelling the relative intensities to entering motherhood in the different countries. Neither of them did a multilevel analysis and took correlation within clusters into consideration. Another possible approach when modelling different covariates' effect of timing of motherhood is using the logistic regression as it will model the probability of parenthood depending on the covariates. However, the normal logistic regression does not take clustering into consideration either, therefore, we found it interesting and important to consider the cluster effect on the relative intensities.

The overall aim of this paper is therefore to analyse the effect of residence area, birth cohort and education on the intensity of entering parenthood among Ethiopian women and how those effects change across the regions whilst taking into account clustering. By performing a multilevel analysis, the study aims for a more accurate interpretation of the effects of the four covariates.

The issues of the study will therefore be:

- How does residence area, birth cohort and education level affect the intensity of entering parenthood for Ethiopian women?
- Does the effect of the three covariates change depend on the region?
- Does accounting for correlation within households change the estimated relative intensities of entering parenthood for women in Ethiopia?

The next section presents the data used in this study. Section 3 presents the statistical methods used to analyse our data. In section 4 we present our results and discuss them in detail in section 5. We present a summary and the conclusions in section 6.

2. Data

2.1 Data source: The 2016 Ethiopian Demographic and Health Survey

The data analysed in this paper is extracted from the final report of the 2016 Ethiopia Demographic and Health Survey (EDHS). The data was collected between January and June 2016 and contains 15,683 women respondents aged 15 to 49 and 487 households. It is based on a nationally represented sample and provides estimates at national and regional levels, as well as for urban and rural areas.

The sample frame used in the collection of the data was from the Ethiopian Population and Housing Census (PHC) and it is a complete list of 84,915 enumeration areas (EA) created in 2007. An enumeration area is a geographic area containing 181 households on average. The

sample frame contains information on the type of residence area, location and the estimated number of residential households. The 2016 EDHS was stratified and then selected in two stages. Each region was stratified into urban and rural areas resulting in 21 strata samples. Later the samples of EAs were selected independently in each stratum in two stages (CSA [Ethiopia] and ICF, 2016). As the data was collected in 2016, the Sidama region was not formed. Therefore, no data was collected specifically from this region, instead, its data is included in the SNNPR region. Although the 2016 EDHS contains 15,683 women, usable records for the purpose of this paper are 15,019 women

2.2 Response variable

The response variable in this study is ‘the intensity of transition to parenthood’ since the aim of this study is to analyse how education, age and residence area affect the transition to motherhood in Ethiopia. The survival time in our dataset is ‘exposure to first birth’ and the event is an indicator named ‘first birth’ that is 0 when censored, otherwise 1. A woman is exposed to childbirth at age 15 and the time variable is the number of months from age 15 and birth of the first child or the survey time - whichever came first. The censored observations in our sample are the women who have not yet given birth to their first child by the time of their interview.

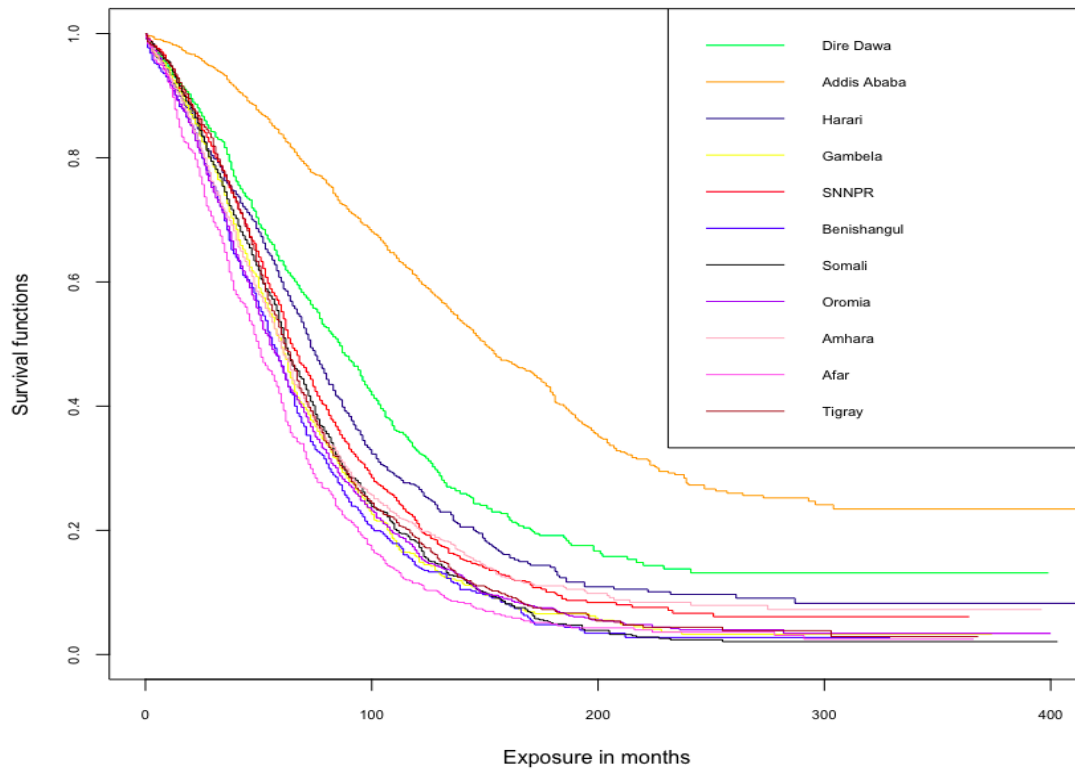
2.3 Explanatory variables

The 2016 EDHS is a report providing data and statistics in numerous areas, many that will be disregarded in this paper. The variables chosen are limited so the variables used are the ones that are known to have a significant impact on women's time to parenthood. Other covariates are of interest when modelling the intensity of transitioning to parenthood, however, interaction terms with specifically age made them inappropriate. The variables chosen are:

- Residence area: This covariate has been transformed into a factor variable that can take on value 1, signifying urban residence, or 2, signifying rural residence.
- Birth cohort: Cohort is an age covariate that specifies during what time period a woman was born. They are in four-year intervals with the first factor, cohort 1 being the youngest indicating that the woman was born between 1997 and 2001. There are seven cohorts in the dataset so the factor can take a value between 1 to 7.
- Education: This covariate signifies the education level of the women. It is transformed into a factor variable with four possible values. Value 1 signifies that the woman has primary education and has completed 8th grade in primary. Value 2 signifies that the woman has secondary education and completed 4th grade in secondary. If a woman has a higher education than that, the education factor takes value 3. Women who haven't completed 8th grade in primary are considered to have no education and the education factor takes value 0.
- Region: Region is a covariate that signifies where the women are residing. It can take 11 possible values where 9 are regions and two are city administrations. The region covariate is used in the Ethiopia models which contain women from all regions and

cities. Figure 1 shows the Kaplan-Meier (1958) estimated functions (proportion of women who are not yet mothers) by age and region.

Figure 1: Survival functions by region: KM estimation



It is evident that the timing of entering parenthood differs across regions. Therefore, a closer look at the effect of other covariates on the intensity of entering parenthood will be made for each region.

We want to consider women with the above explanatory variables and a frailty term to account for the fact that women are clustered within households. Such models are expected to adjust for correlations between women in the same cluster and yield unbiased estimates. In the 2016 EDHS, a household is defined as “A person or group of related or unrelated persons who live together in the same dwelling unit(s), who acknowledge one adult male or female as the head of the household, who share the same housekeeping arrangement, and who are considered a single unit” (CSA [Ethiopia] and ICF, 2016, p.12).

Table 2.1 contains summary statistics of the dataset across covariates of regions.

Table 2.1: Frequency table and the percentage of women who has entered motherhood for the entire country, regions and city administrations

Covariates	Ethiopia	Tigray	Amhar	Amhara	Oromia	Somali	Benishangul	SNNPR	Gambela	Harari	Addis Ababa	Dire Dawa
Residence 1: Urban	5166 (49%)	399 (56%)	186 (60%)	235 (49%)	243 (51%)	323 (63%)	135 (42%)	219 (53%)	322 (58%)	501 (57%)	1793 (41%)	810 (49%)
Residence 2: Rural	9853 (72%)	1213 (67%)	898 (76%)	1399 (69%)	1550 (74%)	1024 (74%)	911 (73%)	1569 (67%)	636 (77%)	369 (77%)	-	284 (78%)
Cohort 1: 1987-2001	3474 (10%)	423 (8%)	262 (19%)	354 (7%)	411 (6%)	316 (15%)	236 (10%)	391 (7%)	205 (13%)	181 (13%)	431 (2%)	264 (7%)
Cohort 2: 1992-1996	2824 (53%)	311 (59%)	227 (77%)	272 (50%)	313 (43%)	250 (69%)	197 (60%)	315 (48%)	186 (64%)	194 (57%)	359 (15%)	200 (37%)
Cohort 3: 1987-1991	2714 (80%)	237 (81%)	222 (92%)	298 (81%)	316 (76%)	227 (89%)	185 (88%)	361 (81%)	179 (88%)	148 (79%)	319 (49%)	222 (72%)
Cohort 4: 1982-1986	2099 (91%)	180 (97%)	137 (97%)	240 (92%)	283 (78%)	179 (96%)	136 (96%)	256 (94%)	150 (98%)	119 (92%)	263 (69%)	156 (85%)
Cohort 5: 1977-1981	1785 (94%)	201 (97%)	97 (96%)	205 (97%)	222 (89%)	163 (88%)	149 (99%)	213 (98%)	111 (98%)	111 (91%)	203 (75%)	110 (92%)
Cohort 6: 1972-1976	1201 (95%)	131 (98%)	68 (96%)	143 (97%)	134 (99%)	129 (97%)	90 (99%)	152 (97%)	70 (94%)	76 (93%)	131 (81%)	77 (90%)
Cohort 7: 1967-1971	922 (96%)	129 (99%)	71 (99%)	122 (97%)	114 (97%)	83 (98%)	53 (100%)	100 (98%)	57 (96%)	41 (88%)	87 (85%)	65 (95%)
Education 0: No Education	6598 (87%)	669 (89%)	790 (83%)	882 (90%)	892 (89%)	1000 (81%)	517 (91%)	770 (90%)	296 (95%)	299 (84%)	144 (67%)	339 (85%)
Education 1: Primary	5040 (51%)	524 (59%)	235 (45%)	483 (44%)	692 (30%)	240 (42%)	374 (56%)	765 (50%)	369 (60%)	301 (64%)	649 (40%)	408 (50%)
Education 2: Secondary	2197 (38%)	317 (97%)	37 (41%)	189 (25%)	151 (32%)	74 (47%)	98 (34%)	184 (29%)	204 (59%)	156 (49%)	556 (38%)	231 (36%)
Education 3: Higher	1184 (39%)	102 (35%)	22 (55%)	80 (29%)	58 (40%)	33 (39%)	57 (21%)	69 (46%)	89 (58%)	114 (46%)	444 (37%)	116 (38%)
Total number of women	15019	1612	1084	1634	1793	1347	1046	1788	958	870	1793	1094
Total number of households	487	452	410	452	466	451	415	459	390	383	442	402

Below the statistical methods used to analyse the dataset are presented.

3. Method

Since our response variable is the intensity of transition to parenthood, the appropriate analysis method is the hazard rate framework of which the Cox proportional hazards model (Cox, 1972) is the most common.

3.1 The survival and intensity functions

One of the main properties of survival data is that the response variable is a non-negative discrete or continuous variable and represents the time from a defined origin until a defined event or end point. Another property is that censoring arises when the starting or ending events are not observed. Right censoring is when the final endpoint is known to surpass a certain value (Moore, 2016), i.e., the start point is precisely observed but the event and end point are not for an individual.

