MOTHER'S WORK STATUS AND INFANT MORTALITY IN ETHIOPIA: A STUDY BASED ON DEMOGRAPHIC AND HEALTH SURVEY DATA

BY

ESAYAS MULETA

A Thesis Submitted to the School of Graduate Studies of

Addis Ababa University in partial fulfillment of the requirements for

the degree of Master of Science in Statistics

February, 2003

Addis Ababa, Ethiopia

I am extremely grateful to the Central Statistical Authority for the opportunity it offered me to undertake my Master of Statistics study in AAU and for providing me the Ethiopia Demographic and Health Survey (EDHS) raw data to use in this thesis.

I am also grateful to my friend Eskinder Tenaw for his devoted assistance, thoroughly reading the manuscript, making comments and valuable suggestions during the period of my research work.

I would like to acknowledge the support I got from my friends and colleagues Nigussie Geletu, Dawit Getnet, Firew Ayalew and Kifle Gebre for their generous co-operation in computer facilities and other related works throughout the study.

Finally, my heart felt gratitude goes to my family, Helen Fisseha and Dr. Fisseha Muleta for the support rendered during my period of study.

TABLE OF CONTENTS

Abstract

Despite its many advantages, the work of women in economic activities in Ethiopia has been associated with increased and decreased mortality of infants. Thus, this study examines whether these conclusions are upheld at the level of the typical Ethiopian mother.

Using data from the Ethiopia Demographic and health Survey (EDHS) in year 2000, the effect of work status of mothers on infant survival is investigated at country level. The study uses information on 15,367 women of age 15-49 included in the survey for the entire country. The effect of work is also evaluated separately by type of work whether the work is professional/technical/clerical or agricultural/manual. Cox regression model is used to examine the association between infant mortality and mother's work stuatus.

Using the Kaplan-Meier estimation technique, the univariate analysis shows that survival of infants is 93.2 percent with standard deviation of 0.0060. The bivariate comparison of infant mortality rates for the period two years before the survey according to mother's work status reveals that mothers who are working had no significant difference on infant mortality from mothers who are not working. These results are largely upheld in the multivariate analysis. However, according to the type of work, the relative risk of infant mortality for agricultural/manual is 1.170 times higher than non-workers. The risk of death is also significantly lower among professional/technical/clerical workers (32 percent) than non-working mothers. Multivariate analysis assesses the strength of apparent association between work status of the mother and infant mortality by controlling other characteristics likely to influence the outcomes.
The relative risks associated with several other variables are statistically significant and in the expected direction. Among these factors the length of the preceding birth interval for infants exerted an expected beneficial effect on the hazard ratio of infants. There is also evidence of a strong detrimental effect of the birth order on the hazard ratio of infant death. As expected, being first born significantly increases the probability of dying at infant stage. The risk of infant death is higher in those who are not married than married; among other variables the probability of infant death will be lower if the mother is educated.

The observation that high frequency of infant death occurred when mothers are engaged in agricultural/manual work does not in any way imply that this type of work should be discouraged. The higher mortality of infant for mothers who work in such low-status work reflects the fact that work for women is an additional duty and burden to their traditionally ascribed responsibilities.

CHAPTER 1

INTRODUCTION

1.1 Background of the study

Employment of women in economic activities has several beneficial effects for women and their families. There are reasons to expect women's participation in the labor force to have beneficial as well as detrimental effects on infant survival. Some beneficial effects of women's labor force participation will derive as a positive income effect of the earnings of mother enhanced by the likelihood that a higher proportion of these earnings compared with earnings of males, will be directed toward infant welfare needs (Kumar 1977, Mencher 1988). In addition, women’s employment may translate into greater control over expending of resources, increased exposure and access to relevant information about childbearing and childrearing practices, and an enhanced ability to manipulate and engage the world outside the home to better meet the nutritive, medical, and survival needs of infants.

Aside from housework most women in the rural areas of the country usually do some kind of work including working as skilled agricultural worker or non-skilled agricultural worker on their own land or someone else's land during the active year or part of the year, etc.

