Survival Models: Application to predicting the Breast-feeding Duration of Mothers in Urban Addis Ababa

By

Daniel Alemayehu

A Thesis Presented to the School of Graduate Studies

Addis Ababa University

in (Part) Fulfilment for the Degree of Master of

Science in Statistics

Addis Ababa, Ethiopia

May, 2001

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

Survival Models: Application to predicting the Breast-feeding Duration of Mothers in Urban Addis Ababa

BY

DANIEL ALEMAYEHU

DEPARTMENT OF STATISTICS

FACULTY OF SCIENCE

Approved by the Examining Board:

Dr. Tadewos Koroto, Advisor ______________________

Prof. Eshetu Wencheko, Examiner ______________________

Dr. Butte Gotu, Examiner ______________________

ACKNOWLEDGMENTS

I feel very much honoured to express my deepest and most sincere gratitude and appreciation to my advisor, Dr. Tadewos Koroto, for his continuous advice and useful comments throughout this work.

My special thanks go to the Central Statistical Authority (CSA) which permitted me to use the 1995 Fertility Survey of Urban Addis Ababa in this thesis.

I would like to extend my thanks to Ato Mulugeta Gebereselassie, Department of Statistics, AAU who assisted me as my co-advisor and Ato Yared Mekonnen, Ethiopian Health and Nutrition Research Institute, whose comments have made a substantial contribution to the thesis. Thanks are also due to Ato Nigussie Geletu for his assistance on computer facilities and related works.

Particular thanks are due to my sister W/t Aster Alemayehu for her unreserved help and encouragement from the beginning to the end of the program of the study. This thesis wouldn't have been possible without the assistance provided by all staff members of the post-graduate program in the Department of Statistics. Particularly my thanks go to Professor Eshetu Wencheko and Dr. Tadewos Koroto for helping me in getting a scholarship to pursue my final year study. The DAAD and CERTWID are also acknowledged for their financial support.

Finally, I would like to express my deepest gratitude to my family and my friends for the entire moral and material support they have given me during the whole program of my study.

ABSTRACT

The benefits of breast-feeding for the health of the infant, as an inexpensive and appropriate source of nutrients, for its simulation of a strong relationship between the mother and the child, and as a major source of protection against pregnancy through suppression of ovulation have been very well documented. But, only few have attempted to examine the factors that determine the duration of breast-feeding in Ethiopia. Thus using the appropriate statistical model this study analyzes the socio-demographic factors that determine the duration of breast-feeding in Urban Addis Ababa.

Data on breast-feeding from the 1995 Fertility Survey of Urban Addis Ababa carried out by the Central Statistical Authority is used for the analysis. The breast-feeding duration of the 31.78% of the mothers is found censored. Therefore, for the overall examination of the factors the Cox-regression model is used.

The prevalence and the median duration of breast-feeding using the Kaplan-Meier method are estimated to be 95% and 14 months, respectively. The log-rank test is used to examine the breast-feeding differentials. To know the joint effect of the factors and to identify the more important factors a multivariate analysis is undertaken. An approach based on statistical modelling without prioritizing the variables, the 'Hierarchic' method is employed.

The final Cox's regression model shows that educational level of the mother is the variable which significantly affects the duration of breast-feeding. It is this variable with a negative association that best determines the breast-feeding duration. The shortest duration is observed for mothers with higher levels of education. Finally, relevant discussion is made and the focus group to which effective breast-feeding promotion campaign should be targeted is recommended.

Table of Contents

CHAPTER I

INTRODUCTION

An important distinguishing characteristic of mammals is that the female has mammae (breasts in human beings), the function of which is to secret milk for the nourishment of new-born offspring. The availability of artificial means (bottles and formula milk) has considerably reduced the dependency of infants on breast milk; however, the new-borns of all other mammalian species remain exclusively dependent on mother's milk for survival in early life. The milk produced by such species is particularly suited to the needs of new-borns of that species. It is not surprising then that breast milk is the most suitable food for the human new-born (Shah and Khanna, 1990).

The new international growth reference prepared by World Health Organization (WHO) recommends that infants should be fed exclusively on breast milk from birth to at least 4 and if possible 6 months of age. But after that they need additional food for their growth. Starting additional food does not mean stopping breast-feeds. Breast-feeding should be continued along with the additional food as long as possible even up to 18-24 months of age (Groff, 1986).

Breast-feeding duration has important effects on a wide array of demographic, maternal and infant health outcomes. Recent evidence suggests that there is currently a trend away from longer breast-feeding duration in Addis Ababa, despite the large number of studies documenting its beneficial effects on child's health and its negative impact on fertility (CSA,1993). There are multitudes of factors that may have influenced the downward course in the duration of breast-feeding. But, little has been done to identify the determinants of breast-feeding duration in Ethiopia. This paper, therefore, using the appropriate statistical model for data with censored observations examines the socio-demographic factors that determine the duration of breast-feeding in urban Addis Ababa.

1.1 Statement of the problem

Breast-feeding has been encouraged by the medical community because of its benefits for both the mother and the child. Health organisations such as WHO and UNICEF have placed breast-feeding promotion high on their agendas. As there is an abundance of literature on the advantages of breast-feeding, some of its benefits are stated below.

Breast milk contains important nutrients such as essential fatty acids for proper digestion, lactose for the proper growth of brain cells and correct balance of aminoacids (the building blocks of proteins) and an enzyme that helps digestion. Breast-feeding also reduces the likelihood of allergic reactions such as asthma and runny noses. As long as the mother is near the baby, the baby's food is always ready at the right temperature and consistency. The more intimate interaction between the breast-feeding mother and child, the warm skin-to-skin contact, and the more immediate satisfaction of the nursing baby's hunger would augur healthier psychological development (Eiger and Olds, 1987).

Breast milk has anti-infective properties that protect the infant from infection in the early months. Besides, breast-feeding is much cheaper than feeding breast milk substitutes. The cost of the extra food needed by the mother is low compared with the cost of feeding milk formula. No utensils or water (which might carry germs) is needed to prepare it.

Recently, considerable importance is being given to the study of breast-feeding practices in different settings in developing and developed countries. There are two main reasons for studying breast-feeding patterns and practices. First, breast-feeding protects infants against higher morbidity and mortality risks, particularly due to diarrhoea disease, which affect children in developing countries, who are bottle-fed. Second, breast-feeding affects birth spacing by lengthening the period of postpartum amenorhhea- the period of temporarily sterility, which follows birth- and by influencing the time from menses to conception. In societies where little contraception is practised, breast-feeding plays a major role in curtailing fertility (Diamond et al. 1986).

Child survival depends upon adequate nutrient intake and the ability of a child to resist or recover from infections. Breast-feeding can affect child survival by its role in nutrient intake, in birth spacing and its anti-infective properties. The overall impact of breast-feeding on child survival is high in developing countries, where non-contaminated food is not available to replace breast milk, where the load of infection is high and health care services are inadequate. Under these circumstances it is possible that breast-feeding could explain some or all of the effects of birth interval on child survival.

The relationship between breast-feeding and infant and child mortality has been amply documented in studies from many areas of the world. Although the magnitudes of the estimates differ from study to study and across cultures, most research in developing countries attest to the importance of breast-feeding as a determinant of child survival. In general the literature indicates that breast-fed children are less susceptible to the risk of infant and child death than artificially fed children.

There is a differential effect of breast-feeding on child survival depending on the child's age. The impact of breast-feeding on child survival through its role in nutrient intake and anti-infective properties is strongest in the early months of life, and, although significant in later months, its effect is at a greatly reduced level (Huffman and Lamphere, 1984). Based on the Nepal Fertility Survey Retherford et al. (1989) found that relative to not breast-feeding, breast-feeding reduced infant mortality (mortality from birth to the age of 18 months) by about 80 percent and child mortality (mortality from age 18 to 60 months) by about 55 percent.

According to the family planning / maternal and child health survey conducted in North-eastern Brazil Goldberg et al. (1984) revealed that the probability of dying in the first year of life was about 1.6 times greater for infants who had never been breast-fed than among infants who had been breast-fed. The proportionate differences in mortality decrease vary regularly as age increases, with the differences during the first few months being much greater than during the first year as a whole. Thus, from the above two studies it is observed that breast milk continues to be the principal source of nutrients for the first few years of life.

Prolonged breast-feeding has the effect of lengthening the time to the next birth by suppressing the return of ovulation, generally signalled by the return of the menses, and safeguarding the health of the baby by postponding the moment when it becomes entirely dependent on adult food. If there is no breast-feeding the average duration of amenorrhoea is 2 months; a duration of breast-feeding of one year is associated with an average duration of amenorrhoea of about 7 months. With 18 months of breast-feeding , amenorrhea lasts on average 10 months, and with 2 years of breast-feeding it lasts between 15 and 18 months. These figures give some indication of the importance of breast-feeding for inhibiting fertility, since a significant portion of the interval between births will consist of a long period of postporatum ammonerrhea resulting from prolonged breast-feeding. The more frequent and intense the episodes of suckling the more effectively are the menses suppressed, but the duration of lactation induced infecundity appears not to be extendible much beyond two years (Bracher and Santow, 1982; Ferry,1981).

The prevalence and duration of breast-feeding widely varies around the world. Increasing levels of modernization and industrialization are among the important contributors to the decline in breast-feeding duration. As living in urban areas by it-self exposes mothers to some other feeding practices due to the availability of breast milk substitutes in the market and also bottle-feeding being considered as a modern activity.

Like other developing countries, breast-feeding is a universal practice not only in Ethiopia but even in the metropolitan where modernization affects the practice. Results of the 1990 National Family and Fertility Survey show that in Addis Ababa 94% of mothers breast-fed their last child with an average of breast-feeding duration of 16.4 months. The corresponding figures for the other urban areas of the country were 95.8% and 20 months, respectively. There are many factors that may have influenced the downward course in the duration of breast-feeding in urban Addis Ababa. A large part of the study is thus essentially devoted to investigate the determining socio-demographic factors in the duration of breast-feeding in urban Addis Ababa and identifying the disadvantaged groups who may need more attention in the promotion of breast-feeding.

Specific studies towards the determining factors of breast-feeding duration are very small in our country. However, primary analysis by organizations like the Central Statistical Authority and few researches were carried out (Yeshewamebrat, 1995 and Abdulahi, 1990). But, considering the seriousness of the problem, its decreasing rate and having a body of data on breast-feeding information for a larger and more heterogeneous population a thorough discussion and application of further statistical analysis is important to find the socio-demographic factors that determine the duration of breast-feeding in urban Addis Ababa.

The dependent variable in the study is the duration of breast-feeding which can be censored by the date of the survey. Thus, there is a proportion of children whose true duration of breast-feeding is unknown. Survival analysis- the appropriate technique for the analysis of data with censored observations- is used. This study of breast-feeding duration is important in several different contexts. The work represents a methodological advance over previous studies of breast-feeding duration. We used the statistical methods that provide readily interpretable estimates of the magnitude of the effects of significant socio-demographic determinants of the duration of breast-feeding.

The need for this study also arises as the population policy of Ethiopia (TGE,1993) considered the promotion of breast-feeding as a means of reducing the existing child mortality and fertility in Ethiopia. Hence, the knowledge of its prevalence, its duration and the focus group to which breast-feeding promotion campaign should address is of paramount importance.

1.2 **Objectives of the study**

The main objective of the study is to identify some socio-demographic determinants in the duration of breast-feeding for those mothers who breast-fed their last child. In light of this the present study has the following specific objectives:

1. To estimate the prevalence and duration of breast-feeding in urban Addis Ababa
2. To use the appropriate statistical model for breast-feeding data with censored observations
3. To gauge the effect of various proximate determinants of duration of breast-feeding i.e. to examine the relative importance, magnitude and direction of the main covariates influencing the duration of breast-feeding

4. To compare the duration of breast-feeding among young versus older age, higher educated versus less educated, and working versus non-working mothers and
5. To provide valuable information for population policy planners and makers.

1.3 The study hypotheses

On the basis of the objectives of the study and review of related studies, the research attempts to test the following hypotheses:

1. Duration of breast-feeding shows a statistically significant difference between working and non-working mothers.
2. Longer breast-feeding duration is practised by less educated mothers than those with higher education (that is, education is inversely related to duration of breast-feeding) and
3. Young mothers breast-feed for shorter duration than older mothers.

1.4 Organization of the Study

Including this introductory chapter that defines the problem, states the objectives of the study and sets the hypotheses, the paper is organized in five chapters. A brief discussion of the revised studies is given in chapter two. Discussion about the data for the analysis , the variables expected to predict the duration of breast-feeding and the methodology applied for the study are covered in chapter three. In chapter four, the results of the data analysis, diagnostics of the fitted model and a thorough discussion of the results will lead us to the end of the study. The final chapter provides the conclusions and recommendations of the study. Finally, necessary figures are appended.

LITERATURE REVIEW

Given the well known important consequences of breast-feeding on postpartum amenorrhea and birth intervals, on the one hand, and infant mortality, on the other, the subject of the determinants of the duration of breast-feeding is getting an increasing interest. Although the birth spacing effects, the child nutrition and health importance of breast-feeding have been very well documented, only few have attempted to examine the factors that determine breast-feeding duration. The World Fertility Surveys (WFS) conducted in countries with low living conditions enabled researchers to obtain data to investigate the factors that determine the duration of breast-feeding. For instance, using breast-feeding data from WFS and Demographic and Health Survey (DHS) Trussell et al. (1992) found that education and residence (urban / rural) of mothers are the most important variables which influence the breast-feeding duration. The more educated and urban mothers were the most likely to be breast-feeding for shorter duration. Assessing the relationship among the different determinants of duration of breast-feeding by considering works done in our country and other studies conducted in other countries is the subject of this chapter.

2.1 Studies in Ethiopia

Yeshewamebrat (1995) studied the patterns of breast-feeding in a specific area, zone 1 in Addis Ababa. She collected breast-feeding data on the two most recent births (penultimate

CHAPTER II

LITERATURE REVIEW

Given the well known important consequences of breast-feeding on postpartum amenorrhea and birth intervals, on the one hand, and infant mortality, on the other, the subject of the determinants of the duration of breast-feeding is getting an increasing interest. Although the birth spacing effects, the child nutrition and health importance of breast-feeding have been very well documented, only few have attempted to examine the factors that determine breast-feeding duration. The World Fertility Surveys (WFS) conducted in countries with low living conditions enabled researchers to obtain data to investigate the factors that determine the duration of breast-feeding. For instance, using breast-feeding data from WFS and Demographic and Health Survey (DHS) Trusell et al. (1992) found that education and residence (urban / rural) of mothers are the most important variables which influence the breast-feeding duration. The more educated and urban mothers were the most likely to be breast-feeding for shorter duration. Assessing the relationship among the different determinants of duration of breast-feeding by considering works done in our country and other studies conducted in other countries is the subject of this chapter.

2.1 Studies in Ethiopia

Yeshewamebrat (1995) studied the patterns of breast-feeding in a specific zone, zone 1 in Addis Ababa. She collected breast-feeding data on the two most recent births (penultimate and ultimate). Data on education level of the mother, contraceptive use, working status, parity, age, ethnicity, religion, husband's level of education, household income and sex of the child were collected to investigate the determinants of breast-feeding duration.

One-way Analysis of Variance (ANOVA) was used to examine breast-feeding differentials independently by socio-economic and demographic characteristics of mothers taking age (15-24, 25-34 and 35-49) as the controlled variable. The relationship between education of mothers and duration of breast-feeding revealed that as the level of education increased, the duration of breast-feeding tended to decrease. Mothers with junior secondary level of education and above breast-fed 6 months shorter than who were illiterate. Similarly, mothers with junior secondary level of education and above breast-fed 5 months shorter than those with primary education. It also decreased as education level of the husband increased. Duration of breast-feeding was shorter among working mothers. Working mothers breast-fed on the average for about 4 months shorter than their non-working counterparts. Increase in income tended to decrease the duration of breast-feeding. The association between contraceptive use and the breast-feeding duration was negative. The duration was found to be shorter for contraceptive users compared to non-users.

The ANOVA results showed no significant variation of duration of breast-feeding by ethnic groups. When the duration was examined in relation to religion, differences were not statistically significant. The ANOVA results also showed a non-significant difference of breast-feeding duration by sex of the child. Besides, there seemed to be inconsistent pattern between the duration of breast-feeding and parity. It may therefore be concluded that, these variables were less important in explaining the variation in breast-feeding duration of the penultimate birth. On the other hand differentials due to education, work status, contraceptive use and household income were found to be statistically significant.