The survival analysis method depends on the survival distribution, which is mainly specified by the survival function and intensity function. *The survival function* is the probability of surviving up to a specific time-point, t . The survival-time T , can take any positive value and is expected to have an underlying density distribution $f(t)$. The survival function $S(t)$ will thereby formally be expressed as the following:

$$S(t) = P(T \geq t) = 1 - P(T < t) = 1 - F(t), \quad 0 < t < \infty \quad (3.1)$$

where $F(t)$ is the cumulative density distribution and the integral of $f(t)$ (Collett, 2015). The probability of surviving up until time-point 0 is 1, therefore the survival function takes value 1 at time 0. The function decreases or remains constant over time and does not go below the value 0 (Moore, 2016).

The intensity function and survival function are closely related to one another and the latter is frequently defined in terms of the former. *The intensity function*, also known as *the hazard function*, is the instantaneous rate at which an event occurs. It is the probability of the event occurring, given that the subject has survived up to a specific time-point t , divided by the length of the time interval. Formally expressed as:

$$h(t) = \lim_{\delta \rightarrow 0} \frac{P(t < T < t + \delta | T > t)}{\delta} \quad (3.2)$$

where δ is the time interval (Moore, 2016). The intensity can casually be interpreted as the probability of the event occurring, conditional on the subject surviving up to that point (Collett, 2015).

The relationship between the intensity function and the survival function can be derived from expression 3.1 and 3.2.

$$h(t) = \frac{f(t)}{S(t)} \quad (3.3)$$

where $f(t)$ is the density function of T (Collett, 2015).

3.2 Cox proportional hazard model and testing the assumption

The Cox proportional hazard model (Cox, 1972), is a popular model fitting right censored survival data similar to the linear and logistic regressions. The model estimates the effect of covariates on the intensity function and assumes that they act multiplicatively on the intensity while making no assumption about any parametric underlying distributions. The Cox PH model is expressed:

$$h_i(t) = h_0(t)\exp(\beta_1 x_{i1} + \dots + \beta_k x_{ik}) \quad (3.4)$$

where $h_0(t)$ is the baseline. Further, x is the k number of fixed covariates of the i :th subject and β their respective coefficient. The interpreted intensity rate will therefore be the exponential of the estimated coefficient and their respective value of the covariate. Taking the logarithm of both sides the model can be rewritten to:

$$\log [h_i(t)] = \alpha(t) + \beta_1 x_{i1} + \dots + \beta_k x_{ik} \quad (3.5)$$

where $\alpha(t) = \log[h_0(t)]$.

The baseline intensity in model 3.5 is an unspecified function except for it being non-negative. Therefore, is the $\alpha(t)$ also unspecified which makes the Cox PH model popular (Allison, 2010). Using parametric models to estimate the intensity rate can be difficult because they require strong assumptions about the underlying survival distribution. The maximum partial likelihood, which is used to estimate the coefficients for covariates, allows for an unspecified baseline survival function for each observed subject (Moore, 2016).

The maximum partial likelihood differs from a regular likelihood because it is the product solely of the failure times while the censored times do not contribute to the factors. Another difference is that the former's factors are conditional probabilities, resulting in the maximum partial likelihood not being a probability (Moore, 2016). The maximum partial likelihood is modelled:

$$L(\beta) = \prod_{j=1}^D \frac{\exp(\beta \times x_i')}{\sum_{j \in R(t_i)} (\beta \times x_j')} \quad (3.6)$$

where $R(t_i)$ is the set of subjects at risk still at risk immediately before the uncensored data point t_i . Further, x_j is a vector of explanatory variables for the j :th individual and x_i is a vector of the same set of variables for the subjects experiencing the event and data point t_i . The vector of the coefficients to be estimated is denoted β and lastly, the D is representing the number of deaths.

Gebrenegus Ghilagaber mentions in lecture five in the course Survival Analysis (2021) that a positive characteristic of the Cox proportional hazards model is that if the baseline intensity follows some parametric distribution, the Cox proportional hazard model's estimations are close to the correct ones, i.e., it is robust to distributional assumptions.

When taking the ratio of two individuals' Cox proportional hazards, i.e., dividing their intensity functions, the baseline intensity will be cancelled out. The model must therefore be assumed to be proportional and constant over time (Allison, 2010). However, some common covariates included in the model might be time-dependent (e.g., income and education) which violates the proportionality assumption. The covariates will change at different rates for different individuals and result in inconsistent intensity ratios. The maximum partial likelihood estimation is not much affected by this problem (Allison, 2010). Allison argues that concerns about the proportional assumption are excessive. By estimating a proportional model when the proportionality assumption is violated for some time-dependent variable, the estimated coefficients for those specific variables are an approximate average effect over the range of times observed in the data. Sometimes the interaction of these variables with time is forceful that ignoring them would be misleading. An extension of the proportional model is then made (Allison, 2010).

To test the proportional hazards assumption, the Schoenfeld residuals method is used. Schoenfeld residuals have separate residuals for each covariate of each individual. It tests if the Schoenfeld residuals are correlated with time or some function of time, significant results signify correlation (Allison, 2010). Education is a time-dependent variable used in this paper. We argue, however, that since our purpose of the paper is not to get a profound analysis of the education level's effect on women's age when entering parenthood, an approximate average of education's effect is satisfactory. Therefore, a Cox proportional model will be used regardless of the results of the Schoenfeld residuals test. Additionally, as Allison implies (2010), the Cox PH model is robust even if the proportional assumption may be violated in some covariates.

The Cox proportional hazard model assumes that the intensity is continuous since time is a continuous variable and under this assumption, subjects cannot have identical survival times,

called tied survival times. Time is always rounded off to the nearest day, month or year which may yield to subjects having tied survival times (Collett, 2015). The partial likelihood formula expressed above must be modified when the dataset contains tied survival times. One solution for the issue with tied continuous survival times is the marginal *method for ties* which averages the numeration of the possible ordering of the failure times. However, large datasets normally contain numerous tied observations and the marginal method becomes burdensome. Moore writes that the Breslow approximation adjusts the terms of the marginal method so they all have the same denominator, corresponding to all subjects at risk. The Breslow approximation, therefore, is a more suitable approach to solve the issue with tied survival times when the datasets are larger (Moore, 2016).

3.3 Accounting for clustering of women in households

3.3.1 The Frailty model

Data with multilevel structure are commonly encountered in many areas of research. A study of mortality of patients is clustered within hospitals, hospitals are clustered within regions etc. (Austin, 2017). Therefore, it is natural to assume that not all covariates that could influence the distribution are included in a model. Individuals with the same value of the covariates may have different distributions, called unobserved heterogeneity since they are nested in different clusters (Balan and Potter, 2020). When individuals are nested in the same cluster, the assumption of independent observations of regular regression models is violated. Outcomes of observations of the same higher-level unit are more likely to be correlated (Austin, 2017).

The Frailty model is similar to the Cox proportional hazard model but with the inclusion of random effects. The random effects account for the unobserved heterogeneity or within-cluster homogeneity. The term frailty model is used when random effects are incorporated into a survival regression model, most commonly a Cox proportional hazards regression model or a parametric survival model. Shared frailty model is a description for these types of models since all subjects in the same cluster share the same random effect. These random effects indicate an increase or decrease in the intensity for distinct classes in the Cox model.

Assuming M clusters, the Cox model is formulated as:

$$h_{ij}(t) = h_0(t) \exp (X_i \beta_i + \alpha_j) \quad (3.7)$$

where similar to the Cox PH model, h_0 indicates the baseline intensity, X_i the vector of explanatory variables of each i :th subject/observation and the β their respective coefficients. Additional to the Cox PH model is the α_j which indicates the random effects associated with the j :th cluster. The shared frailty is indicated by the exponential of the random effect, $\exp(\alpha_j)$. The random effect can therefore be considered as a random intercept that affects the

linear predictor, while the shared frailty term has a multiplicative effect on the reference baseline intensity function.

An assumption needed to perform a Cox shared frailty model is that each observation is only a member of only one cluster (Austin, 2017). Other assumptions are that failure times in a specific cluster are conditionally independent given the frailties. The value of the frailty is constant over time and is common to all individuals in a cluster. The dependence created between event times in a cluster must be positive in the frailty model (Wienke, 2011).

Interpreting the frailty model is different in comparison to the Cox proportional hazard model. The previous models the relative risk at the population level. The frailty model, models the cluster-level relative risk and refers to comparison within the cluster where the individuals share the same frailty or random effect. The parameters are adjusted for the correlation within the cluster and the variance is a measure of the correlation between the survival times in the same group. If there is a within-cluster correlation, the estimations of the parameters from the two models should differ from one another (Wienke, 2011).

Accounting for clustering and using frailty models is essential since it takes the correlation within the clusters into consideration. By disregarding the correlation within observations within the same cluster, the results of the models may lead to less accuracy. Therefore, the frailty model will be used in this paper to account for the correlation between households.

3.3.2 Frailty model and the different distributions

Different distributions have been proposed for the shared frailty term but the most commonly used are the gamma and lognormal distributions (Austin, 2017). In the log-normal frailty model, the frailty terms will have log-normal distribution and the random effects will have a normal distribution (Austin, 2017). In the gamma frailty model, the random effects of the clusters are as the logarithms of independent, identically distributed gamma random variables.

The gamma distribution is asymmetrical on the log-intensity scale while the lognormal is symmetric. In a particular application made by Therneau and Grambsch the gamma distribution had a heavy lower tail implying that there was a portion of clusters with abnormally low risk. Depending on the data, the assumption that a proportion of the clusters having low risk may be inappropriate. If inappropriate, the lognormal distribution for the shared frailty model may be more appropriate. Therneau prefers the lognormal distribution for the random effects since it simplifies the framework for multiple and correlated effects. Wienke writes that the log-normal distribution assumption allows much more flexibility, especially when modelling multivariate correlation structures (Wienke, 2011). There is research on how well the gamma distribution fits the shared frailty terms and how well it fits applicable data, however not how good the distribution is compared to others (Austin, 2017). In this study, both the gamma and lognormal distribution will be used.

All the models and the Schoenfeld residuals test are performed by the statistical software R. and the specific package used to apply the statistical models in this paper is ‘survival’ and is found in the Comprehensive R Archive Network (CRAN) (Therneau, 2021).

In the following section, the results when using the mentioned methods will be presented.

4. Results

4.1 Results of testing the proportional hazards assumption

R has a procedure that can test the proportional hazards assumption of Cox regression models and prints out the p-values of the correlation between the covariate and time. Non-significant correlations are what is desirable. Education is a covariate where at least one level correlates with time in every model. In the Cox PH model containing all the women in all regions all education levels correlate with time, similar results are found in the Tigray Cox PH model. In the Amhara, Oromia, Benishangul, Addis Ababa and Dire Dawa regions, the higher education levels, secondary and higher correlates with time. In the remaining regions, it differs between secondary and higher education correlating with time. Rural has a significant correlation with time in overall Ethiopia- and the Amhara-, Oromia- and Benishangul Cox PH models. Cohort 6 and 7 correlates with time in the Afar Cox PH model and cohort 3 in the Somali Cox PH model. Cohort 4 and 7 correlates with time in the Ethiopia Cox PH model containing the women from all regions. In that same model, the Afar, Amhara, Oromia, Benishangul and Addis Ababa regions correlate with time and violate the proportional hazards assumption.