Proper identification of these associated factors constitute the basis to reducing the level of infant mortality and provide information for policy making on the major factors that lead to infant deaths in the country.

Based on nation-wide representative Demographic and Health Survey Data of the year 2000(EDHS 2000), this study attempts to provide a broader picture on factors influencing survival status of infants from women work status perspective in Ethiopia.

1.2 Statement of the problem

Despite the fact that the determinants of infant mortality have been researched to some extent, the relationship between the survival time of an infant and the work status of a mother has not been investigated. Obviously there are a variety of mechanisms by which work status of a mother could have impact on infant mortality.

The benefits of women's work status are likely to depend critically on the type of work in which women are engaged. For example, income effects are most likely to ensue when women earn cash, and employment is most likely to lead to increased engagement with the wider world when women's work takes them outside the home.

Counterbalancing these effects is the reduced availability of time and consequent likelihood of increased inability of working women to provide personal and timely care for their infants. To the extent that this is true, any negative consequences for the health and welfare of infants are likely to be exacerbated by the lack of appropriate care and the necessity among poor rural women to spend time fetching water and firewood.

More direct effects on the nutrition of infants and shortened breast-feeding among mothers who work also have been noted in different contexts.

The type of work that women engage in will be an important mediating factor for the overall relationship between mother's work status and infant mortality.

Work for most women in the Ethiopian context is far from empowering, concentrated as it is in low-paying, low-productivity occupation, often requiring hard labor, and not necessarily ensuring control over income earned. Three fourths of Ethiopian women who work for cash reported that they alone are mainly responsible for making decisions on how their earnings are spent (EDHS 2000).

The primary reason for expecting a relationship between mother's work status and infant mortality, especially in the rural area, concerns the extent to which the amount of resources flowing into the household might differ by the working status of the mother. If the woman is working more, she has less time to devote to the rearing of her children and obviously this will increase the risk of mortality of babies.

Another possible explanation for a relationship between mother's work status and infant mortality is based on the assumption that, in rural areas, mothers who work are of low economic status.

Therefore, this could also have negative impact on the survival of an infant. Hence this study is an attempt to investigate the impact of work status of a mother on infant mortality.

1.3 Objectives of the study

- **General objectives**

  The purpose of this study is to examine the influence of work status of mothers on infant mortality in Ethiopia.

- **Specific objectives**

  1. To estimate the survival time of infants.
  2. To examine the relationship between the survival times of infant among working versus non-working mothers, professional/technical/clerical versus agricultural/manual workers.
  3. To make relevant recommendations for policy makers, program managers as well as development planners.

1.4 Scope of the study

This study will use secondary data collected by Central Statistical Authority (CSA), i.e., the Ethiopia Demographic and Health Survey of the year 2000(EDHS 2000). The study uses information on 15,367 women of age 15-49 included in the survey for the entire country. The study focuses on the investigation of the effect of work status of these women on the mortality of their infants.

CHAPTER 2

LITERATURE REVIEW

There has been a less than rewarding obsession with dividing women into two categories: “workers” and “nonworkers.” Indeed, a growing body of literature is devoted to demonstrating the falsity of this dichotomy in the Third World context (Rogers, 1980). Work has too often been defined as tasks performed by men. By this criterion, housework is not work even where it entails heavy duties for many hours a day.

Women may perform economic activities within the home, in the informal sector outside the home, in such occupation as farm laborer or domestic worker in other households (in which cases it may be possible to take infants along), or in places where it may not be possible to take infants along. From this point of view it is not only the occupation but also the circumstances in which it is carried out. Women’s economic activities will have a negative impact on childcare only where the activity is incompatible with simultaneous child rearing or where the mother lacks access to another person able to care for the child.

Childrearing is often thought of as being incompatible with women’s economic activities only where women work in the modern formal sector. However, three factors are relevant here: the compatibility of the task itself with child care, the availability of relatives to provide substitute child care of an infant, and the distance between the work place and the home. In the town the women’s work place is likely to be at such a distance from the home that she is not even available for breastfeeding (Manderson, 1982).