In the above bivariate analysis there was lack of estimating net effect of a particular variable on the duration of breast-feeding. To examine the net effect of each explanatory variable the effects of other variables was controlled by applying the multivariate analysis, Multiple Classification Analysis. Among the variables examined in the analysis education was found to be the most important contributor of the variation followed by contraceptive use. The duration of breast-feeding of next to last child tended to be shorter for contraceptive user mothers, for working mothers, for mothers with higher levels of education and for mothers with higher household income.

Yeshewamebrat (1995) used the life-table technique to estimate the duration of breast-feeding of the last birth. While to examine the determinants of breast-feeding she considered the next to the last birth as there were censored observations for the last birth.

In the 1990 National Family and Fertility Survey conducted by the Central Statistical Authority, data on breast-feeding behaviour was also collected on sample basis. Ever-married mothers in the child bearing age (15-49 years) who had breast-fed their last child were the source of information on breast-feeding.

The proportion of mothers who were still breast-feeding at 24 months and longer was higher among rural mothers than urban and among illiterate mothers than mothers with some education. Mothers with higher levels of education were less likely to practice prolonged breast-feeding. As the level of education of mothers increased the proportions of mothers breast-feeding for 24 months or more showed a decline. The proportion of mothers breast-feeding for longer duration did not show a significant difference between working mothers and non-working mothers.

The prevalence / incidence technique (number of children still breast-feeding at the time of the survey / the average number of births per month) was employed to estimate the mean number of months of breast-feeding duration for last births. The average duration of breast-feeding showed variation by socio-demographic characteristics of the mothers. Mothers less than 30 years old tended to breast-feed their children for longer period than mothers more than 30 years old. With respect to education the average duration was shorter for mothers with higher level of education. The rural mothers tended to breast-feed their children for longer duration on average than their urban peers. There was no significant variation in average duration of breast-feeding by work status of the mother.

Abdulahi (1990) examined the breast-feeding practices in Mettu, Alemaya and Addis Ababa. In his study ever-married mothers 15-49 years of age were asked concerning the number of months they had breast-fed their last child. In each study area it was found that the mean length of breast-feeding, in months, increased as the age of the mother advanced. A positive relationship between duration of breast-feeding and age of the mother was observed. The average duration of breast-feeding was lowest in Addis Ababa compared to the rural areas under study.

2.2 Studies in other countries

Adair et al. (1993) used a hazards model to explore the relationship of socio-demographic, biological, health sector and food industry practices on the duration of breast-feeding in a sample of infants from Cebu Longitudinal Health and Nutrition Survey. Their aim was to understand how the different factors affect breast-feeding in a large representative cohort of mothers in rural and urban areas of Cebu, the Philippines. Their analysis represented an advance in their ability to measure and model casually a full set of determinants of breast-feeding behaviour.

The included socio-demographic factors were information on household composition and structure, work and assets, and on maternal education and cultural background. Besides, variables indicating whether the mother had household help during the first six months after giving birth and whether the mother's spouse was present in the household were included. They also incorporated continuous variables indicating the mother's age and year's of education and a dichotomous variable indicating infant's sex, usage of pill and primiparity. Availability of modern conveniences such as a refrigerator, a gas or kerosene stove and water piped into the home was also included.

Most of the socio-demographic factors produced expected results. Urban residents, mothers with history of working in the modern wage sector, mothers with better cooking facilities, mothers with higher levels of education, mothers with hired help for child care, mothers with no spouse present and primiparas were likely to breast-feed for a shorter time. Except for urban residence all of the variables were statistically significant determinants of breast-feeding. With increasing maternal age, the duration of breast-feeding was lengthened. Use of contraceptives decreased the duration of breast-feeding. They found no effect of infant's sex on the duration of breast-feeding.

It was cited in Trusell et al. (1992) that urban residence has the most consistent effect and has been found to be an important determinant of breast-feeding behaviour in a number of studies. In most studies it is negatively associated with breast-feeding. Education of the mother is also an important predictor of breast-feeding but with opposite effects in developed and developing countries. In many western countries educated mothers are more likely to breast-feed and for longest, in many developing countries, education is negatively associated with both initiation and length of breast-feeding.

Socio-economic status (often measured by husband's education or occupation) has similar negative effects on breast-feeding in developing countries. Other intervening variables that are frequently studied include the mother's age, parity, use of contraceptives and employment status. The mean duration of breast-feeding increases with age and parity, though the age effects are stronger in populations in which mothers breast-feed for relatively long periods. Mothers who use modern contraceptives are less likely to be breast-feeding and more likely to breast-feed for shorter duration. The evidence linking a mother's work behaviour to breast-feeding is less clear-cut. Though, it seems obvious that work, especially outside the home, would greatly affect the mothers ability to breast-feed, the evidence in this issue is inconclusive.

Akin et al. (1986) analyzed the data that came from national probability samples of mothers survey as part of the WFS program in Jordan, Tunisia, Yemen and Egypt. Among the variables incorporated in the studies were: mother's education level, mother's work status, pill usage, age of the mother, religion, sex of the child, husband's level of education and residence status.

Using the probit technique they dealt with estimation of the determinants of breast-feeding. They found that work status variables (working and not-working) were very important for the decision to continue breast-feeding for a longer duration. Working away from home had the expected negative effect on prolonging breast-feeding. Urban residence appeared to be associated with a lower probability of breast-feeding for a long duration. But education of parents did not show the expected strong negative effects on breast-feeding. Religion continued to play no role. Boys appeared to be breast-fed for longer duration. There was evidence of a preference for boys in the Middle East. In all countries the probability of longer breast-feeding increased with age of the mother. The use of contraceptive was not associated with decision to continue breast-feeding for a longer duration.

Residence was the variable most strongly associated with differences in breast-feeding in the Middle East countries. Urban residence was the only variable clearly associated with lower probabilities of ever breast-feeding and of continuing for a longer duration. Use of contraceptives was the variable second in apparent importance associated with decrease in the likelihood of ever breast-feeding and of breast-feeding for a shorter duration (but did not have an effect beyond that).

Diamond et al. (1986) analyzed the breast-feeding data of the Pakistan Fertility Survey. The key variables in the analysis were place of residence (urban and rural), region (based on ethnic differentials), maternal literacy (as there was low levels of education in Pakistan years of education were not used), age and parity (1, 2-5 and 6 or more). Using the four explanatory variables described above the non-parametric proportional hazards models were fitted separately for urban and rural dwellers. Each of the four explanatory variables included in the models was a useful predictor of duration of breast-feeding. For literacy and age of mother the trends were similar for both urban and rural dwellers. But, there were some interesting variations for the other two explanatory variables. As one might expect, literate mothers tended to breast-feed for a shorter time than illiterate peers and the likelihood of a mother's breast-feeding increased steadily with her age, with mothers under 20 being twice as likely to wean as their peers over 40 at any given duration. The influence of a mother's region of residence varied according to whether she was urban or rural dweller. Mothers having their first birth were more likely to be breast-feeding than those at higher parities regardless of whether they were urban or rural dwellers. At the higher parities there was some evidence of urban and rural variation. There was little difference in the propensity to wean at any parity above one for urban dwellers, but the rural dwellers at parities between two and five were more likely to breast-feed than their peers at parities six and over.

Smith and Ferry (1984) following the linear regression approach analysed the breast-feeding patterns of 28 countries in Asia, Africa and Latin America. The attributes included were respondent's age, parity, residence, education, work experience, desire of another child, usage of contraceptive and the child's sex. There was an overall pattern of shorter breast-feeding duration among young and low parity mothers and longer duration among older and higher parity mothers. The mean duration of breast-feeding increased by several months from the lowest to the highest age or parity group. Urban migrants and life time urban residents breast-fed for shorter duration than rural mothers.

The mean duration of breast-feeding were higher for mothers working at home than away in most countries. Education was also highly significant, with better educated mothers breast-feeding for shorter duration than the less educated ones. Contraception usage appeared to have pronounced independent effect. In most countries the desire for additional children was associated with shorter breast-feeding duration. Child's sex seemed to be essentially random in its effects. The evidence from the 28 countries examined in their reports suggested that age, parity and residence influenced duration of breast-feeding rather strongly in most countries.

Using a fertility study undertaken in rural Central Java Bracher and Santow (1982) analyzed the data to establish an overall pattern of breast-feeding. Each mother provided information on personal characteristics such as age at marriage, age at confinement, parity, education, contraceptive use and index of wealth. Using the proportional hazards model when they regressed breast-feeding on pairs of covariates the effects of all covariates except those defined on age at confinement and education disappeared. Once these two covariates entered the regression, no other variable provided statistical significance.

Based on World Fertility Survey for Sri Lanka Akin et al. (1981) developed a probit model to find socio-economic and demographic factors that influenced breast-feeding. They selected a set of maternal, child and household characteristics. The selected variables for the analysis of determinants of breast-feeding were education, age, working status, religion, residence, husband's level of education, usage of pill, sex of the child, number of children in different age-sex group, variables which describe the general standard of living and socio-economic status of the rural population, ownership of refrigerator and toilet.

With regard to variables that specifically relate to the mother: mother's age, pill usage and mother's education had a negative association with breast-feeding. Mother's work per se, did not appear to reduce breast-feeding. Rather it was work away from home which seemed to interfere with breast-feeding, especially for extended breast-feeding. There appeared to have a negative effect associated with higher level of husband's education. The four asset variables refrigerator, toilet, being a land lord and owning a farm land had a very clear negative impact on the decision to breast-feed. The family composition variables had very mixed and interesting results. The effect of young children was generally negative while the effect of having older children to substitute for the mother in child care or market work to earn income facilitated it. On the other hand, they observed no effect of the child's sex nor the mother's religion on the duration.

There was no statistically significant difference in the impact of mother's work, education and the other factors between urban and rural areas. This was contrary to a common belief that work and education, interalia, had different effects on breast-feeding in urban and rural areas.

From the cohort analysis of the National Survey of Family Growth, Hirschman and Butler (1981) organized demographic and socio-economic differentials of the duration of breast-feeding of first and second births among American mothers who breast-fed their infants. They studied the effects of social background variables on breast-feeding duration. Effects were expressed as deviations from the mean duration of breast-feeding. The variables included were mother's birth cohort, religio-ethnicity, farm origins, region of origin, educational attainment, region of current residence, wife's occupation and husband's occupation. For second births, the length of breast-feeding of first birth was also incorporated.

After controlling for the length of breast-feeding of first birth, there was no effect of birth cohort on the duration of breast-feeding of second birth. Religio-ethinic differentials were quite modest. Farm origins had a positive but modest effect which was reduced for second births and even further after controls for other variables. For second births region of origin had negligible effects on duration. The effects of education showed a U-shaped pattern, but the strongest impact was for the least educated, not for college graduates. The effects of region of current residence almost disappeared in the comparison of net effects. Respondent's occupation and husband's occupation generated only minor differences in the duration of breast-feeding. Mothers who have not worked, those in farm labours and private household workers breast-fed somewhat longer than average. The effect of mother's working in a high-status occupation (professional or managerial) or being married to a professional was negligible. Net of all predictor variables, the most significant impact on duration of breast-feeding of second child was the length of breast-feeding of the first born.

Jain et al. (1970) studied the interrelations between lactation and postporatum amenorrhea in Taiwan from the reports of married mothers who breast-fed their last child. They measured the respective effects of mother's age, parity, education and rural-urban residence in relation to duration of breast-feeding and postporatum amenorrhea. By excluding mothers who were breast-feeding at the time of the interview, they fitted a multiple regression model. Age was found to be much more important than parity for the lactation period. After adjusting for the effects of age and parity they found that mothers who were better educated and lived in cities breast-fed for shorter periods than were the less educated and the rural mothers. Age had a positive net effect, and education and urban residence had negative net effects on lactation. But, parity had no independent effect on lactation.

2.3 Comments on the revised studies

In some of the studies having breast-feeding data with the exact duration time being unknown (censored observations) inappropriate statistical models were used. In assessing the duration of breast-feeding, the variables were separately analyzed in some of the studies. Even though the variables measure different length of duration of breast-feeding, there exists a strong correlation among them. Therefore, undertaking only univariate / bivariate analysis may cause a problem and hence all the variables must be considered jointly in order to reduce the risk of not rejecting a false hypothesis.

In the studies conducted in our country there were some drawbacks. In Yeshewamebrat's study birth was used to estimate the duration of breast-feeding and births used to determine the factors which influence the duration referred to different births (did not refer to the same birth). In the 1990 National Family and Fertility Survey only bivariate analysis was undertaken to describe the data collected on breast-feeding. Each variable was independently analyzed in relation to the duration of breast-feeding. But, without a multivariate analysis it is not possible to know which factor was more important. In Abdulahi's analysis only the age of mothers who had stopped breast-feeding their last child was analysed (censored cases were ignored).

Whenever we build a statistical model, it is important to examine the adequacy of the fitted model. In none of the revised studies goodness-of-fit of the model and its predictive ability was checked before use. Thus, in this study examination of the fitted model for possible failure will be checked. The appropriate statistical model will be applied and we will emphasise in selecting the most appropriate model that best explains the duration of breast-feeding and avoid the drawbacks appeared in some of the revised studies.

CHAPTER III

THE DATA, VARIABLES AND METHODOLOGY

The study area is urban Addis Ababa, which has six zones and 27 *weredas.* The total number of urban *kebele* is 305 and the number of rural farmers associations is 23. Out of the six zones only two have rural parts. There is no *wereda* that is entirely rural and the number of *weredas* that have rural parts is only five.

Urban Addis Ababa is selected as the study area, mainly due to the availability of a wide variety of artificial means for feeding infants within the metropolitan and this has an influence on the duration of breast-feeding. Besides, the urban female population of Addis Ababa is heterogeneous in terms of socio-demographic characteristics, which may also affect the duration. In this chapter the data used for the analysis, the variables included in the study and the methodology used to analyse the data are discussed.

3.1 The Data

The data for the study were obtained from the 1995 Fertility Survey of Urban Addis Ababa carried out by the Central Statistical Authority. The survey was designed to provide data on fertility and related characteristics of the population of Urban Addis Ababa. Data on breast-feeding practices were also collected. The prime target of the survey were women aged 15-49 years residing in conventional households. In the survey more than 2,000 women were addressed. Residents of collective quarters and the homeless were not covered.

A two-stage sample design was used for sample selection. *Weredas* were treated as strata. Enumeration Areas (EAs) and conventional households formed the primary and ultimate sampling units, respectively. Out of 2,347 EAs of urban Addis Ababa, 54 were sampled. The total sampled EAs were proportionally allocated to *weredas* and within each *wereda* the sampled EAs were selected using probability proportional to size, size being number of households obtained from the 1994 census map work. From within a selected EA, 30 households were selected using systematic sampling technique. All women aged 15-49 in the selected households were then interviewed.

When complex survey designs - surveys involving clustering, stratification and multistage sampling - are used and the sample is small, the estimates and test results based on the Cox's proportional hazards model may be unreliable. Consistent estimators are obtained by appropriately weighing the observations. However, data from large-scale surveys such as the data under consideration produce reliable results, because for large samples asymptotic properties of the estimators remain unchanged (Lehtonen and Pahkinen, 1995).

3.2 Limitations of the study

The study has some limitations, which should be noted, since they may affect our results and their interpretation. To mention the main ones:

(i) Important measures of breast-feeding behaviour, whether the infant was exclusively (full) breast-fed or partial breast-fed (which identifies the extent of breast milk relative to other milk, formula, solid foods, etc.) cannot be addressed with the available data. The necessary questions to make a clear distinction between these two types of breast-feeding were not included. Thus, in our analysis the duration of breast-feeding includes both types of breast-feeding.

(ii) As the survey was focused on fertility and family planning practices biological, health sector and food industry factors which might affect the duration of breast-feeding were not incorporated in the survey. Income of the household and sex of the last child were not included. We should bear this in mind in interpreting our results. Had we excluded important factors, which are not independent of those we have included, their omission would bias the results. However, we believe the available data include a sufficiently broad range of socio-demographic factors (the factors which we want to examine) that influence the breast-feeding duration.

(iii) The work status of a mother was classified into working and non-working groups. In both groups several analytically desirable categories are lumped together. However, this is not expected to produce serious biases in the results.