4.2 Results from standard Cox model for the entire country

When modelling the Cox proportional hazard model in R, the program uses the lowest level of our variables as the baseline. In this analysis, the baseline includes region level 1 (Tigray), residence level 1 (urban), cohort level 1 (born 1997-2001) and education level 0 (no education). In the tables are the intensity ratios of each covariate presented, the estimated parameters - the log-intensity is the logarithm of the numbers presented in the tables. Table 4.1 presents R’s estimated intensity ratios in the standard Cox PH model and the two frailty models.

Table 4.1: The estimated intensity ratios (relative risks) of entry into parenthood from three models

Covariates	Standard Cox	Gamma frailty	Log-normal frailty
Region 2: Afar	1,11893 **	1,1194 **	1,1189 **
Region 3: Amhara	0,89776 **	0,897 **	0,8973 **
Region 4: Oromia	0,97094	0,9699	0,9691
Region 5: Somali	0,84033 ***	0,8377 ***	0,8357 ***
Region 6: Benishangul	1,02999	1,0305	1,0311
Region 7: SNNPR	0,81221 ***	0,8097 ***	0,8079 ***
Region 8: Gambela	1,11724 **	1,1187 **	1,1204 **
Region 9: Harari	0,95922	0,9594	0,9606
Region 10: Addis Ababa	0,57255 ***	0,5707 ***	0,5699 ***
Region 11: Dire Dawa	0,86081 ***	0,8593 ***	0,8582 ***
Residence 2: Rural	1,36534 ***	1,3691 ***	1,3699 ***
Cohort 2: 1992-1996	1,87702 ***	1,8794 ***	1,8788 ***
Cohort 3: 1987-1991	2,17016 ***	2,1742 ***	2,1740 ***
Cohort 4: 1982-1986	2,44622 ***	2,455 ***	2,4582 ***
Cohort 5: 1977-1981	2,24966 ***	2,2537 ***	2,2528 ***
Cohort 6: 1972-1976	2,34318 ***	2,35 ***	2,3526 ***
Cohort 7: 1967-1971	2,12693 ***	2,1301 ***	2,1304 ***
Education 1: Primary	0,84933 ***	0,8478 ***	0,8472 ***
Education 2: Secondary	0,46994 ***	0,4671 ***	0,4656 ***
Education 3: Higher	0,3324 ***	0,3299 ***	0,3286 ***
Frailty p-value	-	0,29	0,2
Variance of random effects	-	0,00266975	0,007359318

***: Estimates significant at the 1% level. **: Estimates significant at the 5% level.

*: Estimates significant at the 10% level

Note that ‘,’ is the decimal separator in the table and ‘.’ elsewhere.

Some of the parameters in table 4.1 are not significant on any level in any of the models. The Oromia, Benishangul and Harari regions are not statistically significantly different from zero at any level and therefore not statistically significant from the Tigray region.

Women from Gambela and Afar have a 12% higher intensity of entering parenthood compared to the women in Tigray. Women from Amhara, Somali, SNNPR regions and the city administration Dire Dawa have a lower intensity compared to the Tigray women. Their intensity is respectively around 90%, 84%, 81% and 86% that of Tigrays. Women from Addis Ababa have the lowest intensity ratio out of all the regions and it is 57% that of Tigray.

Ethiopian women in rural areas enter parenthood at a higher rate than women from urban areas. The intensity ratio is 1.36, indicating that living in rural areas increases the intensity by 36% in Ethiopia.

Women from older cohorts have a higher intensity of entering parenthood compared to the youngest women in Ethiopia. Cohort 4 has the largest intensity ratio at 2.4462, i.e., the intensity of entering parenthood and giving birth to their first child for women born between 1992 to 1986, is more than twice the intensity of the women from the youngest cohort at each time point. More specifically, their intensity is 2.4 times that of the women from the youngest cohort. The smallest intensity ratio is that of cohort 2 at 1.87702. Overall, the rest of the cohort intensity ratios are between 2.1 and 2.3.

Women with education have a lower intensity of entering parenthood compared to women with no education. The intensity ratios are 85%, 47% and 33% respectively, i.e., the intensity rate of women with primary education is 15% lower, the intensity of women with secondary education is 53% lower and the intensity of women with higher education is 67% lower than the intensity of women with no education at each time point in Ethiopia.

4.3 Results from Cox model with frailty for the entire country

Fitting the multilevel frailty models to the data did not change the overall significance of any parameter. Nor does the estimated parameters change notably compared to the Cox PH model. Comparing the estimations of the intensity ratios of the two frailty models, the difference is barely existent as none of the absolute differences exceeds 0.01.

When interpreting the intensity ratios in the frailty models they do not vary from the Cox PH model. The only intensity ratio that differs is that of cohort 4 and cohort 6. The Cox PH model estimates that the rate of entering parenthood of women from the fourth cohort is 2.45 times higher than that of the first and that the sixth cohort's rate is 2.34 higher. The frailty models estimated that the same intensity ratios are 2.46 times higher than the first cohort's intensity and 2.35 times higher respectively. However, as mentioned, those are the only intensity ratios that have slightly different estimated intensity ratios.

The frailties, or random effects, have a low variance in both models and a high p-value indicating that they are not significant.

4.4 Results from standard Cox model for each region

All three tables presented below are also found in appendix A for closer observations.

Table 4.2: The estimated intensity ratios (relative risks) for each region of the Cox proportional hazard model

Covariates	Tigray	Afar	Anhara	Oromia	Somali	Benishangul	SNPR	Gambela	Harari	Addis Ababa	Dire Dawa
Residence 2. Rural	1.0925	1.9131 ***	1.09428	1.28516 **	1.29125 ***	1.11915	0.89073	1.49545 ***	1.73434 ***	-	2.09797 ***
Cohort 2. 1992-1996	2.22057 ***	1.4782 **	2.41886 ***	1.88886 ***	1.49050 **	2.54257 ***	1.97729 ***	1.96573 ***	1.34856	2.5858 **	2.01817 ***
Cohort 3. 1987-1991	2.11281 ***	1.3058	3.18461 ***	1.94759 ***	1.49036 **	3.10862 ***	2.68693 ***	2.21596 ***	1.45865	5.8580 ***	2.53526 ***
Cohort 4. 1982-1986	2.9374 ***	1.6134 ***	3.59185 ***	2.1167 ***	1.43625 **	2.9243 ***	3.1471 ***	2.32428 ***	1.71264 **	6.7866 ***	3.00738 ***
Cohort 5. 1977-1981	2.62651 ***	1.5937 **	3.8326 ***	1.63368 ***	1.06406	2.95042 ***	3.16016 ***	1.99955 ***	1.25142	6.8228 ***	3.926 ***
Cohort 6. 1972-1976	2.97693 ***	1.2597	4.39751 ***	1.82366 ***	0.97952	2.91199 ***	3.69292 ***	1.97458 ***	1.40869	8.7573 ***	2.61256 ***
Cohort 7. 1967-1971	3.1516 ***	1.1398	3.3014 ***	1.52628 **	0.87756	3.22324 ***	3.33903 ***	1.83072 **	1.01629	8.3227 ***	3.36648 ***
Education 1. Primary	0.90824	0.8904	0.74403 ***	0.66227 ***	0.98754	0.99095	0.85396 **	1.1149	1.09077	0.8404	0.88859 ***
Education 2. Secondary	0.43222 ***	0.4664 ***	0.32222 ***	0.34162 ***	0.7536	0.49803 ***	0.32965	0.91349	0.59185 ***	0.4763 ***	0.48614 ***
Education 3. Higher	0.26147 ***	0.4398 ***	0.20207 ***	0.31753 ***	0.3788 ***	0.13485 ***	0.26316 ***	0.54052 ***	0.41244 ***	0.39 ***	0.36003 ***

Table 4.3: The estimated intensity ratios (relative risks) for each region of the Cox model with gamma distributed frailty

Residence 2. Rural	1.0943	2.0045 ***	1.0895	1.2871 **	1.2948 ***	1.1212	0.8902	1.5001 ***	1.7432 ***	-	2.177 ***
Cohort 2. 1992-1996	2.2251 ***	1.5133 **	2.4171 ***	1.8934 ***	1.4926 **	2.5506 ***	1.9785 ***	1.9712 ***	1.3504	2.6233 **	2.0007 **
Cohort 3. 1987-1991	2.116 ***	1.3231 *	3.1894 ***	1.9517 ***	1.4822 **	3.1278 ***	2.6925 ***	2.2213 ***	1.4622	6.0416 ***	2.5791 ***
Cohort 4. 1982-1986	2.9478 ***	1.6541 ***	3.6073 ***	2.1172 ***	1.4378 **	2.9466 ***	3.1563 ***	2.3328 ***	1.7189 **	6.9734 ***	3.0458 ***
Cohort 5. 1977-1981	2.6346 ***	1.6426 ***	3.8398 ***	1.6327 ***	1.0629	2.9655 ***	3.1709 ***	2.0033 ***	1.2508	7.1682 ***	4.0776 ***
Cohort 6. 1972-1976	2.9908 ***	1.2361	4.4229 ***	1.8244 ***	0.9779	2.933 ***	3.7087 ***	1.9778 ***	1.4088	9.2409 ***	2.6756 ***
Cohort 7. 1967-1971	3.1689 ***	1.1491	3.2975 ***	1.5239 **	0.8748	3.2476 ***	3.3518 ***	1.8322 **	1.0155	8.898 ***	3.5422 ***
Education 1. Primary	0.9077	0.8947	0.7367 ***	0.6796 ***	0.9871	0.9929	0.8529 **	1.1151	1.0916	0.829	0.8882
Education 2. Secondary	0.4296 ***	0.4528 ***	0.3167 ***	0.3392 ***	0.7523	0.4963 ***	0.3279 ***	0.9116	0.5903 ***	0.4511 ***	0.4634 ***
Education 3. Higher	0.2595 ***	0.4113 ***	0.1968	0.3154 ***	0.3769 ***	0.1332 ***	0.2615 ***	0.538 ***	0.4103 ***	0.37 ***	0.3412 ***
Frailty p-value	0.34	0.18	0.37	0.54	0.32	0.33	0.86	0.6	0.31	0.1	0.13
Variance of random effects	0.002030249	0.08150213	0.01323474	0.00005	0.001291504	0.004449385	0.0000005	0.00005	0.002121251	0.09709179	0.08474129

Table 4.4: The estimates intensity ratios (relative risks) for each region of the Cox model with log-normal distributed frailty

Residence 2. Rural	1.101	1.9763 ***	1.085	1.2852 **	1.2971 ***	1.1286	0.8898	1.5003 ***	1.7752 ***	-	2.2019 ***
Cohort 2. 1992-1996	2.2272 ***	1.5011 **	2.4131 ***	1.8929 ***	1.4897 **	2.5828 ***	1.979 ***	1.9731 ***	1.3668	2.6256 **	1.9908 **
Cohort 3. 1987-1991	2.1236 ***	1.3164 *	3.1863 ***	1.9508 ***	1.4854 **	3.2408 ***	2.6932 ***	2.2255 ***	1.4877	6.0655 ***	2.5875 ***
Cohort 4. 1982-1986	2.9445 ***	1.6414 ***	3.6112 ***	2.118 ***	1.4409 **	3.0953 ***	3.1575 ***	2.3368 ***	1.7754 **	7.0021 ***	3.0643 ***
Cohort 5. 1977-1981	2.677 ***	1.6257 ***	3.8348 ***	1.6317 ***	1.0628	3.0336 ***	3.1717 ***	2.0035 ***	1.2608	7.2117 ***	4.1384 ***
Cohort 6. 1972-1976	3.027 ***	1.2438	4.428 ***	1.8243 ***	0.9753	3.0624 ***	3.7126 ***	1.9789 ***	1.4605	9.2889 ***	2.7057 ***
Cohort 7. 1967-1971	3.2106 ***	1.1454	3.2858 ***	1.5229 **	0.8728	3.3979 ***	3.3531 ***	1.8366 **	1.0161	8.9953 ***	3.6053 ***
Education 1. Primary	0.9013	0.8926	0.7323 ***	0.6789 ***	0.9885	1.024	0.8527 **	1.1155	1.0955	0.8247	0.8876
Education 2. Secondary	0.4198 ***	0.4569 ***	0.3136 ***	0.3382 ***	0.754	0.4978 ***	0.3278 ***	0.9117	0.5777 ***	0.4467 ***	0.4562 ***
Education 3. Higher	0.2542 ***	0.4204 ***	0.1935 ***	0.3145 ***	0.3769 ***	0.1249 ***	0.2613 ***	0.5382 ***	0.3869 ***	0.3667 ***	0.3355 ***
Frailty p-value	0.27	0.22	0.32	0.39	0.4	0.097	0.35	0.39	0.27	0.051	0.045
Variance of random effects	0.03746311	0.05380972	0.02643687	0.002604167	0.01156039	0.08672587	0.001302063	0.005208333	0.05884628	0.1258594	0.133317

***: Estimates significant at the 1% level. **: Estimates significant at the 5% level.

*: Estimates significant at the 10% level

Note that ‘.’ is the decimal separator in the table and ‘.’ elsewhere.