Village infants may retain being breastfed even when their mothers work over 12 hours a day in the fields, but other aspects of child care may well be neglected and supplementary feeding may be absent or totally unsuitable due to shortage of money.

In Sudan Farah and Preston (1982) found that the mother’s participation in the labor force raised infant mortality by 27 percent in the capital, as compared to 10 percent for the country as a whole, possibly because educated women employed in the capital are more seriously disadvantaged by entrusting child care to illiterate maids or relatives.

Research in India, suggested that women’s employment may have at least one disadvantage: the survival of young children appears to be negatively affected if women work (Basu and Basu 1991; Kishor 1992). Past research to identify the association between women’s employment and child mortality, which has been more comprehensive at the district level than at the household or woman level, show higher rates of child mortality net of several cultural and economic factors (Kishor 1992, 1993). Other researchers found also that work status of women has no relationship with infant mortality (Murthi, Guoio, and Dreze, 1995).

CHAPTER 3

SOURCE OF DATA AND METHODOLOGY

3.1 Source of data

The data for this study were obtained from the 2000 Ethiopia Demographic and Health survey (EDHS) conducted in Ethiopia as part of the worldwide demographic and health survey project. The survey collected Demographic and health information from a nationally representative sample of women in the reproductive age group 15-49.

Using systematic sampling with probabilities proportional to size, 539 enumeration areas (EAS), a unit of land that usually consists 150-200 households delineated for the purpose of enumerating housing unit and population, were selected initially. A complete household listing operation was carried out in each selected EA, and systematic samples of 27 households per EA were selected in all the regions in the second stage. The survey interviewed 15,367 women aged 15-49 in 14,072 households between February and May 2000.

3.2 Study variables

3.2.1 Response variable

In the study duration of survival of the infant, measured in months, is the response variable and is based on information pertaining to the last child of the mother.

Since there are censored observations the coding of the status variable is such that 1 indicates uncensored observation (for those who died before celebrating their first birth day) and 0 for censored observations (for those who died after celebrating their first birthday or those who are alive at the time of the survey).

3.2.2 Explanatory variables

In the survey two questions were used to draw facts about the work status of women. All eligible women were first asked, "As you know, some women take up jobs for which they are paid in cash or kind. Others sell things, have a small business or work on the family farm or in the family business. Are you currently doing any of these things or any other work?" Women who said 'no' were asked 'Aside from housework, have you done any work in the last 12 months?'

Women who answered 'no' to both questions were counted as not employed, women who answered 'yes' to either the first or the second question were then asked details about the work they did including their usual occupation.

Accordingly employed women were asked whether they participated in agriculture or not. And then asked the type of land ownership and the time they spent for those working on agriculture sector. And also for non-agricultural workers the time they spent and for whom they are working was asked. For all employed women the kind of payment and place of work were recorded.

Hence the following explanatory variables are considered in this study.

Table1. Definition of independent variables used in the analysis.

Variable	**Variable description**	**Category**
Work	Type of work	1=Non-working 2=professional/technical/ clerical 3=Agricultural/Manual
Agelach	Age of last child	Age of last child (in months)
Status	Status of the infant	0=censored 1=uncensored

In addition to work status of the mother, the following independent variables, which are expected to have an impact on infant mortality, are noted to see their effect on infant mortality and the interaction they have with the work status of the mother.

Socio economic variables

Variable	**Variable description**	**Category**
Ageres	Age of respondent	In completed years 1: 15-24 years 2: 25- 34 years 3: 35- 49 years
Edures	Literacy status of respondent	1: illiterate 2: literate
Eduhus	Literacy status of the husband	1: illiterate 2: literate
Residenc	Place of residence	1: urban 2: rural
Marital	Marital status	1: married 2: widowed/divorced
Pbinterv	Preceding birth interval	1: two years or less 2: greater than 2 years & 1st birth
Birordr	Birth order	1: one child 2: between two and four children 3: five or more children

3.3 Methodology

3.3.1 Survival analysis

Survival analysis is the phrase used to describe the analysis of data that correspond to the time from a well-defined time origin until the occurrence of some particular event or end-point.