(iv) Some of the variables (such as working status, level of education, etc.) could have changed between birth and interview. But, as it is not a cohort study detailed information on such time varying explanatory variables is not available. Thus, our model is constrained by the assumption that the effect of each covariate is constant over the entire range of failure times (the validity of this assumption is to be checked).

(v) Although one of the important factors that influence the breast-feeding practice is where the mothers grew up, the explanatory variable birth place of the mother i.e. urban or rural was used to see the effect of the social and physical environment of the mother's childhood. When interpreting this sensitive variable it is on the assumption that mothers who were born in rural or urban areas had spent their childhood in the areas where they were born.

3.3 Variables of interest

3.3.1 The dependent variable

In the survey information on breast-feeding was collected for the last birth. Mothers answering 'Yes' to the question 'did you breast-feed your last child?' were asked, 'How many months?' A specific question was also asked to ascertain whether a mother was 'still breast-feeding' at the time of the survey. The duration of breast-feeding of the last child, measured by the months of breast-feeding is the dependent variable in our study.

The restriction to the last birth is imposed as, in a country like Ethiopia where registration of vital events is nearly absent, the accuracy of information on past events depends upon the time of their occurrence. The more recent the occurrence of an event the more likely the information is accurate. Breast-feeding pattern of the last child is chosen, as it is more likely to be accurately recalled for the more recent birth. Thus, using the collected data the duration of breast-feeding of the last child and its socio-demographic determinants are studied on retrospectively reported duration of breast-feeding for weaned children, along the censored duration of breast-feeding for children still being breast-fed at the time of the survey.

The retrospective data for duration of breast-feeding of the last child showed that there is a tendency to report duration in multiples of six months. Although the pattern at those time points may suggest digit preference, it may also reflect cultural norms of breast-feeding. For instance, when mothers in Zone One of Addis Ababa were asked to indicate their attitudes toward the duration of breast-feeding it was found that about 48% and 36% responded a duration of 24 and 36 months, respectively (Yeshewamebrat). This shows that the tradition seems to be breast-feeding for a duration of this length and thus it cannot totally be attributed to digit preference.

A mother who breast-fed her last child is taken as the unit of analysis. To examine the association between the values of certain explanatory variables and the duration of breast-feeding the primary response variable is the time, in months, from starting breast-feeding the last child until discontinuation. Some of the mothers had not weaned by the time the survey was conducted and so these mothers contribute right censored duration times. The coding of the variable 'duration status' of a mother is such that a zero value of the variable denotes a censored observation (mothers who were still breast-feeding at the time of the survey) and mothers who had weaned before / on 36 months are given the value one. The study rule called for the breast-feeding behaviour to be documented for a period of 36 months (Yeshewamebrat, Akin et al. 1986 and 1981, Diamond et al. 1986 and Jain et al. 1970). Thus, mothers who had breast-fed their last child for more than 36 months (3 years) are right censored and are given the value 0.

In the study of socio-demographic determinants our interest is in the duration of breast-feeding for those mothers who breast-fed their last child. Since the duration of breast-feeding can be truncated by the death of the child the influence of mortality on breast-feeding needs to be distinguished from that of socio-demographic characteristics of the mother. Mothers who weaned due to the death of the infant are thus excluded from the analysis. Therefore, we restrict our final analysis to children who survived till the time of the survey.

3.3.2 The explanatory variables

The explanatory variables or covariates included in the study can be classified into two categories:

(i) Demographic factors: includes variables such as the age of the mother and the birth order of the last child (parity)

(ii) Social factors: includes variables like birth place of the mother, ethnicity, religion, marital status, work status, contraceptive usage, status of continuous residence, behaviour of breast-feeding for the previous child, education level of the mother and the husband.

To avoid errors ordered categories are imposed on these predictor variables and for the purpose of making comparisons. A dummy variable coding scheme is used to represent dichotomous variables such as birth place, continuous residence, marital status, work status, contraceptive usage and whether or not the previous child was breast-fed. The variables with their corresponding categories are presented in the table below.

Table1. Variable Description

Variable	Variable Label	Category
agecat	Age of the mother	In completed years 1: 15-24 years 2: 25-34 years 3: 35-49 years
parcat	Birth order of the last child	1:1st 2: 2nd-5th 3: 6th or more
motbirpl	Birth place of the mother	0: urban 1: rural
con_res	Is the mother a permanent resident in Addis Ababa?	0: No 1: Yes
ethnic	Ethnicity of the mother	1: Amhara 2: Oromo 3: Gurage

Variable	Variable Label	Category
		4: Others
religion	Religion of the mother	1: Orthodox 2: Muslim 3: Others
marit	Marital status of the mother	0:Widowed / Divorced 1:Married / Living with a man
working	Work status of the mother	0: Not-working 1: Working
contrace	Contraceptive usage of the mother	0: never used 1: ever used
ptbreast	Was the previous child breast-fed?	0: No 1: Yes
educatio	Educational level of the mother	1: Literacy Program / Non-Formal 2: Primary /Junior High School 3: Senior High School / other higher education
hus_edu	Educational level of the husband	1: Literacy Program / Non-Formal

Variable	Variable Label	Category
		2: Primary /Junior High School 3: Senior High School / other higher education

3.4 **Methodology**

The analysis of factors that influence the duration of breast-feeding is statistically complex. In the standard linear regression model

$$Y_t = \beta' X_t + \varepsilon_t$$

where Y_t is the observed number of months the child is breast-fed, $\mathbf{X}_t$ is a px1 vector of factors that influence the decision, β is a px1 vector of regression coefficients, and ε_t is a random, unobserved disturbance term. Under standard conditions, we assume ε_t to be an independent, identically distributed random variable with zero mean and constant variance. If this is true and if ε_t and $\mathbf{X}_t$ are independent, the Ordinary Least Squares (OLS) method provides the best linear unbiased estimator for β.

Unfortunately, the data do not behave in such an ideal manner for two reasons. First, Y_t can be truncated by the date of the interview in the case of the last birth (open interval since that could still be breast-feeding). Thus we observe the correct duration of breast-feeding only when the mother stops breast-feeding before the interview. If at the time of the interview the child is still breast-fed, we cannot know how much longer this will continue. The effect of this truncation is that OLS estimation cannot be used because the dependent variable is not observed for a substantial portion of the data (31.78%).

The digit preference problem common in retrospective survey data also precludes the use of Ordinary Least Squares as a workable estimation technique (the data revealed a strong tendency to round towards 6 months digits although this might also be due to the normal time of weaning). Statistically if there is a digit preference it means that $\mathbf{X}_t$ and ε_t are likely to be correlated. If we ignored the correlation between $\mathbf{X}_t$ and ε_t and used the Ordinary Least Squares regression method, the resulting estimates of β would be biased and inconsistent (Akin et al. 1986). The results of any analysis of the determinants of breast-feeding based on such estimates (common in the literature) would be thus highly suspect. One solution to the above two statistical problems involves reformulating the continuous variable, duration of breast-feeding as a discrete variable. Data on the occurrences of a point event can frequently be analyzed effectively by logistic or other techniques for such binary response variables. However, these methods do not always make full use of the available data. In survival analysis (also called failure time analysis or event time analysis) we analyze the time to the occurrence of an event as well as the fact of its occurrence (Hinkley et al. 1991).

Allison (1984) also stated that although event histories are ideal for studying the causes of events, they typically possess two features - censoring and time varying explanatory variables

- that create major problems for standard statistical procedures such as multiple regression. The attempt of applying standard methods can lead to severe bias or loss of information. Thus, the methods of survival analysis are used to study the association between explanatory variables and the duration of breast-feeding of the child by the last birth.

3.4.1 **Preliminary analysis techniques**

The term survival data is being used in the context of measuring the duration of time from a well-defined time origin until the occurrence of some particular event or end point. The time taken from the starting point to the occurrence of an event is usually represented by the random variable T which is assumed to be a non-negative random variable; it is called the Survival Time or Failure Time. The realizations generated for the random variable T are referred to as survival data.

For a sample of n independent individuals, the observed data in survival analysis, are often represented by the pair (t_j, σ_j) $j=1,2,\ldots,n$ where t_j is the time an individual is known to have survived before the event and $\sigma_j \in \{0,1\}$ is an indicator of failure, assuming values of $\sigma_j = 1$ if the event of interest is observed at t_j and $\sigma_j = 0$ if the event is right censored at time t_j (event on j^{th} individual has not occurred up to time t_j).

In most areas of statistics, stochastic models for random variables are expressed in terms of densities and distribution functions. In survival analysis the equivalent functions are survivorship and hazard functions. The survivorship function denoted by S(t) is defined as the probability that an individual survives from time origin to some time beyond t (t>0) that is given by

$$S(t) = P[T>t] = 1-F(t)$$

where F(t) is a distribution function of T. A graphical presentation of S(t), called the survival curve, is a continuous function defining the survival rate at any time after zero. The median survival time and other percentiles can be found using this curve. As survival time distributions have skewed distributions the mean is not an appropriate measure of central tendency.

The hazard function of survival time T,h (t), represents the instantaneous or relative failure rate for an individual who has survived to time t. It is defined as the limit of the probability that an individual fails in a very short interval, t to t + dt, given that the individual has survived to time t.

The hazard function h(t) is thus given as

$$h(t)= \lim_{dt\to 0} P(t \le T < t+dt / T \ge t) / dt$$

Most of the survival analysis focuses only on one of the above two functions as there is a relationship between S(t) and h(t) expressed as

$$h(t)= f(t)/S(t) = -d \log S(t) / dt$$

where f(t) is the probability density function of T. If one of the two is given the other is obtained and vice versa. Mostly the hazard function is preferred for its convenience in incorporating the covariates in regression models (Lee, 1980).

As a preliminary analysis, the survival and hazard functions can be estimated from the observed survival times using the Kaplan-Meier method. This method does not require specific assumptions to be made about the underlying distribution of the survival times, that is, it is non-parametric. To apply the Kaplan-Meier method suppose failure and censored times $\{t_1,t_2,\ldots,t_n\}$ are known on n independent individuals in a random sample. Let $t_{(1)},t_{(2)},\ldots,t_{(r)}$, $r \leq n$, be the distinct ordered failure times observed among the n individuals and let d_j (j=1,2,…,r) be the corresponding number of failures with $d_j \geq 1$. The number n_j of individuals at risk is the number of individuals actually observed when failures occur, i.e. the number of individuals who have either failure or censored times greater than or equal to $t_{(j)}$, then probability of survival at time $t_{(j)}$, $p(t_{(j)})$ is estimated by

$$\hat{p}(t_{(j)}) = (n_j - d_j) / n_j$$

Then S(t), the survival probability beyond time $t_{(j)}$, is estimated by the product

$$\hat{S}(t) = \prod_{j/t(j) \leq t} (n_j - d_j) / n_j$$

$\hat{S}(t)$ only changes at time points $t_{(j)}$, j=1,2,…,r when at least a failure occurs. The corresponding graph of $\hat{S}(t)$ versus t is thus a step function. The approximated standard error of $\hat{S}(t)$ is

$$\text{S.e.}\{\hat{S}(t)\} \cong [\hat{S}(t)]\sqrt{\sum_{j=1}^{k}\frac{d_j}{n_j(n_j-d_j)}} \qquad \text{for } t_{(k)} \le t < t_{(k+1)}\ k=1,2,\ldots,r$$

Details of the expressions are given in Collet (1994).

If the hazard function is assumed to be constant between successive failure times, the hazard per unit time (risk of failure for that time) is found by dividing the ratio of failures at that time to the number of individuals at risk at that time by the time interval. Thus, if there are d_j failures at the j^{th} failure time $t_{(j)}$, $j=1,2,\ldots,r$ and n_j at risk at time $t_{(j)}$, the Kaplan-Meier estimate of the hazard function in the interval $t_{(j)}$ to $t_{(j+1)}$ is given by

$$\hat{h}(t) = d_j / n_j\tau_j \text{ for } t_{(j)} \le t < t_{(j+1)} \text{ where } \tau_j = t_{(j+1)} - t_{(j)}.$$

The survivor function of two or more groups of subjects that differ by a given characteristic can also be compared. Comparison of a survivor function of individuals in different categories of a given factor can be tested using the log- rank test (also known as the Mantel-Cox test). Here we want to test whether the estimated survivor function for each category that were found using the Kaplan-Meier procedure are the same or not. That is to check whether the different categories of the factor affect the survival time or not. The null hypothesis is typified by the hypothesis, which specifies that there is no difference between two or more groups. Thus, if there are m categories for a given factor the null hypothesis to be tested at a preassigned level of significance α is

H_o: $S_1(t) = S_2(t) = \ldots = S_m(t)$ for all t

against the alternative hypothesis that there are at least two groups with different survival curves. That is the hazard of failure at any given time for an individual in one group is proportional to the hazard at that time for an individual in the other group.

Suppose there are r distinct failure times $t_{(1)} < t_{(2)} < \ldots < t_{(r)}$, across the m categories, and at time $t_{(j)}$, d_{kj} individuals in group k fail for k=1,2,…, m, j=1,2,…, r. Also suppose that n_{kj} individuals are at risk of failure in the k^{th} category just before time $t_{(j)}$ and consequently, at time $t_{(j)}$. There are $d_j = \Sigma^m_{k=1} d_{kj}$ failures in total out of $n_j = \Sigma^m_{k=1} n_{kj}$ individuals at risk. Under the null hypothesis, the probability of failure at time $t_{(j)}$ does not depend on the category that an individual belongs. The probability of failure at time $t_{(j)}$ is d_j / n_j. Multiplying this by n_{kj} gives the expected number of failures in the category k under the null hypothesis, that is,

$$E(d_{kj}) = e_{kj} = n_{kj}(d_j / n_j) \text{ for } k=1,2,\ldots,m.$$

To combine the information from the failure times and get an overall measure of deviation of the observed values of d_{kj} from their expected values, we sum the differences $d_{kj} - e_{kj}$ over the total number of failure times, r and the resulting log-rank statistics is

$$U_k = \Sigma^r_{j=1} (d_{kj} - e_{kj}) \text{ for } k=1,2,\ldots,m-1.$$

These quantities are expressed in the form of a vector with m-1 components, denoted by $\mathbf{U}_L$ with corresponding variance-covariance matrix $\mathbf{V}_L$ as given in Collet (1994). To test the null hypothesis, we use the test statistic

$$\mathbf{U}_L' \mathbf{V}_L^{-1} \mathbf{U}_L$$

which has a chi-square distribution with m-1 degrees of freedom under the null hypothesis.

3.4.2 **Cox regression**

The Cox regression model is used to examine the effect of a variety of factors on weaning probabilities at various duration time levels. Cox regression model is similar to the well known regression model in that we predict a dependent variable (length of time to the occurrence of the event) as a function of a set of explanatory variables. However, unlike the ordinary regression model, Cox regression model can be used for data that contain censored observations.

Unlike the logistic regression model in which the dependent variable takes the value 1 if the event has occurred and 0 if it has not, the Cox regression takes into account the fact that the probability of experiencing an event differs with duration of exposure to risk. The Cox regression allows us, for instance, to treat two mothers who have identical scores in the explanatory variables except that one was breast-feeding her last child for 12 months and the other for only 6 months differently.

3.4.3 The Cox proportional hazards model

If data on additional variables were also collected for each individual enrolled in the study, the data would be represented by the triple $(t_j, \sigma_j, \mathbf{X}_j)$ j=1,2,…,n where the additional $\mathbf{X}_j=(x_{1j}, x_{2j}, \ldots, x_{pj})'$ is a vector of known explanatory variables for the j^{th} individual. When covariates are considered in a survival analysis there may be covariates which are time dependent, that is their value may vary during the study period and the ratio of the hazard rates for two or more groups may not be constant over time. However, in our case we have covariate information only at the time of the survey. Thus, it can be assumed that the ratio of the hazards for two or more groups to be a constant for all time points. This assumption is known as the assumption of proportional hazards. The log-minus-log (LML) plot, a useful plot for assessing whether the baseline hazard functions are proportional is used to test this assumption (Marubini and Valsecchi, 1995).