4.4.1 Tigray Region

The intensity of Tigray women residing in rural areas entering parenthood is not statistically different from the intensity of Tigray women living in urban areas since the intensity ratio of the previous is not significant at any significance level.

All the cohorts enter parenthood at a higher rate compared to the youngest women. The rate of the women from cohort 2 and 3 is slightly more than twice the rate of the youngest cohort. The intensity ratio of cohort 4, 6 and 7 is even higher, their ratio is around 3 times higher.

Tigray women with no education do not have a statistically significant intensity of entering parenthood compared to the Tigray women with no education, i.e., the hypothesis that the rate of entering motherhood is equal for both groups cannot be rejected at any level. The women with completed secondary education in the region enter parenthood at a slower rate than the women with no education in the same region as their estimated intensity ratio is 43%. The intensity ratio of Tigray women with higher education is even lower as it is 26% that of those with no education.

4.4.2 Afar Region

In the Afar region, women residing in rural areas enter parenthood at almost twice the rate compared to the women living in urban areas.

The intensity ratios by birth cohort, compared to the other regions, do not differ greatly from one another. Neither cohort 3, 6 or 7 are significant at any level, i.e., they do not enter parenthood at another rate than the youngest women in the region. The remaining intensity ratios are 1.48, 1.61 and 1.59 signifying that the rate of entering parenthood of the women belonging to the cohorts are 48%, 61% and 59% higher than the youngest Afar women.

The estimated intensity ratios of Afar women with completed secondary education and Afar women with higher education do not vary greatly as they are 47% and 0.44% respectively. More specifically, the women with completed secondary education enter parenthood at a 53% slower rate compared to the rate of the women with no education at each time point, while the rate of the women with higher education is 56% slower. The intensity ratio of Afar women with no education is estimated to be 89% but the parameter, the log-intensity, is not significant at any level. Therefore, it cannot be concluded that the Afar women with completed secondary education enter parenthood at a different rate compared to those with no education.

4.4.3 Amhara Region

Similarly, to the Tigray region, when fitting a Cox PH model to the dataset containing exclusively women staying in the Amhara region, the estimated intensity ratio of the women residing in rural areas is not statistically significant. The rate of entering parenthood of Amhara women living in rural areas cannot be proven in our dataset to be different from the rate of those living in urban areas.

The estimated intensity ratio increases with each birth cohort in the Amhara region, except for the last one where it decreases to be about the same ratio as the one of cohort 3. Their rates of entering parenthood, going from cohort 2 to 7, are 2.42, 3.18, 3.59, 3.16, 4.40 and 3.30 times higher than the rate of the youngest cohort in the Amhara region. All of the intensity ratios are statistically significant at the 1% significance level.

With every higher level of education, the estimated intensity ratio decreases. Women with higher education in the Amhara region enter motherhood at a lower rate than those with no education. The estimated intensity ratio of women with completed primary education is 74% that of the intensity of the women with no education at each time point. The intensity of women with completed secondary education is 32% and the intensity of those with higher education is 20% that of the women with no education.

4.4.4 Oromia Region

All the estimated intensity ratios are significant at the 5% and 1% level when the Cox PH model is fitted to exclusively the women staying in Oromia.

The estimated intensity ratio for the Oromia women residing in rural areas is 1.29, signifying that the intensity of mentioned women is 29% higher at each time point than the intensity of the Oromia women living in urban areas.

Among the cohorts, the estimated intensity ratios differ and do not follow a particular pattern as the Amhara Cox PH model does. Cohort 4 is the cohort with the highest estimated intensity ratio as they enter parenthood at twice the rate compared to the women from the youngest cohort in the same region. The cohort with the smallest intensity ratio is cohort 7 whose intensity of entering parenthood is 53% higher than the youngest cohort.

Similarly, to the Afar Cox PH model, the intensity ratios of the women with completed secondary education and the women with higher education do not differ much from one another. The intensity of entering parenthood for those with secondary education is 34% compared to those with no education and the intensity of the women with higher education is 32% lower. Women with completed primary education have a lower intensity of entering parenthood compared to the women with no education as it is 68% compared to those with no education. It is the lowest 'primary' intensity ratio out of all the regions.

4.4.5 Somali Region

In the Somali region, the women residing in rural areas enter parenthood at a slightly higher rate as their intensity is 29% higher than the intensity of the women residing in urban areas.

The estimated intensity ratio of the cohorts follows a different pattern compared to the intensity ratios of previous regions. Among the women in the Somali region, the younger cohorts give birth to their first child at a higher rate than the older cohorts as the estimated intensity ratio

decreases with each older cohort level. Neither cohort 5, 6 or 7 are statistically significant at any level, signifying that there is not enough support in our dataset that Somali women from the 5th, 6th and 7th cohort would have a different rate of entering motherhood compared to the youngest women in the same region. The women belonging to cohort 2 and 3 enter parenthood at an almost 50% higher rate than the youngest Somali cohort while the 4th cohort enters parenthood at a 44% higher rate.

The only education level in the Somali Cox PH model that is significant is the highest level of education 'higher'. The estimated intensity ratio is 0.38, i.e., Somali women with higher education enter parenthood at a 62% slower rate than the Somali women with no education. The other education levels are not significant at any level, so the hypothesis that Somali women with primary education enter parenthood at a different intensity than those with no education cannot be rejected at any significance level. The same goes for the Somali women with secondary education.

4.4.6 Benishangul Region

The rate of entering parenthood for women residing in rural areas of the region is not significantly different from the rate of the women residing in urban areas as the estimated intensity ratio in table 4.2 is not statistically significant at any level.

The intensity ratios of the cohorts do not follow a particular pattern. The women belonging to cohort 3 and cohort 7, enter parenthood at 3 times higher rate compared to the youngest women, cohort 1. Cohort 2 rate is 2.5 times higher compared to the youngest cohort.

Education on a primary level is not statistically significant. There is not enough evidence to state that the intensity of entering parenthood for women with primary education is different from the intensity of women with no education. Education on a secondary and a higher level are both statistically significant at all levels. The estimated intensity ratio is 50% for women with secondary education and 13% for women with higher education. More specifically, the groups enter parenthood at a 50% and 87% lower rate compared to the women with no education, respectively.

4.4.7 Southern Nations & Nationalities People's Region (SNNPR)

As the estimated intensity ratios for women residing in the rural areas is not statistically significant, there is not enough evidence to state that there is a difference in the rate for women in the SNNPR region entering parenthood based on if the woman resides in the urban or rural areas.

The estimated intensity ratios for the cohorts are all statistically significant at all significance levels. The rate increases with each cohort up until cohort 6. The 7th cohort has a lower ratio than cohort 6 and enters parenthood at 3,3 times higher rate compared to cohort 1. Whilst cohort 6 enters parenthood at 3.7 times higher rate than cohort 1.

All levels of education are statistically significant on at least significance level 5%. Women with primary education enter parenthood at a 15% lower rate, women with education on secondary level enter parenthood at a 67% lower rate, and women with higher education enter parenthood at a 74% lower rate compared to the not educated women.

4.4.8 Gambela Region

The estimated intensity ratio for women residing in the rural area are significant on all levels, and therefore there is a difference in the rate of entering parenthood if a woman lives in the rural or urban area of the region. The rate is 49% higher for women living in rural area to enter parenthood compared to urban.

Cohort 2 to 6 are statistically significant on all levels. Cohort 7 is significant at significance level 1% and 5%. Women belonging to cohort 2 and 6 enter parenthood at a 97% higher rate compared to cohort 1. While cohort 3, 4 and 5 enter parenthood at an approximately 2 times higher rate compared to women in cohort 1.

In the Gambela region, there is no proof that women with education on a primary and secondary level differ in the rate of entering parenthood from women with no education as the estimated intensity ratios for those levels are not statistically significant.

Women with education on a higher level enter parenthood at a 46% slower rate compared to women with no education at a significance level of 1%.

4.4.9 Harari Region

The estimated intensity ratio indicates that women in the region residing in the rural areas enter parenthood at an almost 2 times higher rate than women residing in the rural areas, at a significance level of 1%.

The only significant cohort for this region is level 4. Women belonging to the cohort enter parenthood at nearly 2 times higher rate than the youngest women at a significance level of 5%. The other cohorts are not significant and are interpreted the same i.e., there is no statistical evidence that women belonging to the other cohorts enter parenthood at another rate than women belonging to cohort 1.

Women with primary education do not differ from women with no education in the event of entering parenthood as the estimated intensity ratio for primary education is not statistically significant and therefore there is not enough evidence to state otherwise. Women with secondary education and higher education are statistically significant on all levels and have a difference in entering parenthood compared to women with no education. Women with secondary education enter parenthood at a 41% slower rate and women with higher education enter parenthood at a 59% slower rate, both individually compared with women with no education.

4.4.10 Addis Ababa City Administration

Addis Ababa is the capital of Ethiopia and is considered as urban i.e., there are no women from Addis Ababa that are considered as residing in a rural area and therefore there is no estimation for that level.

Estimated intensity ratios for all levels of birth cohort are significant, at least a significance level of 5%. Women belonging to any birth cohort but cohort 1 will enter parenthood at a higher rate regardless. Cohort 6 enters parenthood at almost 9 times higher rate than cohort 1. Cohort 7 enters parenthood at an 8 times higher rate than cohort 1. Cohort 5 and 4 are very similar with a higher rate of nearly 7 times than cohort 1. Cohort 3 enters parenthood at nearly 6 times higher rate whilst cohort 2 enters parenthood with a nearly 3 times higher rate compared to cohort 1.

The estimated intensity ratio of women with primary education is not statistically significant meaning there is no statistical evidence that there is a difference in the rate of entering parenthood for women with no education and women with primary education.

The parameters for women with secondary and higher education are both significant at all significance levels. Women with secondary education enter parenthood at a 52% lower rate and women with higher education enter parenthood at a 61% lower rate, both levels in comparison with women with no education.

4.4.11 Dire Dawa City Administration

Dire Dawa is the only region where all estimated intensity ratios are statistically significant on all levels.

The intensity ratio indicates women enter parenthood at a 2 times higher rate, at a significance level of 1%, if residing in the rural areas compared to women living in the urban areas.