Survival data are not amenable to standard statistical procedures used in data analysis. One of the features of survival data that renders standard methods inappropriate is that survival times are frequently censored. The survival time of an individual is said to be censored when the end-point of interest has not been observed for that individual.

In practice, the actual date recorded will be the date on which each individual enters the study, and the date on which each individual dies or was last known to be alive. The survival time in days, weeks or months, whichever is the most appropriate, can then be calculated.

Survivor function and hazard function are the two functions of central interest in summarizing survival data. The actual survival time of an individual, t, can be regarded as the value of a random variable T, which can take any non-negative value. The different values that T can take have a probability distribution, and we call T the random variable associated with the survival time. When the random variable T has a probability distribution with underlying probability density function f(t), the distribution function of T is then given by

$$F(t) = P(T<t)$$

and represents the probability that the survival time is less than some value t.

The survivor function, S(t), is defined to be the probability that the survival time is greater than or equal to t, and so

$$S(t) = P(T \geq t) = 1 - F(t).$$

The survivor function can therefore be used to represent the probability that an individual dies at time t, conditional on having survived to that time. That is, the function represents the instantaneous death rate for an individual surviving to time t. Thus, the hazard function, h (t), is defined as

$$h(t) = \lim_{\delta t \to o} \left\{ p\left(\frac{t \leq T < t + \delta t / T \geq t}{\delta t} \right) \right\}$$

From this definition the relationship between the survivor and hazard function, can be expressed as

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

where f(t) is the probability density function of T.

The survival and hazard functions are estimated using the Kaplan-Meier method as a preliminary analysis. This method is non-parametric or distribution-free, since it does not require specific assumption to be made about the underlying distribution of the survival times.

To apply the Kaplan-Meier method suppose that there are n independent individuals in a random sample with observed survival times $t_1, t_2, \ldots, t_n$. The distinct ordered failure times observed among the n individuals are $t_{(1)}, t_{(2)}, \ldots, t_{(r)}$, $r<n$ as there are more than one individual with the same observed survival time and some of the observations may be right-censored, i.e., the survival status of the individual might not be known at the time of the analysis.

The probability of survival at time $t_{(j)}$, $P(t_{(j)})$ is then estimated by

$$P(t_{(j)}) = (n_j - d_j)/n_j$$

where n_j is the number of individuals who are alive just before time $t_{(j)}$ and d_j is the number who die at this time.

Consequently the estimated probability of surviving beyond $t_{(j)}$, S (t) is

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

With the approximated standard error given by

$$\text{s.e}\left\{\hat{S}(t)\right\} = \hat{S}(t)\left\{\sum_{j=1}^{k}\frac{d_j}{n_j(n_j - d_j)}\right\}^{1/2}$$

for $t_{(k)} \leq t < t_{(k+1)}$

The ratio of the number of deaths at a given death time to the number of individuals at risk at that time is used to estimate the hazard function for ungrouped survival data. But if the hazard function is assumed to be constant between successive death times, the hazard per unit time can be found by further dividing by the time interval τ_j. Thus, if there are d_j failures at the j^{th} failure time, t_j and n_j at risk at time $t_{(j)}$, the hazard function in the interval from $t_{(j)}$ to $t_{(j+1)}$ can be estimated by

$$\hat{h}(t) = \frac{d_j}{n_j \tau_j}$$

for $t_{(j)} \leq t < t_{(j+1)}$ where $\tau_j = t_{(j+1)} - t_{(j)}$

When the survival distributions of two or more groups of survival data are to be compared the null hypothesis to be tested is that there is no difference between two or more groups against the alternative that there are at least two groups with different survival curves.