When the dependent variable is the cumulative survival function, i.e. the proportion of individuals 'surviving' at time t, the model is

$$S(t, \mathbf{X}_j)= [S_o(t)]^{\exp(\beta' x_j)} \qquad \text{for } j=1,2,\ldots,n$$

where $S_o(t)$ is the baseline survival function and β is a px1 vector of regression coefficients. The baseline survival function is similar to the constant term in multiple regression. It is the reference value that is increased or decreased depending on the values of the explanatory variables and their relationship with the dependent variable. Variables with positive coefficients are associated with decreased survival times while variables with negative coefficients are associated with increased survival times.

The Cox regression model defined above can also be written in terms of the hazard function. This model assumes that the hazard function for failure time t for an individual j with covariate vector $\mathbf{X}_j$ included in the study is

$$h(t, \mathbf{X}_j)=h_o(t)\exp(\beta' \mathbf{X}_j) \quad \text{for } j=1,2,\ldots,n$$

The arbitrary non-negative unspecified baseline hazard function, $h_o(t)$ is a function of time(t) only, not dependent on covariates and assumed to be the same for all subjects. The Cox model expressed in terms of the hazard function is called the ***Cox proportional hazards model***.

The cox regression model is not a fully parametric model since it does not specify the form of $h_o(t)$. It does, however, specify the hazard ratio for any two individuals with covariate vectors $\mathbf{X}_i$ and $\mathbf{X}_j$, $i \neq j$, and thus it is defined as a semi-parametric model. The function $\exp(\beta' \mathbf{X}_j)$ is a function of the values of the vector of explanatory variables for the j^{th} individual. This function can be interpreted as the hazard at time t for an individual whose vector of explanatory variables is $\mathbf{X}_j$ relative to the baseline hazard function (hazard function for an individual for whom the values of all the explanatory variables that make up the vector $\mathbf{X}_j$ are zero).

3.4.4 The parameter estimation and interpretation

Since the functional form of $h_0(t)$ is unspecified, it is not possible to use an ordinary likelihood to estimate the regression coefficients. Thus, the Cox's partial likelihood function which is discussed below is used in the analysis.

Consider a sample of n individuals with r distinct failures and the observed ordered failure times be $t_{(1)} < t_{(2)} < \ldots < t_{(r)}$, so that $t_{(j)}$ is the j^{th} ordered failure time. Let $R(t_{(j)})$ be the set of individuals who are at risk at time $t_{(j)}$ and $\mathbf{X}_j$ be a vector of covariates of the individual who fails at $t_{(j)}$. The probability that an individual with covariates $\mathbf{X}$ fails in a small interval (t,t + dt), given the set at risk at t is h(t, $\mathbf{X}$)dt. Thus, on the condition that one individual is observed to fail at $t_{(j)}$, the probability that it is an individual with covariates $\mathbf{X}_j$ is

$$\frac{h(t_{(j)}, x_j)dt}{\sum_{l \in R(t_{(j)})} h(t_{(j)}, x_l)dt}$$

The function describing the failure pattern is the product of r terms. Thus the relevant likelihood function for the proportional hazards model is given by

$$L(\beta) = \prod_{j=1}^{r} \frac{h(t_{(j)}, x_j)dt}{\sum_{l \in R(t_{(j)})} h(t_{(j)}, x_l)dt}$$

For the data consisted of n observed survival times let δ_j be a censoring indicator which is zero if the j^{th} survival time is right censored, and unity otherwise. The above likelihood function thus can be expressed in the form

$$L(\beta) = \prod_{j=1}^{n} \left[\frac{h(t_{(j)}, x_j)dt}{\sum_{l \varepsilon R(t_{(j)})} h(t_{(j)}, x_l)dt} \right]^{\delta_j}$$

$$= \prod_{j=1}^{n} \left[\frac{\exp(\beta' x_j)}{\sum_{l \varepsilon R(t_{(j)})} \exp(\beta' x_l)} \right]^{\delta_j}$$

And the corresponding Log-likelihood function is

$$LL(\beta) = \sum_{j=1}^{n} \delta_j \left\{ \beta' x_j - \log \sum_{l \varepsilon R(t_{(j)})} \exp(\beta' x_l) \right\}$$

The β parameters in the proportional hazards model can be obtained by maximizing the Log-likelihood function using the Newton-Raphson procedure which can be handled by the SPSS software. The likelihood function used is not a true likelihood since it does not make direct use of actual censored and uncensored survival times. Thus, the estimated β parameters are usually called the maximum partial likelihood estimates.

The proportional hazards model for survival data assumes that the hazard function is continuous, and under this assumption, tied survival times are not possible. However, the survival times are usually recorded to the nearest day, month or year (in this study to the nearest month) and so tied survival times can arise as a result of this rounding process. In order to accommodate the case of tied observations, the above partial likelihood function has thus to be modified. Details of the method of partial likelihood with tied observations are given in Kalbfleisch & Prentice (1980) and Collet (1994).

When individuals fall into one of m groups , $m \geq 2$ which correspond to categories of an explanatory variable, the groups being indexed by the levels of a factor (in our study there are only factors which take a limited set of values). Under the proportional hazards model, the hazard function for an individual in the k^{th} group of the factor is given by

$$h_k(t) = \exp(\gamma_k) h_o(t)$$

where γ_k is the effect due to the k^{th} level of the factor, and $h_o(t)$ is the baseline hazard function. Usually, we take $\gamma_1=0$ so that the baseline hazard function corresponds to the hazard of failure at time t for an individual in the first group. The ratio of the hazards at time t for an individual in the k^{th} group, $k \geq 2$ relative to an individual in the first group (reference group) is then $\exp(\gamma_k)$. Thus, the parameter γ_k is the logarithm of this relative hazard given as

$$\gamma_k = \log\left(\frac{h_k(t)}{h_o(t)}\right)$$

The coefficients of the explanatory variables in the model can thus be interpreted as logarithms of the ratio of the hazard of failure to the baseline hazard. A model which contains the terms γ_k, $k= 1,2,\ldots, m$ with $\gamma_1 = 0$ can be fitted by defining m-1 indicator variables $x_2, x_3, \ldots, x_m$. Fitting this model leads to estimators $\hat{\gamma}_1, \hat{\gamma}_3, \ldots, \hat{\gamma}_m$ with their corresponding standard errors. The estimated logarithm of the relative hazard for an individual in group k, relative to an individual in group 1 is $\hat{\gamma}_k$.

If the hazard ratio relative to the level of a factor other than the first (reference category) level is needed the hazard functions for individuals at level i and k of the factors are respectively $\exp(\gamma_i)h_o(t)$ and $\exp(\gamma_k)h_o(t)$, and so the hazard ratio for an individual at level i relative to the one at level k, is $\exp(\gamma_i-\gamma_k)$. The log hazard ratio is then $\gamma_i-\gamma_k$ which is estimated by $\hat{\gamma}_i-\hat{\gamma}_k$.

3.4.5 **Variable selection procedures**

To determine on which of the explanatory variables the hazard function depends, a number of different models will need to be fitted and the results compared. The situation where all explanatory variables are on an equal footing is considered with the aim to identify subsets of variables on which the hazard function is fitted and to build a model that excludes the variables that do not appear to be good predictors. Many possible combinations of terms paying due regard to the hierarchic principle (when a model contains an interaction term, the corresponding lower order terms should also be included) are considered. Comparisons among a number of possible models were made on the basis of the Akaikes Information Criterion (AIC) given as

$$AIC = -2 \text{ Log } L + \alpha q$$

where Log L is the logarithm of the maximized likelihood,

q is the number of unknown β parameters in the model and

α is a predetermined constant (usually taken to be between 2 and 6).

The smaller the value of the AIC, the better the model. Its value will tend to increase when unnecessary terms are added to the model. The choice α=3 is equivalent to using a 5% level of significance in judging the difference between the values of -2 Log L for two models. This is the value of α recommended for general use.

The automatic routines for variable selection have a number of disadvantages:

(i)They lead to the identification of one particular subset, rather than a set of equally good ones.

(ii)The subsets found by these routines often depend on the variable selection process that has been used, i.e. whether it is forward selection, backward elimination or the stepwise procedure.

(iii) Generally they do not tend to take any account of the hierarchic principle.

(iv)They depend on the stopping rule that is used to determine whether a term should be included in or excluded from a model.

Thus, the automatic routines have a limited role in model selection in survival analysis (Collet, 1994). Instead of using automatic variable selection procedures, the following general strategy for model selection was used:

1. First, models that contain each of the variables was fitted one at a time. To determine which variables on their own significantly reduced the values of -2 Log L, the values of this statistic for the models were then compared with that for the null model.

2. The variables which appeared to be important from step 1 were then fitted. In the presence of certain variables, others may cease to be important. Thus, those variables which did not significantly increase the value of -2 Log L when they were omitted from the model were discarded. The change in the value of -2 Log L when each variable on its own was omitted from the set was computed. Only those that led to a significant increase in the value of -2 Log L were retained in the model. Once a variable was dropped, the effect of each of the remaining variables in turn were also examined.
3. Variables which were not important on their own and were not considered in Step 2, but which became important in the presence of others (reduced the AIC significantly) were added to the model from Step 2 and are retained in the final model. Interactions which seemed important and satisfied the hierarchic principle were also examined.
4. Finally, it was checked to ensure that no term can be omitted without significantly increasing the value of the AIC, and, that no term not included significantly reduces the AIC.

3.4.6 **Model checking**

After a model has been fitted to an observed set of survival data, the adequacy of the fitted model should be investigated. Here in the assumption of a Cox regression model the adequacy of the fitted proportional hazards model has to be assessed. This can also be checked by examining the *martingale residuals.*

The Cox-Snell residual for the j^{th} individual, $j = 1,2,\ldots,n$ is given by

$$r_{cj} = \exp(\mathbf{b}'\mathbf{x}_j)H_o\hat{}(t_j)$$

where $H_o\hat{}(t_j) = -\log(S_o\hat{}(t_j))$ is the estimated cumulative baseline hazard function at time t_j, the observed survival time of that individual, and **b** is a vector of estimated parameters. The Cox-Snell residual is thus just the estimated cumulative hazard. The censored observations lead to residuals that cannot be treated the same as residuals from uncensored observations. Therefore, the Cox-snell residuals need modification to take an explicit account for censoring. Thus, the martingale residual (modified Cox-Snell residual) for the j^{th} individual is of the form

$r_{Mj} = -r_{cj}$ for censored observations

$r_{Mj} = 1 - r_{cj}$ for uncensored observations

The martingale residuals are similar to residuals encountered in linear regression analysis. In large samples they are uncorrelated with one another and have an expected value of zero. However, the martingale residuals are not symmetrically distributed about zero. Plots of the martingale residuals against the values of the explanatory variables or the linear predictor score ($X'\beta$), which is the sum of the products of the coefficients and the variables for a case, are useful in detecting any violations of the proportional hazards model. If most of the points in the plots fall in a horizontal band around zero, then the fitted model is taken as satisfactory. The plot of the martingale residuals against the linear predictor score would also be used for checking the presence of an outlier observation.

To identify influential observations that significantly affect the estimates of the coefficients, a statistic which estimates the change in the j^{th} coefficient with and without a case called Dfbeta, computed for each case is used. Cases with outlying values for Dfbeta are identified as influential observations and have to be examined. The graphical plot of Dfbeta against the case identification is used for this purpose.

CHAPTER IV

RESULTS AND DISCUSSION

This chapter presents results of the analysis of the data. On the basis of the objectives of the study, the data analysis among others comprises estimation of the prevalence and duration of breast-feeding of the last child in urban Addis Ababa. Undertaking comparisons between different levels of a factor and examination of the socio-demographic variables which are expected to jointly affect the variation in the duration of breast-feeding of the last child are also discussed in this chapter.

After editing the data- univariate, bivariate and multivariate statistical techniques were employed. Using the Kaplan-Meier estimate, univariate analysis on the duration of breast-feeding was made. Bivariate analysis based on the log-rank statistic served as a preliminary investigation of each variable's relationship to the duration of breast-feeding of the last child before the appropriate model was fitted. As the breast-feeding duration of a certain proportion (31.78%) of mothers was censored, for the simultaneous examination of the factors the Cox regression model was employed.

4.1 **Results of the univariate analysis**

Prevalence of breast-feeding of the last child may be defined as the proportion of mothers who practised breast-feeding their last child irrespective of the duration. In our study it was found to be 95%. Using the Kaplan-Meier technique, the probability of weaning for mothers who breast-fed their last child at various duration was estimated. Estimates of cumulative survival probability based on this method were 0.7643 (s.e.=0.0141), 0.5058 (s.e.=0.0178), 0.1542 (s.e.=0.0151) and 0.0218 (s.e.=0.0081) for months 6, 12, 24 and 36, respectively. The corresponding estimates for the months 1 up to 36 are presented in table 2.

Table 2. Kaplan-Meier estimates of the duration of breast-feeding of the last child in urban Addis Ababa

Time	**Cumulative Survival**	*Standard error*	Time	**Cumulative Survival**	*Standard error*
1	**0.9505**	*0.0063*	17	**0.4426**	*0.0181*
2	**0.9268**	*0.0084*	18	**0.3952**	*0.0182*
3	**0.8912**	*0.0101*	19	**0.3869**	*0.0182*
4	**0.8473**	*0.0118*	20	**0.3852**	*0.0182*
5	**0.8259**	*0.0125*	21	**0.3834**	*0.0182*
6	**0.7643**	*0.0141*	22	**0.3798**	*0.0182*
7	**0.7413**	*0.0146*	24	**0.1542**	*0.0151*
8	**0.7103**	*0.0153*	25	**0.1523**	*0.0150*
9	**0.6769**	*0.0159*	26	**0.1345**	*0.0144*
10	**0.6531**	*0.0163*	27	**0.1255**	*0.0141*
11	**0.6274**	*0.0167*	28	**0.1122**	*0.0138*

12	**0.5058**	*0.0178*	29	**0.1012**	*0.0135*
13	**0.5015**	*0.0179*	30	**0.0399**	*0.0101*
14	**0.4676**	*0.0180*	31	**0.0363**	*0.0098*
15	**0.4585**	*0.0180*	32	**0.0290**	*0.0091*
16	**0.4554**	*0.0180*	36	**0.0218**	*0.0081*

The median was found to be 14 months (s.e.=0.58) showing that 50% of the mothers were breast-feeding their most recent child until at most the child was 14 months old. While the remaining 50% breast-fed their last child for more than 14 months at the time of the survey.

A certain proportion of continued breast-feeding is evident from the pattern of the cumulative survival function. About 2.18% of the mothers continued to breast-feed their last child up to 36 months. Thus, from the survival analysis of the duration of breast-feeding of the last child it can be stated that weaning is a slow process for some of the mothers.

The first quartile, $Q_1 = 7$ months (s.e. = 0.39), locates the month at which 75% of the mothers continued to breast-feed their most recent child was at the age of 7 months. While the upper quartile, $Q_3 = 24$ months (s.e.=0.13), shows the month at which 25% of the mothers continued to breast-feed their last child was at the age of 24 months. The semi-interquartile range(SIQR) was found to be 8.5 months .Thus, it seems that there is a wide variation in the duration of breast-feeding in the study area.

4.2 **Bivariate findings**

For each level of the factor in the study survival functions were computed. To examine whether a given variable has an effect on the duration of breast-feeding a log-rank test was undertaken. The null hypotheses that the survival functions are the same for the different groups of the predictor variables were tested using the log-rank statistic and the results of the test are given in Table 3.

Table3. Test statistics for testing equality of survival functions

Variable	Log-rank statistic	df	significance
con_res	1.64	1	0.2002
motbirpl	44.49	1	0.0000*
educatio	14.03	2	0.0000*
ethnic	5.54	3	0.1361
religion	1.8	2	0.4059
marit	0.19	1	0.6640
ptbreast	1.3	1	0.2537
contrace	19.88	1	0.0000*
working	3.77	1	0.0500*
hus-edu	84.57	2	0.0009*
agecat	23.43	2	0.0000*
parcat	14.74	2	0.0006*

* *Significant at 0.05 level.*

The log-rank test showed the null hypothesis that the survival functions are equal among the groups was rejected at the 0.05 level for the variables : mother's birth place, educational level of the mother, usage of contraception, work status of the mother, educational level of the husband, age of the mother and birth order of the last child. In other words, for these variables the assumption that the hazard of weaning at any given time for a mother in one group is proportional to the hazard at that time for a mother in the other group holds. But, the remaining variables did not show a significant association with the duration of breast-feeding of the last child. From this result, it can be stated that the duration of breast-feeding of the last child showed a statistically significant difference between working and non-working mothers. Our earlier hypothesis concerning working status of mothers, i.e. there is a statistically significant difference on the duration of breast-feeding of the last child between working and non-working mothers cannot thus be rejected.