The estimated intensity ratios for birth cohorts are all significant at a level of 1%. All cohorts enter parenthood at a higher rate than cohort 1. Cohort 5 enters parenthood at approximately 4 times higher rate than cohort 1. Cohort 7 and 2 enter parenthood at a 3 times higher rate than cohort 1.

Women with primary education enter parenthood at an 11% lower rate than women with no education. Women with secondary education enter parenthood at a 51% lower rate and higher education at a 64% lower rate.

4.5 Results from Cox model with frailty for each region and city administration

Table 4.3 and 4.4 presents the estimated intensity ratios of the gamma- and log-normal frailty models in every region.

4.5.1 Results from Cox model with gamma distributed frailty

When assuming that the frailties have a gamma distribution, the estimated intensity ratios in the gamma model barely vary from the estimated intensity ratios in the Cox PH models of the regions. This is the case in the regions Tigray, Amhara, Oromia, Somali, Benishangul, SNNPR, Gambela and Harari, where the estimated intensity ratios do not have noticeable differences. The p-value in the gamma frailty model for those regions are very high and indicates that the correlation within the households is small and insignificant. The random effects of each household are not statistically different. When assuming that the frailties are gamma distributed, the variance of the households in mentioned regions are small indicating that there is no difference between the intensities of entering parenthood between the households.

For the remaining three regions, Afar, Dire Dawa and Addis Ababa there is a difference in the estimated parameters that are worth mentioning. The estimated intensity ratio of the Afar women living in rural areas increases slightly by 0.09 in the gamma frailty model and the intensity ratio of cohort 3 increases by 0.05 in the Afar gamma frailty model compared to the Afar Cox PH model. The remaining estimated intensity ratios of the Afar gamma frailty model remain indifferent from the Cox PH model. The p-value of the household effects in Afar is 0.18 which is high and is not significant.

For the city of Dire Dawa, there is an increase in the estimated intensity ratio in the gamma frailty model compared to the Cox PH model. For women residing in rural areas, the increase is 0.08, for women belonging to cohort 5 the increase is 0.15 and for women belonging to cohort 7 the increase is 0.18. The p-value, 0.13 in the frailty gamma model, is close to being significant at the 10% level but is not.

The capital Addis Ababa has the biggest difference compared to all regions with some increasing estimated intensity ratios in the frailty gamma model compared to the Cox PH model. The increase is in the cohorts where all cohorts, besides cohort 2, have an increase that is observable. Women belonging to cohort 3 have an increase of 0.18, 4th cohort has an increase of 0.19, 5th cohort has an increase of 0.35, 6th cohort has an increase of 0.48 and 7th cohort has an increase of 0.58. The p-value is 0.1 which is at a significance level of 10% and significant. The variance is 0.097 and is the highest variance detected amongst the regions indicating that the intensity ratio of the covariates differs depending on the household. At an acceptable significance level, accounting for correlation within households changes the estimated intensities of entering parenthood.

4.5.2 Results from Cox model with log-normal distributed frailty

When assuming that the frailties have a log-normal distribution, the estimated intensity ratios in the log-normal model vary more than the gamma model when comparing the estimated intensity ratios in the Cox PH model. In the regions Tigray, Afar, Amhara, Oromia, Somali, SNNPR, Gambela and Harari the estimations are barely noticeable. The p-values in the above-mentioned regions are high and the correlation within households are not significant. The

remaining three regions, Benishangul, Addis Ababa and Dire Dawa, have differences in the estimates that are worth mentioning.

In Benishangul an increase in the estimated intensity ratio, from the Cox PH model to the log-normal frailty model, is seen in cohort 3 up to cohort 7. In cohort 3 the increase is 0.13, cohort 4 0.17, cohort 5 0.08, cohort 6 0.15 and cohort 7 0.17. The p-value is 0.097 indicating that the correlation within households is statistically significant at the 10% level. The variance of the household effects is 0.09 signifying that the rate of entering parenthood differs depending on the household.

Addis Ababa has increased estimated intensity ratios in the log-normal frailty model. Similar to the Benishangul region, the differences are distinguishable from cohort 3 up to cohort 7. In cohort 3 the increase is 0.21, cohort 4 0.22, cohort 5 0.39, cohort 6 0.51 and cohort 7 0.67. The p-value is 0.051 indicating that the correlation within households is statistically significant at the 10% level. The variance of the household effects is 0.13 signifying that the rate of entering parenthood differs depending on what household the woman belongs to.

Dire Dawa has similar results as the two previous regions, where the increase in the estimated intensity ratio in the log-normal frailty model (once again compared to the Cox PH intensity ratio estimates) can be seen in cohort 3 to 7. Cohort 3 has an increase of 0.05, cohort 4 0.06, cohort 5 0.21, cohort 6 0.09 and cohort 7 0.24. The p-value is 0.045 indicating that the correlation within households is statistically significant at the 5% level in Dire Dawa. The variance of the household effect is 0.13 signifying that the rate of entering parenthood differs depending on what household the woman belongs to.

The following section will discuss the results presented above.

5. Detailed discussion of results

5.1 Validity of the proportional hazards assumption

The results when testing the proportional hazards assumption in section 4.1 indicates that the Cox PH model may not be the best suitable when covariate education is included in the regression model. This covariate was a variable that was known to violate the assumption before using the Schoenfeld residuals method to get the significance levels of the correlations. However, somewhat unexpected was that all the other covariates used in the modelling were proven to be correlated with time in at least one region Cox regression model. Allison writes, as mentioned in section 3.2, that concerns about these violations may be considered excessive as estimated coefficients for the variables that do not achieve the proportionality assumption are an approximate average effect over the range of times observed in the data. This implies that the Cox PH model is robust, even if the proportional hazards assumption is violated.

The frailty models are a version of the Cox PH model as it includes a random effect, in this study the random effect is the effect of the different households on the intensity of women giving birth to their first child. As the Cox model with mixed effects is a nested model of the Cox PH model, performing the later models enables simpler comparison between the two and analyses if household effect is something that needs to be taken into consideration when modelling similar models. Therefore, are the Cox PH models still performed in this study, regardless of the assumption violations.

If the purpose is to study what model fits the dataset and explains the intensity and timing of first birth for women in Ethiopia, another model may be suggested. When the underlying distribution of the exposure times are unknown and the proportionality assumptions of the covariates are violated, there is another model called extended Cox model. Although an interesting issue about what model would fit the data the best, it is not investigated in this study as the purpose is to see the covariates and household effect on the timing of motherhood for Ethiopian women.

5.2 Regional similarities and differences in intensity of transition to parenthood

The intensity of entering parenthood for women residing in rural areas is not significantly different from the women residing in urban areas in the Tigray, Amhara, Benishangul and SNNPR region. This is not the case for women in the Afar, Gambela, Harari and Dire Dawa regions where there is a significant difference for women residing in the rural areas at a 1% significance level. i.e., the women residing in rural areas enter parenthood at a different rate compared to the women residing in urban areas. Being the capital city, Addis Ababa is considered as urban and therefore rural is left blank. Dire Dawa is the region where the estimated intensity ratio of women residing in rural areas, compared to the urban, is highest across regions as their intensity is twice as high. The lowest estimated intensity ratio is the one estimated in SNNPR where the women residing in the rural areas enter parenthood at an 11% lower rate compared to those residing in urban areas. Note that the estimated intensity ratio in SNNPR is not significant at any levels.

All cohorts are significant for the regions Tigray, Amhara, Oromia, Benishangul, SNNPR, Gambela, Addis Ababa and Dire Dawa. Nearly all of them are significant at all levels. The remaining regions Afar and Somali have three significant estimated intensity ratios each, whilst Harari only has one estimation that is significant.

Addis Ababa differs the most from the rest of the regions where the intensity increases greatly with each cohort in comparison to the other regions. In cohort 2 the estimated intensities do not vary much between the regions. For instance, the estimated intensity of cohort 2 in Addis Ababa is 2.6 times higher, 2.5 times higher in Benishangul and 2.4 times higher in Amhara compared to the youngest cohort of each respective region. The same cohort has an approximately 2 times higher intensity in the Tigray, Oromia, SNNPR, Gambela and Dire

Dawa regions. In cohort 3 the differences between regions in the estimated intensity ratios is evident as it is nearly 6 in Addis Ababa. Whilst the second highest estimated intensity ratio in cohort 3 is in the Amhara region at 3. The lowest ratio among the regions in cohort 3 is found in the Somali region with an estimation intensity ratio of 1.49, which translates to 50% higher intensity compared to the youngest women. Addis Ababa continues to have a remarkably high estimated intensity ratio with each birth cohort level compared to all the regions.

When observing all the regions, excluding Addis Ababa, the estimated intensity ratios of the cohorts are most similar in the Amhara, SNNPR and Dire Dawa regions. The estimated intensity ratios of the cohorts are similar in the Tigray and Benishangul regions as well. Out of all regions, Oromia and Gambela are the most similar regions when comparing estimated intensity ratios of the cohorts.

All the statistically significant education level intensity ratios are below 1, i.e., women of all education levels enter parenthood at a lower rate compared to the women with no education in all regions. All levels of education are significant at the 5% level in Amhara, Oromia, SNNPR and Dire Dawa. Primary level is not significant in Tigray, Afar, Benishangul, Harari and Addis Ababa. Secondary and higher education level is significant at the 1% level in all regions except in Gambela where secondary education is not significant at any level. Both primary and secondary levels are not significant in Somali and Gambela where only the highest level of education is significant on 1% significance level.

With each level of education, the estimated intensity ratio decreases in all regions. The lowest significant estimated intensity ratio of primary education is in the Oromia region where the intensity of entering parenthood is 32% lower compared to those not educated in the region. Dire Dawa has the highest intensity ratio of primary education at 88%. At the primary education level, the estimated intensity ratios for region SNNPR and Dire Dawa are similar.

The lowest intensity ratio of secondary level education is in the Amhara region where the estimated intensity ratio is 32%. The highest intensity ratio is found in the Harari region, where the estimated intensity ratio is 59%. The Amhara, SNNPR and Oromia regions have similar estimated intensity ratios for the secondary education level. Afar, Addis Ababa, Dire Dawa and Benishangul are other regions with similar estimates.

The estimated intensity ratio for the higher education level varies slightly between regions. Benishangul is the region with the lowest estimated intensity ratio at 0.13 and Gambela is the region with the highest at 0.54. Tigray and SNNPR have the most similar intensity ratio estimates.

5.3 Effect of clustering of women within households

An interesting result is that the inclusion of random effects into the models barely change the estimated intensity ratios, indicating that households do not have a significant effect on the timing of birth of the first child in the majority of the regions and not in Ethiopia overall. When analysing the model containing women from all regions, this issue may be because there is an

insufficient number of households in the dataset considering the number of women. The approximate average number of women in a household is 31 and possibly fewer women in each household would change the results. However, after dividing the dataset into regions and reducing the average amount of women in a household considerably, significant household effects are only achieved in the Addis Ababa region with the gamma frailty model and Addis Ababa, Dire Dawa and Benishangul with the log-normal frailty model. The results of the region divided frailty models contradict the hypothesis that lower average women per household would increase the variation within households and increase their significance.

Overall, it is evident that households do not have a significant effect on women's timing of parenthood in most regions. Dire Dawa and Addis Ababa are both cities and therefore might be heterogeneous as they are inhabited by a large number of people from different ethnic groups and religions. Because of the diverse ethnic background, households have a significant effect in these cities. This also applies to the region of Benishangul where the population is heterogeneous and could be the reason why accounting for households in the log-normal frailty model had an effect in the region. The households in the different regions may not be entirely homogeneous, but they are not as diverse as the cities. For example, the Somali region is almost entirely inhabited by Somali ethnic groups, the Oromia region is almost entirely inhabited by the Oromo ethnic group, the Afar region is almost entirely inhabited by the Afar ethnic group, the Amhara region by the Amhara ethnic group etc. This may be the cause of the low variance and insignificant results of the households even after reducing the average amount of women per household by performing the frailty models at regional levels.

Including a household effect may be more efficient when there is a belief that there is a multilevel effect that should be taken into consideration. When the dataset is from a diverse set of women with different backgrounds, like city administrations, the inclusion of the household effect is beneficial when analysing the effect of different covariates on women's intensities of entering motherhood. If the region is homogeneous in ethnic groups and background, accounting for clustering may not be a necessity.

Although the intensity ratio in the majority of regions and city administrations models barely changed with the inclusion of household effect, it should be mentioned that, although small, the absolute value of each parameter increased. By ignoring the effect of clustering, the effects of the covariates on the intensity of transitioning to parenthood may be underestimated and therefore should be analysed when doing these types of studies.

5.4 Choice of distribution for the frailty term

Another interesting result is that the estimated intensity ratios do not vary much between the gamma- and the log-normal frailty model, some coefficients do, however. The Benishangul region is where the estimated coefficients vary the most and the p-value changes drastically from being non-significant to significant when fitting the log-normal frailty model instead of the model with gamma distributed frailties. The intensity ratios of the cohorts are the ones that change the most as cohort 3 to 7 increase by 0.07 to 0.15 when fitting the former model instead

of the latter. The estimated intensity ratio of Cohort 7 in the Addis Ababa frailty models increases by 0.1 when fitting the log-normal frailty model instead of the gamma frailty model. In the Dire Dawa models cohort 7 and 5 increase by 0.6.

The remaining intensities do not increase or decrease by more than 0.05. Most household effect's p-values decrease when fitting the log-normal frailty model instead of the gamma one. The exceptions are the Afar and Somali models.

With our results, it could be implied that misidentifying the underlying distribution of the random effects does not affect the results substantially. The log-normal frailty model seems to identify the household variation slightly better compared to the gamma frailty model, however, this is not applied for all regions as the Somali and Afar region seem to be better fitted with the gamma frailty model when it comes to identifying the household effect on the intensity.

The following and last section presents a summary and the conclusions of this study.

6. Summary and conclusions

When a woman decides to enter motherhood has a direct economic consequence not only for the mother and child but for the whole society as well. Therefore, it is important to study the underlying factors influencing a woman's decision under different conditions. This paper attempted to do an empirical study of the effects of residence area, birth cohort and education level on the rate Ethiopian women enter parenthood whilst taking account of clustering of women in households on a regional level and the entire country. We performed survival analysis on the intensity of entering parenthood in Ethiopia by performing the Cox proportional hazard model and the frailty models on the entire country and on the regional level. The data was provided by the 2016 Ethiopian Demographic and Health Survey.

Our results show that the effects of the covariates vary depending on the region. In the regions overall, women residing in rural areas enter parenthood at a higher intensity compared to women residing in urban areas, the older cohort enters parenthood at a higher intensity compared to the youngest birth cohort and education has a negative effect on the estimated intensity of entering parenthood, as the estimated intensity ratio of the women decreases with each level of education. The results of this study clearly show that in the 2016 EDHS dataset, the inclusion of a household effect does not primarily change the estimated intensity ratios of the covariates. When accounting for clustering in the frailty models, the estimated intensity ratios of the covariates only differ in the regions where the populations are heterogeneous, compared to the estimates in the standard Cox model. The regions are the city administrations Addis Ababa and Dire Dawa and Benishangul are the only regions where household effects are significant. The other regions are more homogeneous.

Based on the above results we can draw the following conclusions:

- Women residing in rural areas enter parenthood at a higher rate compared to women residing in urban areas.
- All cohorts enter parenthood at a higher rate compared to the youngest cohort born 1997 to 2001.
- Education has a decreasing effect on the intensity of entering parenthood
- The multilevel survival analysis is the best fit on the heterogeneous cities and regions

In future studies, more covariates can be used. Interesting covariates to analyse in heterogeneous cities and regions are religion and wealth for example. Performing a multilevel analysis on those covariates could provide insight and interesting results. Another suggestion is to analyse different distributions of the frailties and to analyse what distribution fits the data best. Having a dataset with more households would increase the variance of the household effect and result in significant household effects.

References

- Alemayehu, A. (2019). *Why should Ethiopians care about urbanization? Jobs, infrastructure, and formal land and housing*. Worldbank. Available at: <https://blogs.worldbank.org/african/why-should-ethiopians-care-about-urbanization-jobs-infrastructure-and-formal-land-and-housing> [Accessed 2021-03-25]
- Allison, P D (2010). *Survival Analysis Using SAS: A Practical Guide, Second Edition*, Cary, North Carolina: SAS Institute Inc.
- Austin, P C (2017). A Tutorial on Multilevel Survival Analysis: Methods, Models and Applications, *International Statistical Review*, 85, 2, 185-203
- Balan, T A and Potter, H (2020). *A tutorial on frailty models*. Available at: <https://doi.org/10.1177/0962280220921889> [Accessed: 2021-03-20]
- Bui, Q and Miller, C C (2018). *The age that women have babies: How a gap divides America*. The New York Times. Available at: <https://www.nytimes.com/interactive/2018/08/04/upshot/up-birth-age-gap.html> [Accessed 2021-04-21]
- Central Statistical Agency (CSA) [Ethiopia] and ICF (2016). *Ethiopia Demographic and Health Survey 2016*. Addis Ababa, Ethiopia, and Rockville, Maryland, USA: CSA and ICF.
- Chernet, A G, Shebeshi, D S and Banbeta, A (2019). *Determinant of time-to-first birth interval after marriage among Ethiopian women*. BMC Women's Health 19, 157. Available at: <https://doi.org/10.1186/s12905-019-0858-3> [Accessed 2020-04-20]
- Collett, D (2015). *Modelling Survival Data in Medical Research, Third Edition*, Florida: CRC Press
- Cox, D R (1972). *Regression Models and Life Tables (with Discussion)*. Journal of the Royal Statistical Society, Series B 34:187—220.
- Eurostat (2020). *Women are having their first child at an older age*. Available at: <https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20200515-2> [Accessed 2021-04-20]
- Ghilagaber, G (2021). *Regression-type models - Cox Ph Model*. Lecture 5. In the course: Analysis of survival data with Demographic. Provided by Stockholm University
- Hidayat, R, Sumarno, H and Nugrahani, E H (2014). *Survival Analysis in Modeling the Birth Interval of the First Child in Indonesia*. Open Journal of Statistics, 4, 198-206. Available at:

https://www.researchgate.net/publication/269796257_Survival_Analysis_in_Modeling_the_Birth_Interval_of_the_First_Child_in_Indonesia [Accessed 2021-03-29]

Kaplan, E L, Meier P (1958). Nonparametric estimation from incomplete observations. *J.Amer.Statist.Assoc.* Vol 53, 457-481

Kunnuji, M, Eshiet, I and Nnorom, C P (2018). *A survival analysis of the timing of onset of childbearing among young females in Nigeria: are predictors the same across regions?*

Available at:

https://www.researchgate.net/publication/328326492_A_survival_analysis_of_the_timing_of_onset_of_childbearing_among_young_females_in_Nigeria_Are_predictors_the_same_across_regions [Accessed 2021-04-12]

Mills, M, Rindfuss, R R, McDonald, P, Velde, E (2011). *Why do people postpone parenthood? Reasons and social policy incentives.* Available at:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3529638/> [Accessed 2021-04-12]

Moore, D F (2016). *Applied Survival Analysis using R.* Switzerland: Springer International Publishing

OECD (2019). *Age of mothers at childbirth and age-specific fertility.* Available at:

https://www.oecd.org/els/soc/SF_2_3_Age_mothers_childbirth.pdf [Accessed 2021-03-05]

Sahl, A (2020). *New study: Millennial women are delaying having children due to their careers.* Forbes. Available at: <https://www.forbes.com/sites/ashleystahl/2020/05/01/new-study-millennial-women-are-delaying-having-children-due-to-their-careers/?sh=40acc8ca276a> [Accessed 2021-05-01]]

Therneau T (2021). *A Package for Survival Analysis in R.* version 3.2-11. Available at:

<https://CRAN.R-project.org/package=survival> [Accessed 2021-05-13]

UNICEF (2018). *Learning and development.* Available at:

<https://www.unicef.org/ethiopia/learning-and-development> [Accessed 2021-03-25]

UNICEF (2019). *Early childbearing.* Available at: <https://data.unicef.org/topic/child-health/adolescent-health/> [Accessed 2021-03-24]

USAID (2017). *Developing Ethiopia's youth.* Available at:

https://www.usaid.gov/sites/default/files/documents/1860/Fact_Sheet_Developing_Ethiopia's_Youth_Jul_2017.pdf [Accessed 2021-03-25]

Wienke, A (2011). *Frailty Models in Survival Analysis.* Florida: Chapman n & Hall/CRC

Wilkie, J (1981). The Trend toward Delayed Parenthood. *Journal of Marriage and Family*, 43(3), 583-591. doi:10.2307/351759 [Accessed 2021-04-12]

World Health Organisation (2020). *Adolescent pregnancy*. Available at:
<https://www.who.int/news-room/fact-sheets/detail/adolescent-pregnancy> [Accessed 2021-03-25]

Worldbank (2021), *The Worldbank in Ethiopia*. Available at:
<https://www.worldbank.org/en/country/ethiopia/overview> [Accessed 2021-03-18]

Appendix A: Tables

Table 4.2: The estimated intensity ratios (relative risks) for each region of the Cox proportional hazard model

Covariates	Tigray	Afar	Amhara	Oromia	Somali	Benishangul	SNNPR	Gambela	Harari	Addis Ababa	Dire Dawa
Residence 2: Rural	1.09255	1.9131 ***	1.09428	1.28516 **	1.29125 ***	1.11915	0.88073	1.49545 ***	1.73434 ***	-	2.09797 ***
Cohort 2: 1992-1996	2.22057 ***	1.4782 **	2.41886 ***	1.88886 ***	1.49050 **	2.54257 ***	1.97729 ***	1.96573 ***	1.34856	2.5858 **	2.01817 ***
Cohort 3: 1987-1991	2.11281 ***	1.3058	3.18461 ***	1.94759 ***	1.49036 **	3.10862 ***	2.68693 ***	2.21596 ***	1.45865	5.8580 ***	2.53526 ***
Cohort 4: 1982-1986	2.9374 ***	1.6134 ***	3.59185 ***	2.11167 ***	1.43625 **	2.9243 ***	3.1471 ***	2.32428 ***	1.71264 **	6.7866 ***	3.00738 ***
Cohort 5: 1977-1981	2.62651 ***	1.5937 **	3.8326 ***	1.63368 ***	1.06406	2.95042 ***	3.16016 ***	1.99955 ***	1.25142	6.8228 ***	3.926 ***
Cohort 6: 1972-1976	2.97693 ***	1.2597	4.39751 ***	1.82366 ***	0.97952	2.91199 ***	3.69292 ***	1.97458 ***	1.40869	8.7573 ***	2.61256 ***
Cohort 7: 1967-1971	3.1516 ***	1.1398	3.3014 ***	1.52628 **	0.87756	3.22324 ***	3.33903 ***	1.83072 **	1.01629	8.3227 ***	3.36648 ***
Education 1: Primary	0.90824	0.8904	0.74403 ***	0.68227 ***	0.98754	0.99095	0.85396 **	1.1149	1.09077	0.8404	0.88859 ***
Education 2: Secondary	0.43222 ***	0.4664 ***	0.32222 ***	0.34162 ***	0.7536	0.49803 ***	0.32965 ***	0.91349	0.59185 ***	0.4763 ***	0.48614 ***
Education 3: Higher	0.26147 ***	0.4398 ***	0.20207 ***	0.31753 ***	0.3788 ***	0.13485 ***	0.26316 ***	0.54052 ***	0.41244 ***	0.39 ***	0.36003 ***