Pair-wise log-rank tests were undertaken for the variables with more than two categories to determine which of the groups were significantly different from each other. All comparisons for ethnicity and religion were not significant. All comparisons for educational level of the mother and birth order of the last child were significant. Except the comparisons between literacy program / non-formal versus primary / junior high school and between age group 15-24 versus age group 25-34, the remaining comparisons for educational level of the husband and age of the mother were significant.

When un-adjusted (singularly considered) Cox regression model was fitted for each variable, the variables found to be significant were only those which were found to be so using the log-rank test. The significant variables obtained using the two methods were one and the same. This is shown in Table 4.

Table 4. Un-adjusted Cox regression model

Variable	Exb(β)	Significance	-2 Log Likelihood	AIC
agecat (15-24)		0.0001*	8128.378	8134.378
25-34	0.9573	0.6562		
35-49	0.6702	0.0002		
con_res(no)	1.1129	0.2471	8146.448	8149.448
contrace (never used)	1.3943	0.0001*	8131.352	8134.352
hus_edu(liter./non for)		0.0035*	7347.811	7353.811
primary/juniorhigh sh.	1.0687	0.5345		
Seniorhigh sch./higher	1.3633	0.0022		
ethnic(Amhara)		0.2123	8144.314	8153.314
Oromo	1.1802	0.0919		
Gurage	0.9400	0.6040		
others	1.1353	0.2867		
educatio(liter./non for)		0.0000*	7089.784	7095.784
primary/juniorhigh sch.	1.5106	0.0002		

Variable	Exb(β)	Significance	-2 Log Likelihood	AIC
Senior high sch./higher	2.3004	0.0000		
marit(widowed/divor.)	1.0335	0.6978	7626.058	7629.058
motbirpl(urban)	0.5439	0.0000*	5950.267	5953.267
parcat(1st)		0.0026*	8135.85	8141.85
2nd - 5th	0.8212	0.0509		
6th or more	0.6718	0.0006		
ptbreast(no)	0.8120	0.3011	6186.048	6189.048
religion (Orthodox)		0.4761	8146.327	8152.327
Muslim	1.1357	0.3808		
others	1.1645	0.3636		
Working (no)	1.1467	0.0509*	8144.704	8147.704

* *Significant at 0.05 level of significance.*

Levels in parentheses are the reference categories.

The relative risk (hazard ratio) for mothers with the age groups 25-34 and 35-49 were 0.96 and 0.67, respectively when compared with the age group 15-24. This showed that the duration of breast-feeding of the last child increased as the age of the mother advanced. But the relative risk for mothers with primary / junior high school level of education was found to be 1.51, while for mothers with senior high school / other higher education was found to be 2.30 when compared with mothers whose level of education was literacy program / non-formal. From this it can be stated that the duration of breast-feeding of the last child decreased as the level of education of the mother increased. The hazard ratio of weaning for working mothers was found to be 1.15 times that of the non-working. This implies that the working mothers breast-fed their last child for a shorter duration when compared with mothers who were non-working. From the above results it can be concluded that our earlier three hypotheses were not found to be rejected at the 0.05 level of significance.

It was also found that the relative risk of weaning for mothers who used contraceptive was higher than those who never used. As educational level of the husband increased the hazard ratio also increased. Mothers who were born in rural areas had a smaller relative risk of weaning when compared with those who were born in urban areas. As the order of the last birth (parity) increased the hazard ratio of weaning was found to decrease. Mothers who breast-fed their previous child had a smaller relative risk of weaning than those who did not do so.

4.3 Multivariate Findings

In the above bivariate analysis, the associations observed for each factor do not take account of any other factors. Without a multivariate analysis it is difficult to know which factor is more important. Thus, in order to explore the adjusted effect of the socio-demographic variables on the duration of breast-feeding of the last child in urban Addis Ababa a multivariate analysis was also undertaken. An approach based on statistical modelling without prioritizing the variables, the 'Hierarchic' method as discussed in Section 3.4.5 was used. When a proportional hazards model which contained no explanatory variables, that is, the null model was fitted the value of -2 Log L (AIC) was found to be 8147.762. In this model the hazard of weaning did not depend on any of the variables and it was the same for all mothers in the study.

Models which contained each of the 12 explanatory variables on their own were also fitted. Their -2 Log L and AIC values were given in Table 4. From these variables agecat, contrace, educatio, hus_edu, marit, motbirpl, parcat and ptbreast led to the largest reduction in -2 Log L when each of these variables was added to the null model one at a time. All of these reductions were found to be significant at the 5% level when compared with percentage point of the chi-squared distribution with 1 degree of freedom. This suggests that not all of the explanatory variables were required to explain the duration of breast-feeding.

The next step was to fit the model that contained all of the above eight variables together which led to a value of -2 Log L of 3528.895. The effect of omitting each of the variables in turn from this model and also the values of -2 Log L and AIC for other models also is given in Table 5. When parcat was omitted, the increment in -2 Log L was by only 0.246. But when each of the other variables was omitted the value of -2 Log L increased tremendously. Each of the remaining changes in the value of -2 Log L was significant when compared with percentage points of a chi-squared distribution with 1 degree of freedom. Since parcat did not appear to be needed in the model in the presence of the remaining seven variables it was therefore excluded.

If each variable was omitted from the model which contained the 7 variables agecat, contrace, educatio, hus_edu, marit, motbirpl and ptbreast (model 3) the increment in the value of -2 Log L was not significant when the variable contrace was excluded. Thus the model which was further studied was the model which contained the variables agecat, educatio, hus_edu, marit, motbirpl and ptbreast (model 15). When each of the remaining variables in model 15 was omitted the variable agecat had insignificant increment. Thus, the model which was comprised of the five variables educatio, hus_edu, marit, motbirpl and ptbreast (model 21) was the selected model. When each of the variables in model 21 was omitted the increment in the value of -2 Log L was found to be significant. Any variable in model 21 cannot be excluded from the model without significantly increasing the value of the -2 Log L statistic of this model.

Finally, we wanted to see if any of the variables con_res, ethnic, religion and working which were not significant on their own whether they were important in the presence of the variables selected so far. The value of -2 Log L decreased significantly when the variable working was added to the model which contained the variables educatio, hus_edu, marit, motbirpl and ptbreast. Thus our tentative model was model 31. When the AIC was calculated for each of the previously fitted models, this model was found to be the one with the smallest AIC value.

Table 5. Values of -2 Log L and AIC for some of the fitted models

Number	Model	-2 Log L	AIC
1	agecat + contrace + educatio + hus_edu + marit + motbirpl + parcat + ptbreast	3528.895	3564.895

Number	Model	-2 Log L	AIC
2	agecat + contrace + educatio + hus_edu + marit + motbirpl + parcat	4427.024	4460.024
3	agecat + contrace + educatio + hus_edu + marit + motbirpl + ptbreast	3529.141	3559.141
4	agecat + contrace + educatio + hus_edu + marit + parcat + ptbreast	4897.821	4930.821
5	agecat + contrace + educatio + hus_edu + motbirpl + parcat + ptbreast	3560.698	3593.698
6	agecat + contrace + educatio + marit + motbirpl + parcat + ptbreast	4034.730	4064.730
7	agecat + contrace + hus_edu + marit + motbirpl + parcat + ptbreast	3816.188	3846.188
8	agecat + educatio + hus_edu + marit + motbirpl + parcat + ptbreast	3535.570	3568.570
9	contrace + educatio + hus_edu + marit + motbirpl + parcat + ptbreast	3535.722	3565.722
10	agecat + contrace + educatio + hus_edu + marit + motbirpl	4427.185	4454.185
11	agecat + contrace + educatio + hus_edu + marit + ptbreast	4897.845	4924.845
12	agecat + contrace + educatio + hus_edu +	3560.962	3587.962

Number	Model	-2 Log L	AIC
	motbirpl + ptbreast		
13	agecat + contrace + educatio + marit + motbirpl + ptbreast	4037.232	4061.232
14	agecat + contrace + hus_edu + marit + motbirpl + ptbreast	3816.718	3840.718
15	agecat + educatio + hus_edu + marit + motbirpl + ptbreast	3530.729	3557.729
16	contrace + educatio + hus_edu + marit + motbirpl + ptbreast	3536.493	3560.493
17	agecat + educatio + hus_edu + marit + motbirpl	4428.506	4452.506
18	agecat + educatio + hus_edu + marit + ptbreast	4899.437	4923.437
19	agecat + educatio + hus_edu + motbirpl + ptbreast	3563.210	3587.21
20	agecat + educatio + marit + motbirpl + ptbreast	4042.308	4063.308
21	educatio + hus_edu + marit + motbirpl + ptbreast	3533.627	3554.627
22	agecat + hus_edu + marit + motbirpl + ptbreast	3818.098	3839.098

Number	Model	-2 Log L	AIC
23	educatio + hus_edu + marit + motbirpl	4434.062	4452.062
24	educatio + hus_edu + marit + ptbreast	4903.643	4921.643
25	educatio + hus_edu + motbirpl + ptbreast	3566.058	3584.058
26	hus_edu + marit + motbirpl + ptbreast	3820.833	3835.833
27	educatio + marit + motbirpl + ptbreast	4049.927	4064.927
28	educatio + hus_edu + marit + motbirpl + ptbreast + con_res	3533.627	3557.627
29	educatio + hus_edu + marit + motbirpl + ptbreast + ethnic	3529.825	3559.825
30	educatio + hus_edu + marit + motbirpl + ptbraest + religion	3532.482	3559.482
31	educatio + hus_edu + marit + motbirpl + ptbreast + working	3528.766	3552.766
32	educatio + hus_edu + marit + motbirpl + ptbreast + working + educatio x hus_edu	3520.263	3556.263
33	educatio + hus_edu + marit + motbirpl + ptbreast + working + educatio x working	3532.301	3562.301
34	educatio + hus_edu + marit + motbirpl + ptbreast + working + educatio x ptbreast	3525.425	3555.425
35	educatio + hus_edu + marit + motbirpl + ptbreast + working + educatio x marit	3531.348	3561.348

Number	Model	-2 Log L	AIC
36	educatio + hus_edu + marit + motbirpl + ptbreast + working + educatio x motbirpl	3528.400	3558.4
37	educatio + motbirpl + educatio xmotbirpl	5043.607	5048.607

The relevant interaction terms that have a large importance in predicting the duration of breast-feeding should also be included in our model. Applying the hierarchic principle the interaction effects of the combination of some of the variables were thus considered. The interaction effect of the educational level of the mother with husband's educational level, her working status, her marital status, her place of birth and breast-feeding condition of the previous child were not seen easily. When the interaction effect of these variables were incorporated to the model which contained the variables educatio, hus_edu, marit, motbirpl, ptbreast and working, the changes in the values of -2 Log L were not significant when they were compared with a chi-square distribution with corresponding degrees of freedom. The value of the AIC for each fitted model with an interaction (models 32-36) did not reduce from that of model 31. Thus the model (model 31), without the interaction terms having the smallest AIC was preferred.

When our final model that contained the variables educatio, hus_edu, marit, motbirpl, ptbreast and working was fitted the obtained results are given in Table 6 below.

Table 6. Summary of the parameter estimates of the final model

Variable	B	S.E.	Wald	df	Sig	Exp(B)
PTBREAST	-.4271	.2728	2.4514	1	.1174	.6524
MOTBIRPL	-.3423	.1401	5.9659	1	.0146	.7102
HUS_EDU			2.0657	2	.3560	

HUS_EDU (1)	-.0274	.1510	.0329	1	.8561	.9730
HUS_EDUC(2)	.1600	.1430	1.2518	1	.2632	1.1735
EDUCATIO			29.7401	2	.0000	
EDUCATIO(1)	.3967	.1431	7.6882	1	.0056	1.4868
EDUCATIO(2)	.8463	.1560	29.4505	1	.0000	2.3311
MARIT	.0188	.1263	.0223	1	.0814	1.0190
WORKING	.1221	.1143	1.1415	1	.2853	1.1298

In the above fitted model the difference between educated and less educated mothers in urban Addis Ababa is among the largest observed. Husband's educational level, mother's work status, marital status and breast-feeding condition of previous child did not appear to be as strongly associated with duration of breast-feeding as that of the educational level of the mother. We found that educational level of the mother and mother's birth place were the two variables which significantly affected the duration of breast-feeding in a decreasing order. When the model which contained only these two variables and their interaction was fitted (model 37) the value of -2 Log L increased significantly. This showed that the rest of the variables were necessary in the presence of the variables level of education of the mother and birth place of the mother.

From the above table it was observed that a mother with an educational level of senior high school / other higher education had 2.33 times higher risk of weaning as compared to a mother with an educational level of literacy program / non-formal adjusted for other variables in the model. Similarly, a mother with an educational level of primary / junior high school had 1.49 times higher risk of weaning when compared with mothers who had the lowest level of education. It was also observed that the hazard of weaning increased for husbands with senior high school / other higher education while it was almost the same for a husband with primary / junior high school level of education and literacy program / non-formal.

Mothers who were born in rural areas breast-fed for a longer duration and had 0.71 times lower risk of weaning as compared to mothers who were born in urban areas. The relative risk for the working mothers was 1.13. This implied that the estimated risk of weaning is 1.13 times greater for a mother who was working as compared to a non-working mother. The estimated risk of weaning for a mother who was married / living with a man was 1.02 times greater than for a mother who was widowed / divorced. And as it was expected, the hazard of weaning decreased for mothers who breast-fed their previous child.

4.4 Results of the model adequacy measures

The likelihood-ratio test for the hypothesis that all parameters in the final proportional hazards model are zero is obtained by comparing -2 Log L for this model with that of a model in which all the coefficients are 0. The difference in -2 Log L between the initial and final proportional hazards models was significant at the 5% level of significance. The test of this null hypothesis can also be based on the score statistic (overall chi-square). This statistic for the final proportional hazards model was also significant at the 0.05 level. Hence we rejected the null hypothesis that all the parameters in the model are zero. Thus, there is no statistical evidence to conclude that this model did not show the best combination of the variables.

In order for the assumption of proportional hazards for two or more groups to be fulfilled the cumulative survival curves and the hazard curves for groups should not cross each other. The survival curves for the factors in the final model are thus given in Figures 3-8 in the annexes. The graphical representation of the overall survival function and the hazard function for this model are also depicted in Figures 1 and 2.

For checking the assumption of the proportional hazards, the log minus log plot for each covariate in the final model is shown in Figures 9-14. The lines for individual categories, for each covariate, are almost parallel. This implies that the assumption of proportional hazards is justifiable for the variables included in the model.

Plots of the martingale residuals against the values of the explanatory variables and the plot of the martingale based residuals versus the linear predictor score (X'β) are also given in Figures 15-21. The fitted model is thus satisfactory as most of the points in the plots fell in a horizontal bound around zero. For the economy of space the plots of Dfbetas versus case numbers are not shown.

4.5 **Discussion**

The descriptive results showed that only 5% of the mothers who gave birth reported not to have breast-fed their last child. This indicates that breast-feeding is a universal practice (a norm) in the metropolitan. According to the study made by Abdulahi (1990) the average duration of breast-feeding of the last child in urban Addis Ababa was found to be 18 months. In 1990 it was found to be 16.4 months (CSA, 1993). The study made by Yeshewamebrat (1995) observed a duration of 15.2 months. This study revealed that the median duration of breast-feeding of the last child was 14 months. As time passed, slight variation in the average duration of breast-feeding is observed.

The log-rank test was employed as the primary assessment to identify factors significantly associated with the duration of breast-feeding. Results of this test showed that the variables permanent residence of the mother in urban Addis Ababa, ethnicity, religion, marital status of the mother and condition of breast-feeding of the previous child did not have a significant association with the duration of breast-feeding. Without the presence of other variables marital status and condition of breast-feeding of the previous child were thus not found to be associated with the duration of breast-feeding.

The un-adjusted Cox regression revealed that mothers: born in rural areas, with low level of education, never used a contraceptive, older, higher parity, not working and with low levels of their husbands' education were found to have a longer duration of breast-feeding when compared with their respective counterparts. The level of education and the birth place of the mother were the two most significant variables ($p<0.0001$) which affect the duration of breast-feeding of the last child.