Table 4.3: The estimated intensity ratios (relative risks) for each region of the Cox model with gamma distributed frailty

Covariates	Tigray	Afar	Amhara	Oromia	Somali	Benishangul	SNNPR	Gambela	Harari	Addis Ababa	Dire Dawa
Residence 2: Rural	1.0943	2.0045 ***	1.0895	1.2871 **	1.2948 ***	1.1212	0.8902	1.5001 ***	1.7432 ***	-	2.177 ***
Cohort 2: 1992-1996	2.2251 ***	1.5133 **	2.4171 ***	1.8934 ***	1.4926 **	2.5506 ***	1.9785 ***	1.9712 ***	1.3504	2.6233 **	2.0007 **
Cohort 3: 1987-1991	2.116 ***	1.3231 *	3.1894 ***	1.9517 ***	1.4922 **	3.1278 ***	2.6925 ***	2.2213 ***	1.4622	6.0416 ***	2.5791 ***
Cohort 4: 1982-1986	2.9478 ***	1.6541 ***	3.6073 ***	2.1172 ***	1.4378 **	2.9466 ***	3.1563 ***	2.3328 ***	1.7189 **	6.9734 ***	3.0458 ***
Cohort 5: 1977-1981	2.6346 ***	1.6426 ***	3.8398 ***	1.6327 ***	1.0629	2.9655 ***	3.1709 ***	2.0033 ***	1.2508	7.1682 ***	4.0776 ***
Cohort 6: 1972-1976	2.9908 ***	1.2361	4.4229 ***	1.8244 ***	0.9779	2.933 ***	3.7097 ***	1.9778 ***	1.4088	9.2409 ***	2.6756 ***
Cohort 7: 1967-1971	3.1669 ***	1.1491	3.2975 ***	1.5239 **	0.8748	3.2476 ***	3.3518 ***	1.8322 **	1.0155	8.898 ***	3.5422 ***
Education 1: Primary	0.9077	0.8947	0.7367 ***	0.6796 ***	0.9871	0.9929	0.8529 **	1.1151	1.0916	0.829	0.8882
Education 2: Secondary	0.4296 ***	0.4528 ***	0.3167 ***	0.3392 ***	0.7523	0.4963 ***	0.3279 ***	0.9116	0.5903 ***	0.4511 ***	0.4634 ***
Education 3: Higher	0.2595 ***	0.4113 ***	0.1968 ***	0.3154 ***	0.3769 ***	0.1332 ***	0.2615 ***	0.538 ***	0.4103 ***	0.37 ***	0.3412 ***
Frailty p-value	0.34	0.18	0.37	0.54	0.32	0.33	0.86	0.6	0.31	0.1	0.13
Variance of random effects	0.002030249	0.08150213	0.01323474	0.00005	0.001291504	0.004449385	0.0000005	0.00005	0.002121251	0.09709179	0.08474129

Table 4.4: The estimates intensity ratios (relative risks) for each region of the Cox model with log-normal distributed frailty

Covariates	Tigray	Afar	Amhara	Oromia	Somali	Benishangul	SNNPR	Gambela	Harari	Addis Ababa	Dire Dawa
Residence 2: Rural	1,101	1,9763 ***	1,085	1,2852 **	1,2971 ***	1,1266	0,8898	1,5003 ***	1,7752 ***	-	2,2019 ***
Cohort 2: 1992-1996	2,2272 ***	1,5011 **	2,4131 ***	1,8929 ***	1,4897 **	2,5828 ***	1,979 ***	1,9731 ***	1,3668	2,6256 **	1,9908 **
Cohort 3: 1987-1991	2,1236 ***	1,3164 *	3,1863 ***	1,9508 ***	1,4954 **	3,2408 ***	2,6932 ***	2,2235 ***	1,4977	6,0655 ***	2,5875 ***
Cohort 4: 1982-1986	2,9445 ***	1,6414 ***	3,6112 ***	2,118 ***	1,4409 **	3,0953 ***	3,1575 ***	2,3368 ***	1,7754 **	7,0021 ***	3,0643 ***
Cohort 5: 1977-1981	2,677 ***	1,6257 ***	3,8348 ***	1,6317 ***	1,0628	3,0336 ***	3,1717 ***	2,0035 ***	1,2608	7,2117 ***	4,1384 ***
Cohort 6: 1972-1976	3,027 ***	1,2438	4,428 ***	1,8243 ***	0,9753	3,0624 ***	3,7126 ***	1,9789 ***	1,4605	9,2689 ***	2,7057 ***
Cohort 7: 1967-1971	3,2106 ***	1,1454	3,2858 ***	1,5229 **	0,8728	3,3979 ***	3,3531 ***	1,8356 **	1,0161	8,9953 ***	3,6053 ***
Education 1: Primary	0,9013	0,8926	0,7323 ***	0,6789 ***	0,9885	1,024	0,8527 **	1,1155	1,0955	0,8247	0,8876
Education 2: Secondary	0,4198 ***	0,4569 ***	0,3136 ***	0,3382 ***	0,754	0,4978 ***	0,3278 ***	0,9117	0,5777 ***	0,4467 ***	0,4562 ***
Education 3: Higher	0,2542 ***	0,4204 ***	0,1935 ***	0,3145 ***	0,3769 ***	0,1249 ***	0,2613 ***	0,5362 ***	0,3969 ***	0,3667 ***	0,3355 ***
Frailty p-value	0,27	0,22	0,32	0,39	0,4	0,097	0,35	0,39	0,27	0,051	0,045
Variance of random effects	0,03746311	0,05380972	0,02643687	0,002604167	0,01156039	0,08672587	0,001302083	0,005208333	0,05684628	0,1258594	0,133317

Appendix B: R-codes

```
#####  
##                                EXAMENSARBETE                                ##  
#####  
# Transporting the data into R and naming it ethiopia16  
library(readxl)  
ethiopia16 <- read_excel("C:/Users/Amanda/Desktop/Uppsats Statistik VT21/GG-Data-to-  
Amanda-Elelta.xlsx")  
View(ethiopia16)  
#####  
#####  
# Creating subsets for each region  
Tigray <- subset(ethiopia16, Region == 1)  
Afar <- subset(ethiopia16, Region == 2)  
Amhara <- subset(ethiopia16, Region == 3)  
Oromia <- subset(ethiopia16, Region == 4)  
Somali <- subset(ethiopia16, Region == 5)  
Benishangul <- subset(ethiopia16, Region == 6)  
SNNPR <- subset(ethiopia16, Region == 7)  
Gambela <- subset(ethiopia16, Region == 8)  
Harari <- subset(ethiopia16, Region == 9)  
Addis_Ababa <- subset(ethiopia16, Region == 10)  
Dire_Dawa <- subset(ethiopia16, Region == 11)  
#####  
##                                1. Kaplan-Meier estimated survival curves by region                                ##  
#####  
library(survival)  
  
RegionTigray <- which(ethiopia16$Region==1)  
regionT <- ethiopia16[RegionTigray,]  
km.regionT <- survfit(Surv(Exposure, Status)~1, data=regionT)  
  
RegionAfar <- which(ethiopia16$Region==2)  
regionAf <- ethiopia16[RegionAfar,]  
km.regionAf <- survfit(Surv(Exposure, Status)~1, data=regionAf)  
  
RegionAmhara <- which(ethiopia16$Region==3)  
regionAm <- ethiopia16[RegionAmhara,]  
km.regionAm <- survfit(Surv(Exposure, Status)~1, data=regionAm)  
  
RegionOromia <- which(ethiopia16$Region==4)  
regionO <- ethiopia16[RegionOromia,]
```

```

km.regionO <- survfit(Surv(Exposure, Status)~1, data=regionO)

RegionSomali <- which(ethiopia16$Region==5)
regionS <- ethiopia16[RegionSomali,]
km.regionS <- survfit(Surv(Exposure, Status)~1, data=regionS)

RegionBenishangul <- which(ethiopia16$Region==6)
regionB <- ethiopia16[RegionBenishangul,]
km.regionB <- survfit(Surv(Exposure, Status)~1, data=regionB)

RegionSNNPR <- which(ethiopia16$Region==7)
regionSN <- ethiopia16[RegionSNNPR,]
km.regionSN <- survfit(Surv(Exposure, Status)~1, data=regionSN)

RegionGambela <- which(ethiopia16$Region==8)
regionG <- ethiopia16[RegionGambela,]
km.regionG <- survfit(Surv(Exposure, Status)~1, data=regionG)

RegionHarari <- which(ethiopia16$Region==9)
regionH <- ethiopia16[RegionHarari,]
km.regionH <- survfit(Surv(Exposure, Status)~1, data=regionH)

RegionAddis <- which(ethiopia16$Region==10)
regionAA <- ethiopia16[RegionAddis,]
km.regionAA <- survfit(Surv(Exposure, Status)~1, data=regionAA)

RegionDire <- which(ethiopia16$Region==11)
regionDD <- ethiopia16[RegionDire,]
km.regionDD <- survfit(Surv(Exposure, Status)~1, data=regionDD)

plot(km.regionDD, main="Figure 1: Survival functions by region: KM estimation", col =
"green",
     cex.axis=0.7, xlab = "Exposure in months", ylab = "Survival functions", conf.int = "none")
lines(km.regionAA, col = "orange", conf.int = "none")
lines(km.regionH, col = "darkblue", conf.int = "none")
lines(km.regionG, col = "yellow", conf.int = "none")
lines(km.regionSN, col = "red", conf.int = "none")
lines(km.regionB, col = "blue", conf.int = "none")
lines(km.regionS, col = "black", conf.int = "none")
lines(km.regionO, col = "purple", conf.int = "none")
lines(km.regionAm, col = "pink", conf.int = "none")
lines(km.regionAf, col = "violet", conf.int = "none")
lines(km.regionT, col = "brown", conf.int = "none")

```

```

legend("topright", c("Dire Dawa", "Addis Ababa", "Harari", "Gambela",
"SNNPR", "Benishangul", "Somali", "Oromia", "Amhara", "Afar", "Tigray"),
      lty = c(1:1), cex=0.7, col = c("green", "orange", "darkblue", "yellow", "red", "blue",
"black", "purple", "pink", "violet", "brown"))

```

```

#####
##          2. COX proportional models for the whole country and for each region      ##
#####

```

```

coxall <- coxph(Surv(Exposure, Status) ~ as.factor(Region) + as.factor(Residence) +
as.factor(Cohort) + as.factor(Education), method="breslow", data = ethiopia16)
summary(coxall)

```

```

coxTigray <- coxph(Surv(Exposure, Status) ~ as.factor(Residence) + as.factor(Cohort) +
as.factor(Education), method="breslow", data = Tigray)
summary(coxTigray)

```

```

coxAfar <- coxph(Surv(Exposure, Status) ~ as.factor(Residence) + as.factor(Cohort) +
as.factor(Education), method="breslow", data = Afar)
summary(coxAfar)

```

```

coxAmhara <- coxph(Surv(Exposure, Status) ~ as.factor(Residence) + as.factor(Cohort) +
as.factor(Education), method="breslow", data = Amhara)
summary(coxAmhara)