The final Cox regression model indicated that numerous factors affect the duration of breast-feeding. Among the widely studied variables that determine the duration of breast-feeding is the educational level of the mother. In our study level of education of the mother had the greatest explanatory power of the six variables fitted in the model. As many other studies in developing countries,it was found that the educational level of a mother is negatively associated with the duration of breast-feeding. The same pattern was obtained in the bivariate analysis. An inverse relationship was also found in the 1990 National Family and Fertility Survey. The negative relationship between the educational level of mothers and duration of breast-feeding may be due to the fact that educated mothers have more exposure to the mass media and also have access to read written materials which advertise the superiority of bottle-feeding on breast-feeding. Thus, educated mothers are supposed to be the lead group, the first to replace the traditional breast-feeding behaviour with bottle-feeding by considering bottle-feeding as a modern behaviour.

Urban residence is one of the most important variables that determine the duration of breast-feeding. Several studies documented that urban mothers breast-feed for shorter duration than their rural counterparts. As this study was restricted only to an urban population it did not indicate how the residence, urban versus rural, would affect the duration. Thus, the mother's urban-rural birth place difference was observed instead to explore whether mothers of rural origin currently living in the metropolitan maintained breast-feeding practices. The study revealed that migrants to the metropolitan from rural areas breast-fed longer than urban migrants and native urban mothers. Immigration to urban areas of rural mothers who maintain their traditional breast-feeding practices may increase the duration of breast-feeding in urban areas.

It seemed likely that part of the effects of social background on last parity breast-feeding was mediated by previous birth feeding practices. Moreover, the actual experience of breast-feeding including mothers reaction is likely to play a critical role in her decision whether to continue or change breast-feeding practices. A positive association was observed between breast-feeding of previous child and last child. It was thus observed that if a mother breast-fed her previous child, she is very likely to continue with her next child while it is very rare for a mother who did not breast-feed her previous child to try with the last. Hirschman and Butler (1981), in their analysis on the probability and duration of breast-feeding of second births, found that the effect of the length of breast-feeding of first birth was the single most powerful variable. Most of the mothers who tried breast-feeding for a very short period with their first birth did not continue with their second birth. It was also found that the longer breast-feeding duration of the first birth, the more likely is continuation of breast-feeding of subsequent births. A positive experience with breast-feeding the first child might be the major reason for deciding to continue with subsequent births. Thus, it may be stated that an unsuccessful experience of breast-feeding the first child promotes adoption of bottle-feeding for subsequent births.

Mothers who worked tend to breast-feed for a shorter period than mothers who did not work. The work status of mothers perhaps makes a major difference in the duration of breast-feeding as it demands leaving the infant at home during working hours. As breast-feeding requires time and competes with full time employment, working mothers tend to breast-feed for shorter duration.

Marital status indicates availability of spouse to provide economic and social support for the mother. The duration of breast-feeding for married mothers / mothers living with a man was observed to be slightly shorter. A stable relationship with the child's father may provide more financial security for the mother to buy breast-milk substitutes.

There appeared to be a negative association between duration of breast-feeding and level of husband's education. While husband's level of education is not, strictly speaking, the same as income, it is highly correlated and probably represents it. It is a proxy for socio-economic status. As long as the child is on the breast milk, it receives assured nourishment independently of father's socio-economic status as reflected by father's level of education. When the child is weaned, however, its health then also depends on nourishment apart from breast milk and the availability and quality of such nourishment is influenced by the socio-economic status of the family as reflected by father's level of education. Thus mothers with a lower economic status breast-fed for a longer duration.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

The objective of using the proportional hazards model in the prsent study was to determine whether a relationship exists between the mother's socio-demographic characteristics and subsequent duration of breast-feeding of the last child and if so, to estimate its size and direction. The study offers important insights into the relative importance of some of the socio-demographic factors that affect the duration of breast-feeding. These insights should prove valuable in developing specific breast-feeding promotion efforts needed to enhance the health and well-being of infants and mothers.

5.1 **Conclusions**

Education was found to be highly significant, with more educated mothers breast-feeding for shorter periods than the less educated ones. Based on the experience of developing countries increasing levels of urbanization and modernization are among the important contributors to the decline in the duration of breast-feeding, their impact being primarily upon educated mothers. Educated and working mothers are the forefront for the declining of breast-feeding duration. The heavy commercial promotion of formula feeding probably also contributed to the discouragement of breast-feeding, the more educated mothers being the most likely to be affected first. Thus, in order breast-feeding promotion campaigns to be fruitful they should be targeted at these mothers.

Highly educated mothers are more likely to read progressive infant care literature and are likely to be contacted by voluntary organizations that encourage breast-feeding. Since it is this group of mothers whose patterns is followed by the less educated ones, campaigns directed toward the better educated is required more. Such breast-feeding campaigns will be effective, when also aimed at health and family planning practitioners. Public awareness on the importance of breast-feeding should also be created in the metropolitan through the mass media.

Two large groups with special health risks are, infants who are weaned very early and children weaned late who are often provided with nutritionally inadequate supplements. Concerned health institutions, governmental and non-governmental organizations, and health workers should encourage the promotion and protection of breast-feeding even the second year of life (the recommended breast-feeding duration by UNICEF / WHO). A duration of breast-feeding more than two years is not significant, perhaps because beyond this age a child's health status is determined more by the environmental conditions under which it plays, is fed and lives. Hence, breast-feeding is less important beyond this age. Thus, efforts should include the development of positive societal attitudes towards two years breast-feeding duration.

After birth, advice from health professionals can influence feeding situations strongly. Mothers should receive the necessary support, advice and encouragement to initiate and continue breast-feeding. Appropriate polices and well-designed programmes have positive effects on breast-feeding behaviour. Particularly, younger mothers who gave their first birth should be treated with a special attention. Successful experience of breast-feeding the first child promotes its adoption for subsequent births. We therefore suggest that the advice and support of the medical profession is critical in the initiation and continuation of breast-feeding. Successful lactation should be initiated with the necessary social support and encouragement for the primiparas. Breast-feeding education should also be established to counteract early supplementation.

5.2 **Recommendations**

The promotion of breast-feeding require an integrated approach including favourable public policies, attention to the health care providers and the mothers themselves and monitoring the results. Explicit public policies that discourage bottle-feeding and encourage breast-feeding need to be formulated. The study recommends the re-education of health personnel, especially those in the maternal and child health clinics (MCH) to educate and train the mothers on the importance of breast-feeding their children. Breast-feeding education programme has to be conducted with the purpose of promoting more frequent and intensive breast-feeding especially during the first six months, to postpone weaning and to confront local myths that are contrary to sound breast-feeding practices.

Thus the following specific recommendations are essential:

- Official encouragement of breast-feeding by medical organizations
- Provision of advise and support to breast-feeding beginners

- Vigorous campaign on the higher educated mothers to promote longer breast-feeding duration
- Build a sustained breast-feeding promotion programme through the mass media
- Allowing working mothers to have a maternity leave for at least 4 months
- Display breast-feeding posters on MCH institutions.

REFERENCES

Abdulahi Hasen (1990). Breast-feeding Practices and Fertility in Mettu, Alemaya and Addis Ababa. *Population and Development Bulletin*, Vol. 1, No. 2, 1-8.

Adiar, L.S., Popkin, B.M. and Guilkey, D.K. (1993). The Duration of Breast-feeding: How Is It Affected by Biological, Socio-demographic, Health Sector and Food Industry Factors?. *Demography,* Vol. 30, No.1, 63-80.

Akin, J.S., Bilsborrow, R.E., Guilkey, D.K. and Popkin, B.M. (1986). Breast-feeding Patterns and Determinants in the Near East: An Analysis for Four Countries. *Population Studies,* Vol. 40, No.2, 247-262.

Akin, J., Bilsborrow, R., Guilkey, D., Popkin, B.M., Benoit, D., Cantrelle, P., Garenne, M. and Levi, P. (1981). The Determinants of Breast-feeding in Sri Lanka. *Demography,* Vol.18, No.3, 287-307.

Allison, P.D. (1984). *Event history analysis*. Sage Publications, Newbury Park.

Bracher, M. and Santow, G. (1982). Breast-feeding in Central Java. *Population Studies,* Vol. 36, No.3, 413-429.

Central Statistical Authority (1993). The 1990 National Family and Fertility Survey Report, Central Statistical Authority, Addis Ababa.

Collet, D. (1994). *Modelling Survival Data in Medical Research.* Chapman and Hall, London.

Diamond, I.D., Mcdonald, J.W.and Shah, I.H. (1986). Proportional Hazards Models for Current Status Data: Application to The Study of Differentials in Age at Weaning in Pakistan. *Demography,* Vol.23, No.4, 607-620.

Eiger, M.S. and Olds, S.W. (1987). *The Complete Book of Breast-feeding.* Workman Publishing, New York.

Ferry, B. (1981). Breast-feeding. *WFS Comparative Studies,* No.13, 1-42.

Goldberg, H.I., Rodrigues, W., Thome, A.M.T., Janowitz, B. and Morris, L. (1984). Infant Mortality and Breast-feeding in North-Eastern Brazil. *Population Studies* Vol.38, No.1, 105-115.

Groff, P. (1986). *Maternal and Child Health for Health Assistants in Ethiopia.* Ministry of Health, Addis Ababa.

Hinkley, D.V., Ried, N. and Snell, E.J. (1991). *Statistical Theory and Model Building.* Chapman and Hall, London.

Hirschman, C. and Butler, M. (1981). Trends and Differentials in Breast-feeding: An Update. *Demography,* Vol.18, No.1, 39-54.

Huffman, S.L. and Lamphere, B.B. (1984). Breast-feeding Performance and Child Survival. in *Population and Development Review,* Vol.10, 93-116.

Jain, A.K., Hsu, T.C., Freedman, R. and Chang, M.C. (1970). Demographic Aspects of Lactation and Postpartum Amenorrhea. *Demography,* Vol.7, No.2, 255-271.

Kalbfleisch, J.D. and Prentice, R.L.(1980). *The Statistical Analysis of Failure Time Data.* John Wiley and Sons, New York.

Lee, E.T. (1980). *Statistical Methods for Survival Data Analysis.* Lifetime Learning Publications, Belmont.

Lehtonen, R. and Pahkinen, E.J. (1995). *Practical Methods for Design and Analysis of Complex Surveys.* John Wiley and Sons, Chichester, England.

Marubini, E. and Valsecchi, M.G. (1995). *Analysing Survival Data from Clinical Trials and Observational Studies.* John Wiley and Sons, Chichester.

Retherford, R.D., Choe, M.K., Thapa, S. and Gubhaju, B.B.(1989). To what Extent Does Breast-feeding Explain Birth-Interval Effects on Early Childhood Mortality. *Demography,*Vol.26, No.3, 439-450.

Shah, I.H. and Khanna, J. (1990). Breast-feeding, Infant Health and Child Survival in the Asia-Pacific Context. In *Asia-pacific Population Journal,*Vol.5,No.1, 25-44.

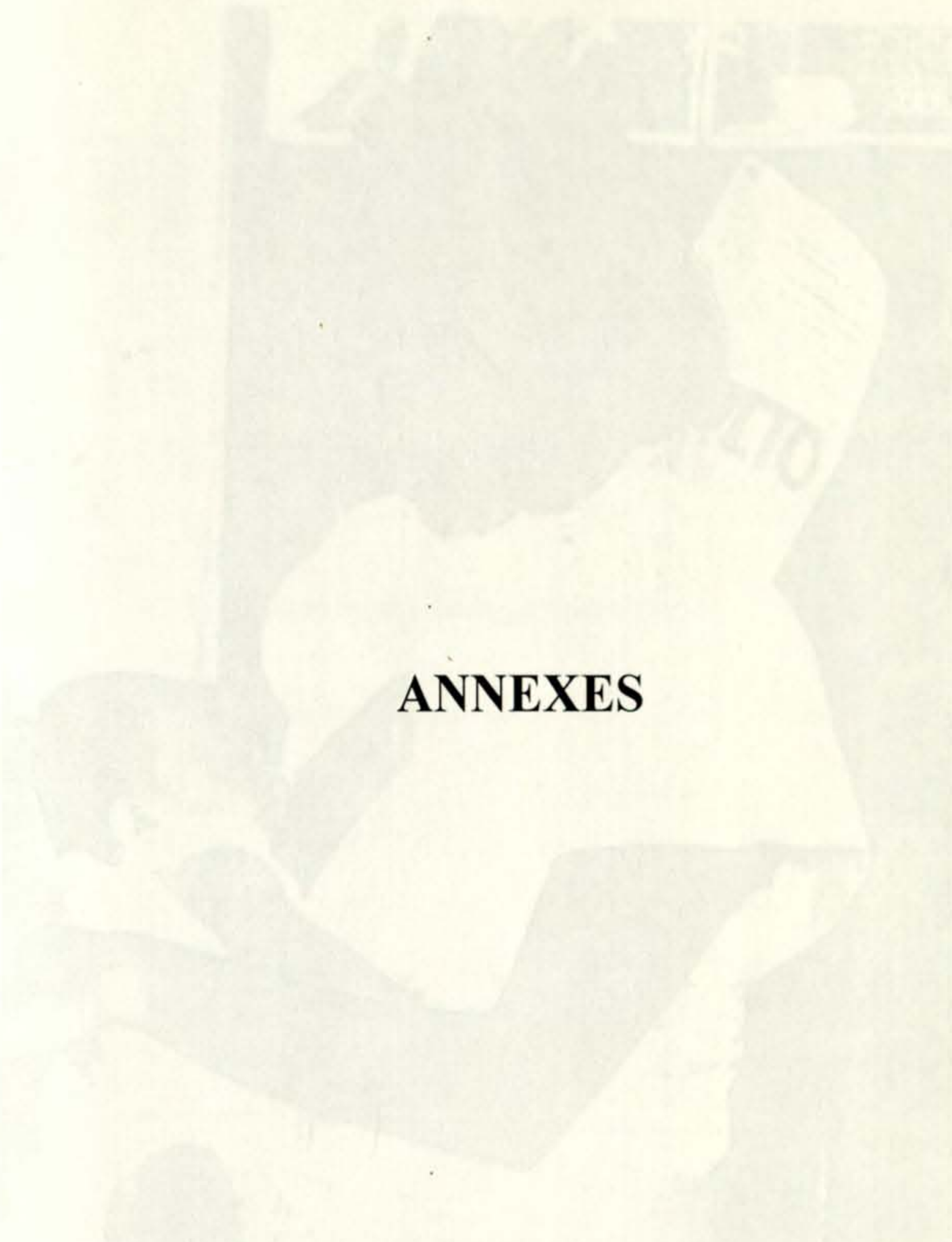
Smith, D.P.and Ferry, B. (1984). Correlates of Breast-feeding. *Comparative Studies*, No.41, 1-23.

The Transitional Government of Ethiopia (TGE) (1993). National Population Policy of Ethiopia. Office of the Prime Minister, Addis Ababa.

Trussell, J., Strawn, L.G., Rodriguez, G. and Vanlandingham, M. (1992). Trends and Differentials in Breast-feeding Behaviour: Evidence from the WFS and DHS. *Population Studies,* Vol. 46, No.2, 285-307.

Yeshewamebrat Ejigsemahu (1995). Breast-feeding: Determinants and Birth Interval Effects in Addis Ababa- A Case of Zone One. Unpublished M.Sc. thesis, Addis Ababa University

ANNEXES



One of life's precious moments - the time a mother spends breastfeeding her baby.

One of life's precious moments – the time a mother spends breast-feeding her baby.