```

```

coxOromia <- coxph(Surv(Exposure, Status) ~ as.factor(Residence) + as.factor(Cohort) +
as.factor(Education), method="breslow", data = Oromia)
summary(coxOromia)

```

```

coxSomali <- coxph(Surv(Exposure, Status) ~ as.factor(Residence) + as.factor(Cohort) +
as.factor(Education), method="breslow", data = Somali)
summary(coxSomali)

```

```

coxBenishangul <- coxph(Surv(Exposure, Status) ~ as.factor(Residence) + as.factor(Cohort)
+ as.factor(Education), method="breslow", data = Benishangul)
summary(coxBenishangul)

```

```

coxSNNPR <- coxph(Surv(Exposure, Status) ~ as.factor(Residence) + as.factor(Cohort) +
as.factor(Education), method="breslow", data = SNNPR)
summary(coxSNNPR)

```

```

coxGambela <- coxph(Surv(Exposure, Status) ~ as.factor(Residence) + as.factor(Cohort) +
as.factor(Education), method="breslow", data = Gambela)
summary(coxGambela)

```

```
coxHarari <- coxph(Surv(Exposure, Status) ~ as.factor(Residence) + as.factor(Cohort) +
as.factor(Education), method="breslow", data = Harari)
summary(coxHarari)
```

```
coxAddis_Ababa <- coxph(Surv(Exposure, Status) ~ as.factor(Cohort) +
as.factor(Education), method="breslow", data = Addis_Ababa)
summary(coxAddis_Ababa)
```

```
coxDire_Dawa <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)+ as.factor(Cohort) +
as.factor(Education), method="breslow", data = Dire_Dawa)
summary(coxDire_Dawa)
```

```
#####
##                               3. Testing the proportional hazards assumptions                               ##
#####
```

```
cox.zph(coxall)
cox.zph(coxTigray)
cox.zph(coxAfar)
cox.zph(coxAmhara)
cox.zph(coxOromia)
cox.zph(coxSomali)
cox.zph(coxBenishangul)
cox.zph(coxSNNPR)
cox.zph(coxGambela)
cox.zph(coxHarari)
cox.zph(coxAddis_Ababa)
cox.zph(coxDire_Dawa)
```

```
#####
##                               4. Frailty models for the whole country and for each region                               ##
#####
```

```
#Gamma- and log-normal frailty distributed models
#Ethiopia
```

```
GammaAll <- coxph(Surv(Exposure, Status) ~ as.factor(Region) + as.factor(Residence) +
as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,distribution="gamma"),
method="breslow", data = ethiopia16)
summary(GammaAll)
```

```
lognormalAll <- coxph(Surv(Exposure, Status) ~ as.factor(Region) + as.factor(Residence) +
as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,distribution="gaussian"),
method="breslow", data = ethiopia16)
summary(lognormalAll)
```

```
#Tigray
```

```
GammaTigray <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)+ as.factor(Cohort) +
as.factor(Education) + frailty(HouseHold,distribution="gamma"), data=Tigray)
```

```
summary(GammaTigray)
```

```
lognormTigray <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)+ as.factor(Cohort) +  
as.factor(Education) + frailty(HouseHold,distribution="gaussian"), data=Tigray)  
summary(lognormTigray)
```

```
#Afar
```

```
GammaAfar <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)+ as.factor(Cohort) +  
as.factor(Education) + frailty(HouseHold,distribution="gamma"), data=Afar)  
summary(GammaAfar)
```

```
lognormAfar <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)+ as.factor(Cohort) +  
as.factor(Education) + frailty(HouseHold,distribution="gaussian"), data=Afar)  
summary(lognormAfar)
```

```
#Amhara
```

```
GammaAmhara <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)+ as.factor(Cohort) +  
as.factor(Education) + frailty(HouseHold,distribution="gamma"), data=Amhara)  
summary(GammaAmhara)
```

```
lognormAmhara <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)+ as.factor(Cohort)  
+ as.factor(Education) + frailty(HouseHold,distribution="gaussian"), data=Amhara)  
summary(lognormAmhara)
```

```
#Oromia
```

```
GammaOromia <-coxph(Surv(Exposure, Status) ~ as.factor(Residence)+ as.factor(Cohort) +  
as.factor(Education) + frailty(HouseHold,distribution="gamma"), data=Oromia)  
summary(GammaOromia)
```

```
lognormOromia <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)+ as.factor(Cohort) +  
as.factor(Education) + frailty(HouseHold,distribution="gaussian"), data=Oromia)  
summary(lognormOromia)
```

```
#Somali
```

```
GammaSomali <-coxph(Surv(Exposure, Status) ~ as.factor(Residence)+ as.factor(Cohort) +  
as.factor(Education) + frailty(HouseHold,distribution="gamma"), data=Somali)  
summary(GammaSomali)
```

```
lognormSomali <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)+ as.factor(Cohort) +  
as.factor(Education) + frailty(HouseHold,distribution="gaussian"), data=Somali)  
summary(lognormSomali)
```

```
#Benishangul
```

```
GammaBenishangul <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)+  
as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,distribution="gamma"),  
data=Benishangul)  
summary(GammaBenishangul)
```

```
lognormBenishangul <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)+  
as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,distribution="gaussian"),  
data=Benishangul)  
summary(lognormBenishangul)
```

```
#SNNPR
```

```
GammaSNNPR <-coxph(Surv(Exposure, Status) ~ as.factor(Residence)+ as.factor(Cohort) +  
as.factor(Education) + frailty(HouseHold,distribution="gamma"), data=SNNPR)  
summary(GammaSNNPR)
```

```
lognormSNNPR <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)+ as.factor(Cohort)  
+ as.factor(Education) + frailty(HouseHold,distribution="gaussian"), data=SNNPR)  
summary(lognormSNNPR)
```

```
#Gambela
```

```
GammaGambela <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)+ as.factor(Cohort)  
+ as.factor(Education) + frailty(HouseHold,distribution="gamma"), data=Gambela)  
summary(GammaGambela)
```

```
lognormGambela <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)+ as.factor(Cohort)  
+ as.factor(Education) + frailty(HouseHold,distribution="gaussian"), data=Gambela)  
summary(lognormGambela)
```

```
#Harari
```

```
GammaHarari <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)+ as.factor(Cohort) +  
as.factor(Education) + frailty(HouseHold,distribution="gamma"), data=Harari)  
summary(GammaHarari)
```

```
lognormHarari <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)+ as.factor(Cohort) +  
as.factor(Education) + frailty(HouseHold,distribution="gaussian"), data=Harari)  
summary(lognormHarari)
```

```
#Addis Ababa
```

```
GammaAddis_Ababa <- coxph(Surv(Exposure, Status) ~ as.factor(Cohort) +  
as.factor(Education) + frailty(HouseHold,distribution="gamma"), data=Addis_Ababa)  
summary(GammaAddis_Ababa)
```

```
lognormAddis_Ababa <- coxph(Surv(Exposure, Status) ~ as.factor(Cohort) +
as.factor(Education) + frailty(HouseHold,distribution="gaussian"), data=Addis_Ababa)
summary(lognormAddis_Ababa)
```

```
#Dire Dawa
```

```
GammaDire_Dawa <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)+
as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,distribution="gamma"),
data=Dire_Dawa)
summary(GammaDire_Dawa)
```

```
lognormDire_Dawa <- coxph(Surv(Exposure, Status) ~ as.factor(Residence)+
as.factor(Cohort) + as.factor(Education) + frailty(HouseHold,distribution="gaussian"),
data=Dire_Dawa)
summary(lognormDire_Dawa)
```

```
#####
```

```
##                               5. The different frequencies                               ##
```

```
#####
```

```
# Ethiopia
```

```
table(ethiopia16$Residence)
table(ethiopia16$Cohort)
table(ethiopia16$Education)
table(ethiopia16$Residence,ethiopia16$Status)
table(ethiopia16$Cohort,ethiopia16$Status)
table(ethiopia16$Education,ethiopia16$Status)
```

```
#Tigray
```

```
table(Tigray$Residence)
table(Tigray$Cohort)
table(Tigray$Education)
table(Tigray$Residence,Tigray$Status)
table(Tigray$Cohort,Tigray$Status)
table(Tigray$Education,Tigray$Status)
```

```
#Afar
```

```
table(Afar$Residence)
table(Afar$Cohort)
table(Afar$Education)
table(Afar$Residence,Afar$Status)
table(Afar$Cohort,Afar$Status)
table(Afar$Education,Afar$Status)
```

```
#Amhara
```

```
table(Amhara$Residence)
table(Amhara$Cohort)
```



```
table(Amhara$Education)
table(Amhara$Residence,Amhara$Status)
table(Amhara$Cohort,Amhara$Status)
table(Amhara$Education,Amhara$Status)
```

```
#Oromia
table(Oromia$Residence)
table(Oromia$Cohort)
table(Oromia$Education)
table(Oromia$Residence,Oromia$Status)
table(Oromia$Cohort,Oromia$Status)
table(Oromia$Education,Oromia$Status)
```

```
#Somali
table(Somali$Residence)
table(Somali$Cohort)
table(Somali$Education)
table(Somali$Residence,Somali$Status)
table(Somali$Cohort,Somali$Status)
table(Somali$Education,Somali$Status)
```

```
#Benishangul
table(Benishangul$Residence)
table(Benishangul$Cohort)
table(Benishangul$Education)
table(Benishangul$Residence,Benishangul$Status)
table(Benishangul$Cohort,Benishangul$Status)
table(Benishangul$Education,Benishangul$Status)
```

```
#SNNPR
table(SNNPR$Residence)
table(SNNPR$Cohort)
table(SNNPR$Education)
table(SNNPR$Residence,SNNPR$Status)
table(SNNPR$Cohort,SNNPR$Status)
table(SNNPR$Education,SNNPR$Status)
```

```
#Gambela
table(Gambela$Residence)
table(Gambela$Cohort)
table(Gambela$Education)
table(Gambela$Residence,Gambela$Status)
table(Gambela$Cohort,Gambela$Status)
table(Gambela$Education,Gambela$Status)
```

```
#Harari
table(Harari$Residence)
table(Harari$Cohort)
table(Harari$Education)
table(Harari$Residence,Harari$Status)
table(Harari$Cohort,Harari$Status)
table(Harari$Education,Harari$Status)
```

```
#Addis Ababa
table(Addis_Ababa$Residence)
table(Addis_Ababa$Cohort)
table(Addis_Ababa$Education)
table(Addis_Ababa$Residence,Addis_Ababa$Status)
table(Addis_Ababa$Cohort,Addis_Ababa$Status)
table(Addis_Ababa$Education,Addis_Ababa$Status)
```

```
#Dire Dawa
table(Dire_Dawa$Residence)
table(Dire_Dawa$Cohort)
table(Dire_Dawa$Education)
table(Dire_Dawa$Residence,Dire_Dawa$Status)
table(Dire_Dawa$Cohort,Dire_Dawa$Status)
table(Dire_Dawa$Education,Dire_Dawa$Status)
```

```
length(unique(ethiopia16$HouseHold)) #487
length(unique(Tigray$HouseHold)) #452
length(unique(Afar$HouseHold)) #410
length(unique(Amhara$HouseHold)) #452
length(unique(Oromia$HouseHold)) #466
length(unique(Somali$HouseHold)) #451
length(unique(Benishangul$HouseHold)) #415
length(unique(SNNPR$HouseHold)) #459
length(unique(Gambela$HouseHold)) #390
length(unique(Harari$HouseHold)) #383
length(unique(Addis_Ababa$HouseHold)) #442
length(unique(Dire_Dawa$HouseHold)) #402
#####
```