Fig.1 The overall Survival Function

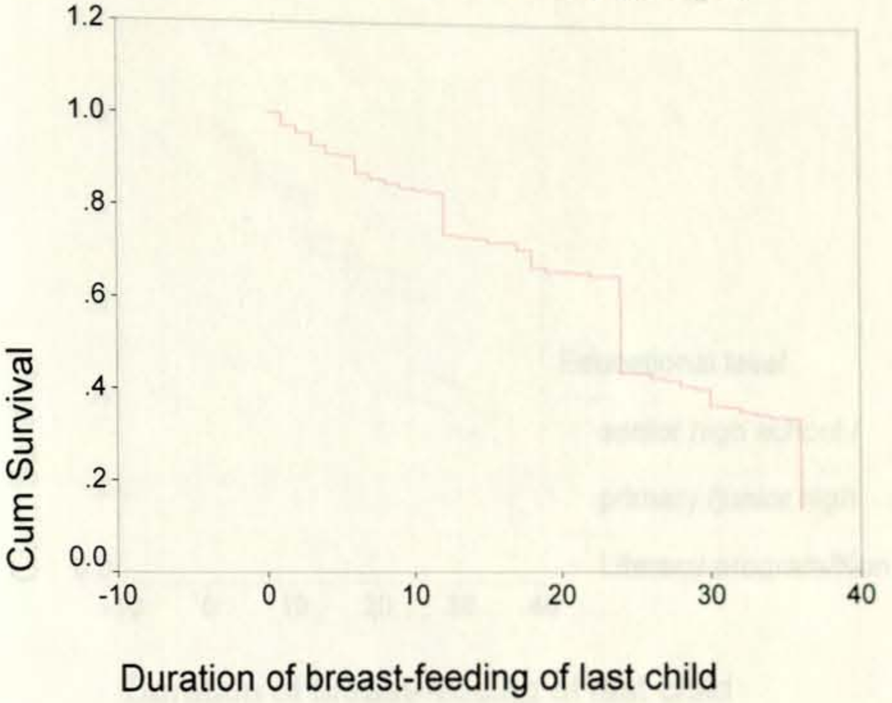


Fig.2 The Hazard Function

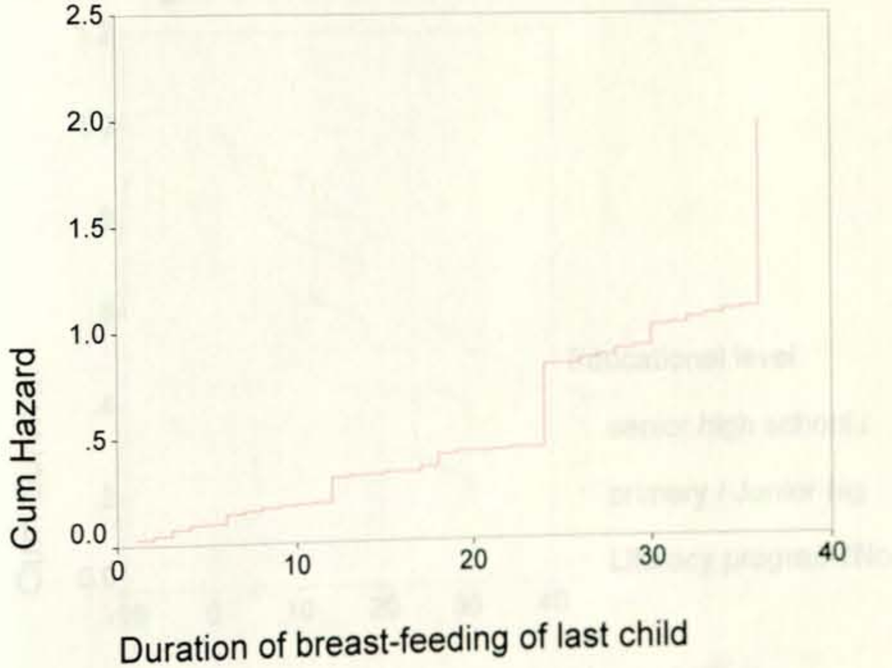


Fig.3 Survival Function for hus_edu

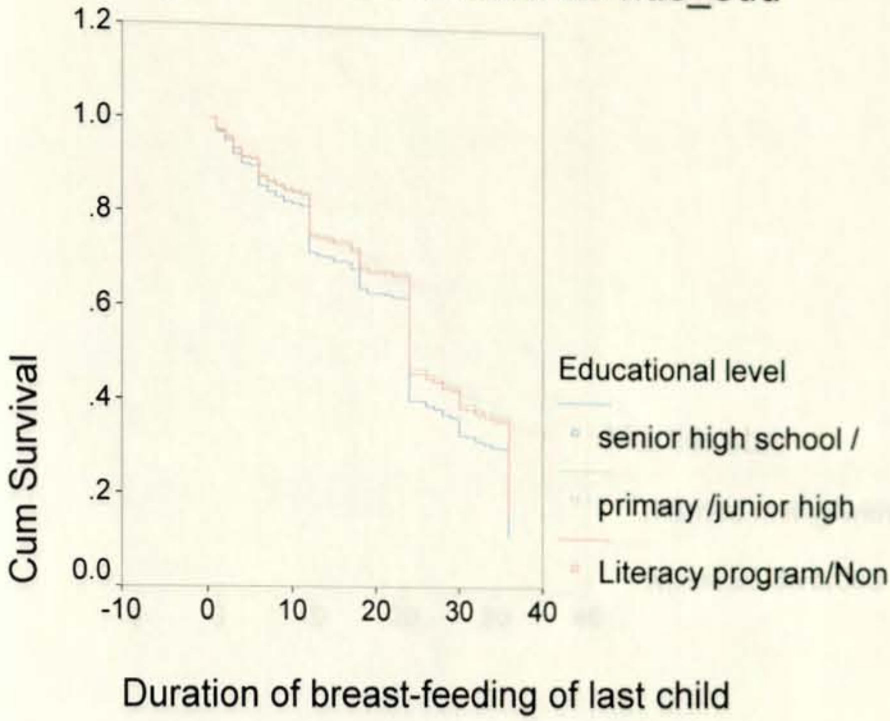


Fig.4 Survival Function for educatio

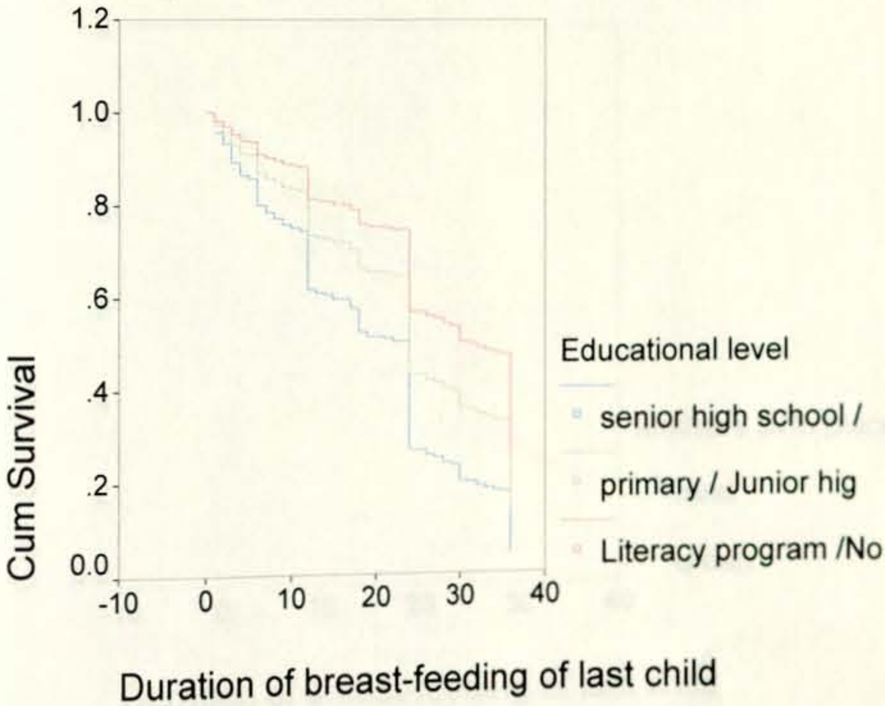


Fig.5 Survival Function for marit

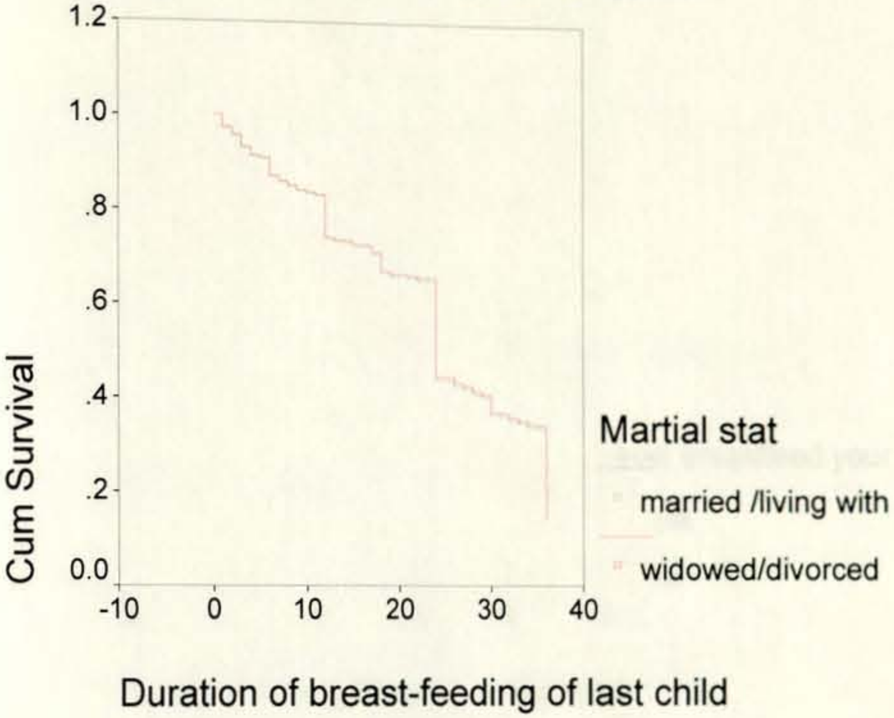


Fig.6 Survival Function for motbirpl

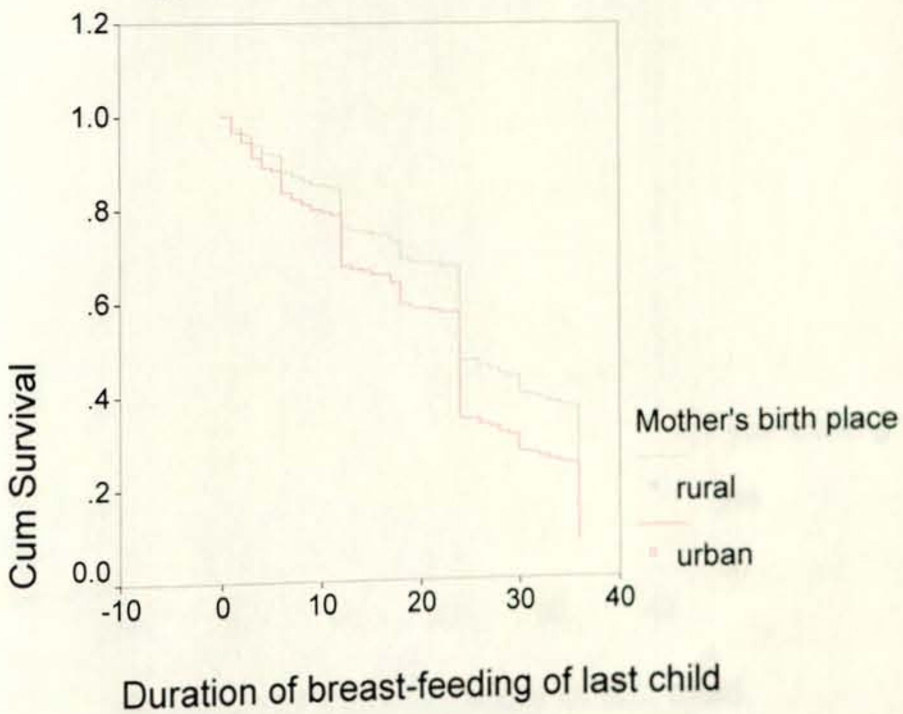


Fig.7 Survival Function for ptbreast

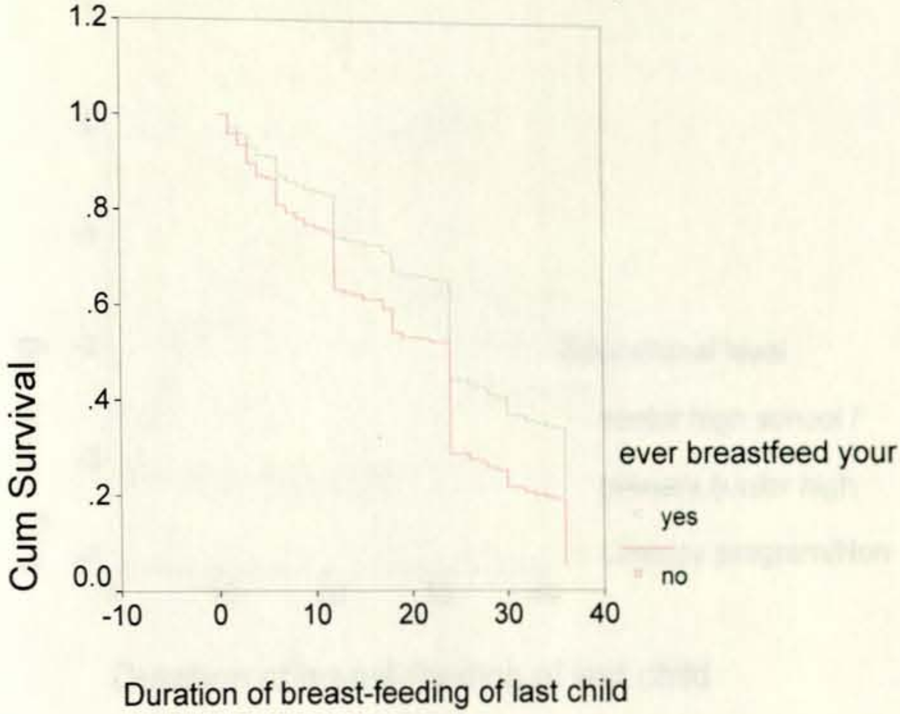


Fig.8 Survival Function for working

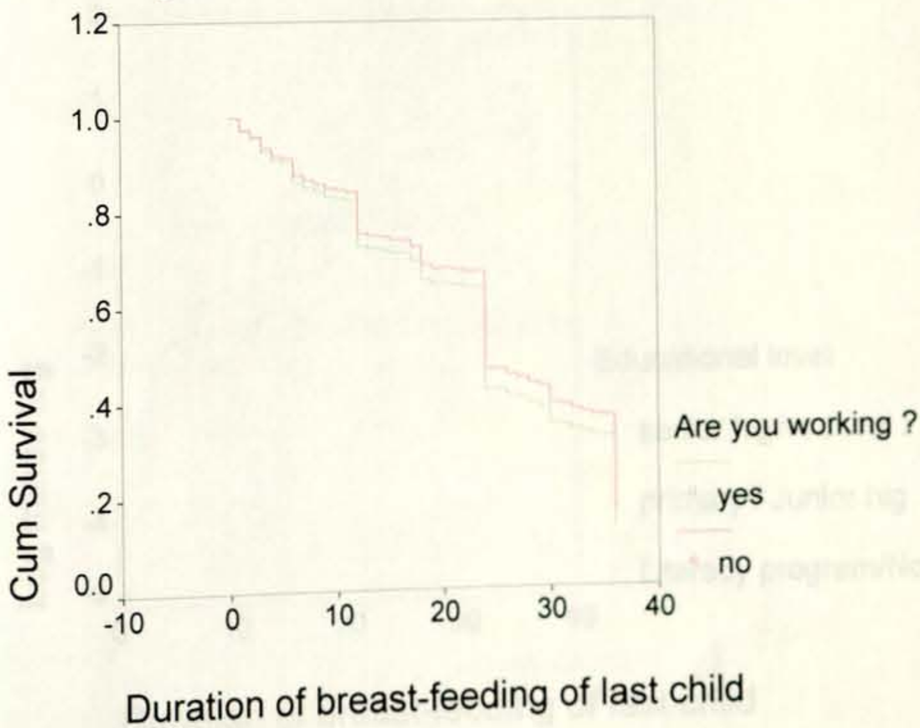


Fig.9 The LML Function for hus_edu

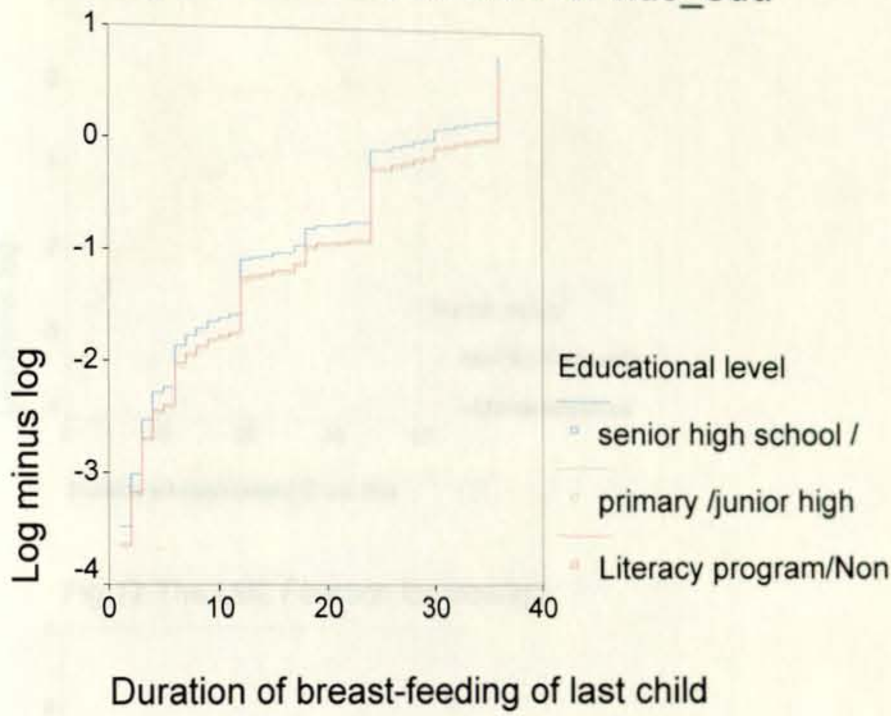


Fig.10 The LML Function for educatio

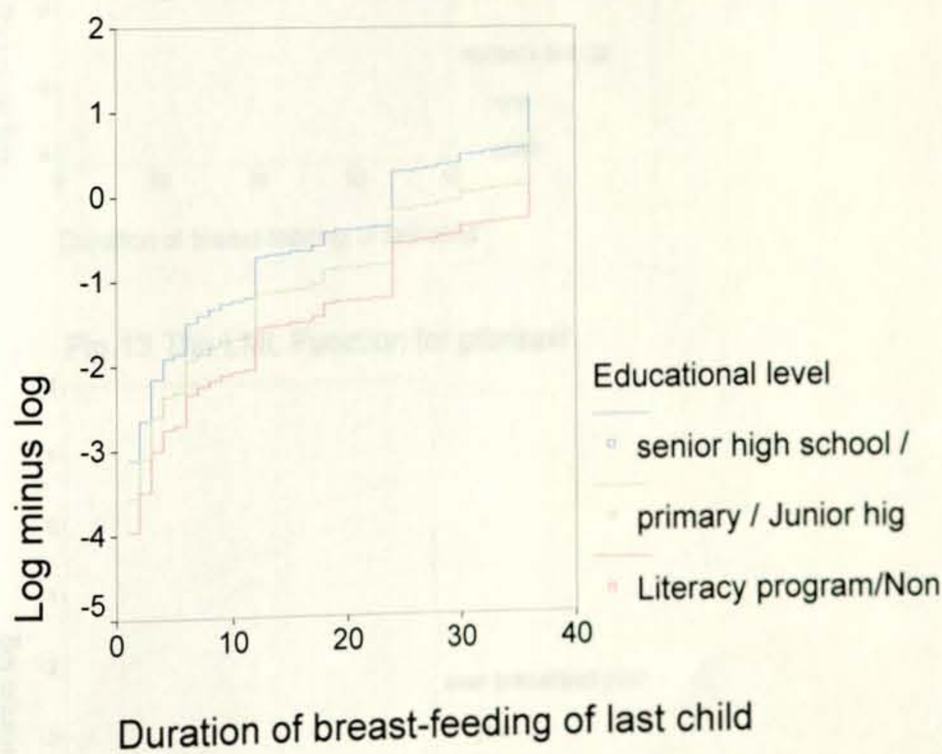


Fig.11 The LML Function for marit

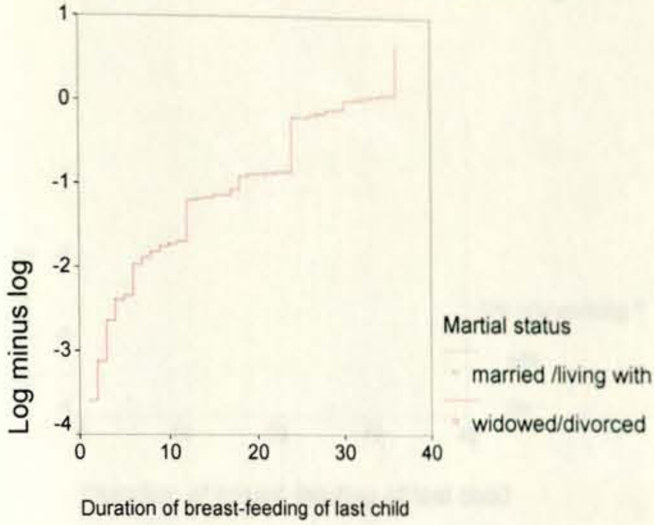


Fig.12 The LML Function for motbirpl

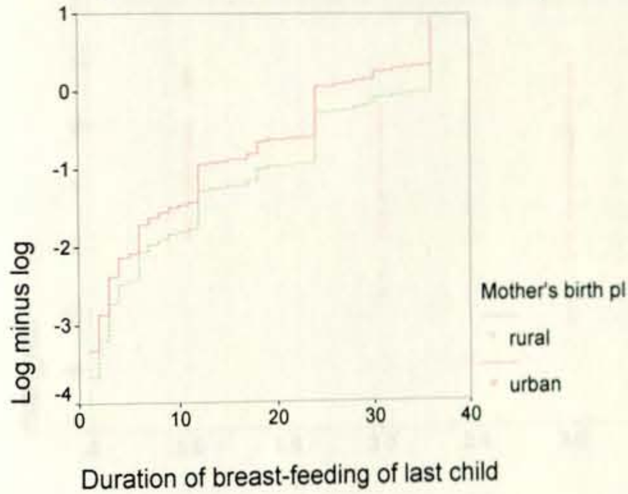


Fig.13 The LML Function for ptbreast

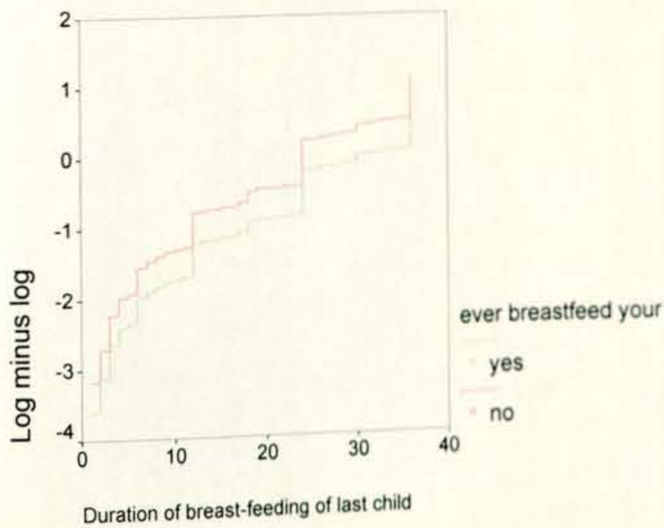
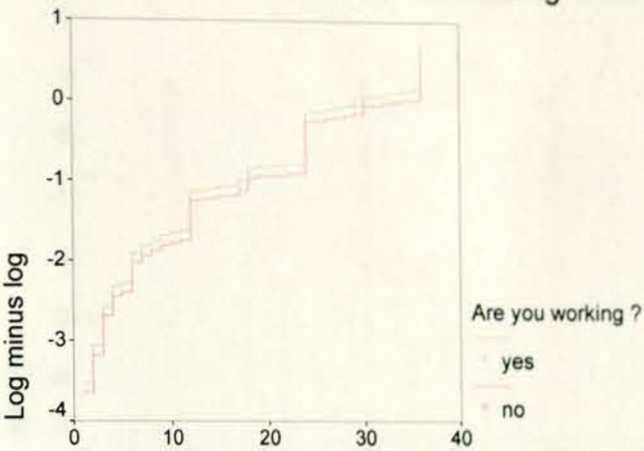
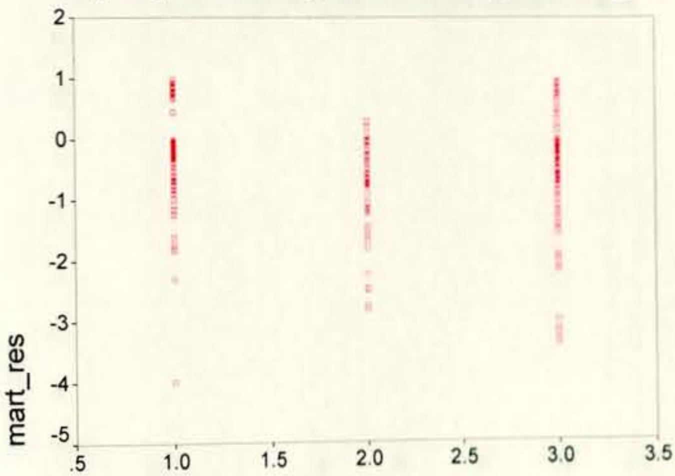


Fig.14 The LML Function for working



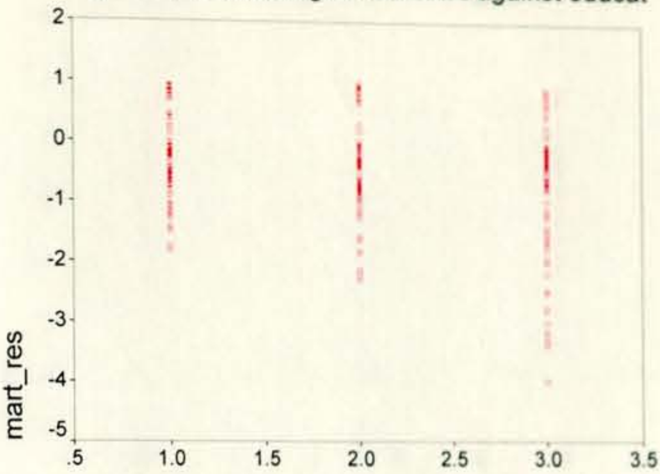
Duration of breast-feeding of last child

Fig.15 plot of matingale residuals against hus_edu



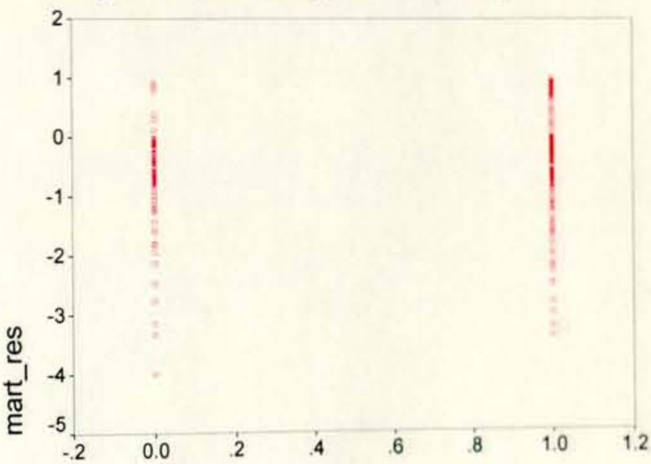
Educational level

Fig.16 Plot of martingale residuals against educat



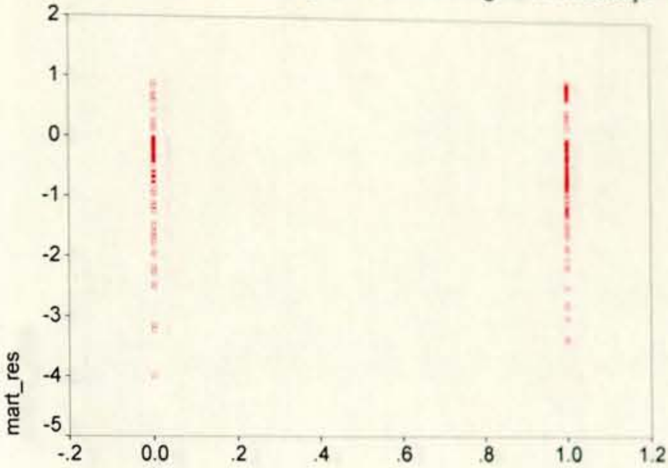
Educational level

Fig. 17 Plot of martingale residuals against marit



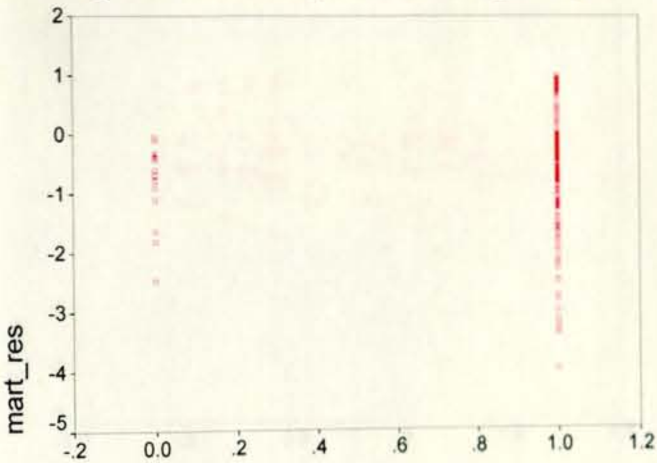
Martial status

Fig.18 Plot of martingale residualsagainst motbirpl



Mother's birth place

Fig.19 Plot of martingale residuals against ptbreast



ever breastfeed your previous child

Fig.20 Plot of martingale residuals against working

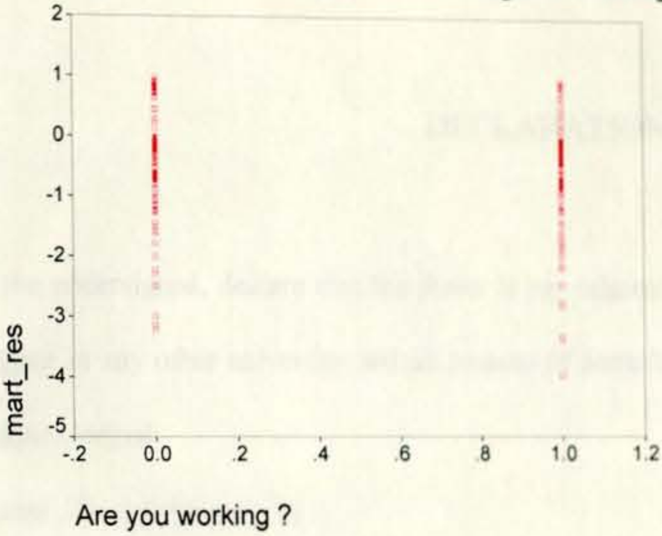
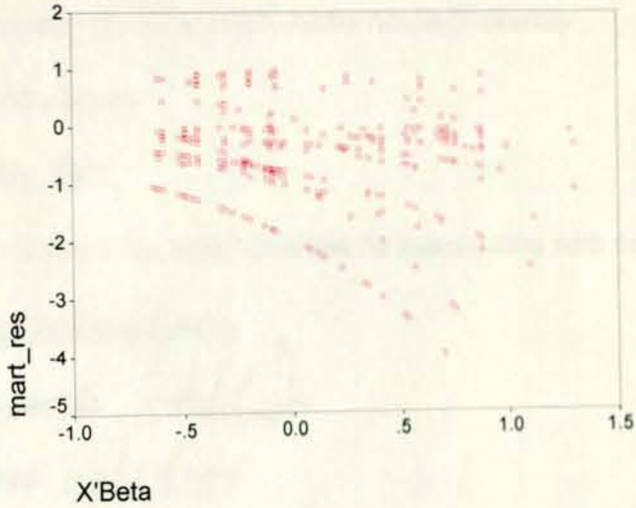


Fig.21 plot of martingale residuals against x'Beta



DECLARATION

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

Name Daniel Alemayehu

Signature_____________

Department of Statistics, Addis Ababa University

Addis Ababa

May 2001

This thesis has been submitted for examination with my approval as a university advisor.

Dr. Tadewos Koroto

Signature_____________

Date 02/07/2001