Let $t_{(1)} < t_{(2)} < \ldots < t_{(r)}$ be r distinct failure times across g categories, and at time $t_{(j)}$, d_{kj} individuals in group k fail for $k = 1, 2, \ldots, g$, $j = 1, \ldots, r$. Also suppose that n_{kj} individuals are at risk of failure in the k^{th} category just before time $t_{(j)}$. There are $d_j = \sum_{k=1}^{g} d_{kj}$ failures in total out of $n_j = \sum_{k=1}^{g} n_{kj}$ individuals at risk. Under the null hypothesis, 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,

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

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 statistics is given by

$$U_k = \sum_{j=1}^{r} \left(d_{kj} - e_{kj}\right), \quad \text{for } k = 1,\ldots,g-1$$

In order to test the null hypothesis of no group differences, we use the test statistic

$$\mathbf{U}^{1}{}_{L}\mathbf{V}_{L}^{-1}\mathbf{U}_{L}$$

which has a chi-square distribution with g-1 degrees of freedom, where $\mathbf{U}_L$ is a vector with g-1 components and $\mathbf{V}_L$ is the variance-covariance matrix (Collet, 1994).

3.3.2 Cox regression model

The Cox regression model is used to determine which combination of explanatory variables affect the form of the hazard function. Also the model is used to obtain an estimate of the hazard function itself for an individual which may be of interest in its own right. But in addition, from the relationship between the survivor function and hazard function, an estimate of the survivor function can be found. This will in turn lead to an estimate of quantities such as the median survival time, which will be a function of the explanatory variables in the model.

Cox regression model is somewhat different in form from linear models encountered in regression analysis and in the analysis of data from designed experiments, where the dependence of the mean response, or some function of it, on certain explanatory variables is modeled. However, many of the principles and procedures used in linear modeling carry over to the modeling of survival data. The Cox regression model can be used for data that contain censored observations. The model also takes into account the fact that the probability of experiencing an event differs with duration of exposure to risk.

3.3.3 Cox proportional hazards model

The Kaplan-Meier and log-rank methods described are useful in the analysis of a single sample of survival data, or in the comparison of two or more groups of survival times. However, in most studies which give rise to survival data, supplementary information will also be recorded on each individual which are referred to as explanatory variables. The basic model to be considered here is the proportional hazards model.

The assumption of proportional hazards is that the hazard of death 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. When there are covariates in the analysis, which are time dependent, this assumption may not hold. But in this study it can be assumed that the ratio of the hazards for two or more groups are constant for all time points since we have covariate information only at the time of the survey.

A useful plot for assessing whether this assumption holds or not is the log-minus-log (LMC) plot (Marubini and Valseechi, 1995). The set of values of the explanatory variables in the proportional hazards model will be represented by vector X. Let $h_o(t)$ be the hazard function for an individual for whom the values of all explanatory variables that make up the vector X are zero. The function $h_o(t)$ is called the baseline hazard function. The hazard function for the individual can then be written as

$$h_i(t) = h_0(t) \exp(\beta' X)$$

where β is a p x 1 vector of regression coefficients. Suppose that data are available for n individuals, amongst whom there are r distinct failure times and n - r right-censored survival times so that there are no ties in the data. The r ordered death times will be denoted by $t_{(1)} < t_{(2)} < \ldots < t_{(r)}$, so that $t_{(j)}$ is the j^{th} ordered death time.

The set of individuals who are at risk at time t_j is the j^{th} ordered death time. The set of individuals who are at risk at time t_j will be denoted by R ($t_{(j)}$), so that R ($t_{(j)}$) is the set of individuals who are alive and uncensored at a time just prior to t_j.

The relative likelihood function for the proportional hazards model is given by

$$L(\beta)=\prod_{i=1}^{r}\frac{\exp(\beta' X_i)}{\sum_{l \varepsilon R\,(t_{(j)})}\exp(\beta' X_l)}$$

Now suppose that the data consist of n observed survival times, denoted by t_1, t_2, …, t_n and that δ_i is a censoring indicator which is zero if the i^{th} survival time t_i, i = 1, 2,…., n is right-censored, and unity otherwise. The likelihood function can then be expressed in the form

$$\prod_{i=1}^{n}\left[\frac{\exp(\beta' X_i)}{\sum_{l \varepsilon R(t_{(i)})'}\exp(\beta' X_l)}\right]^{\delta_i}$$

where R ($t_{(i)}$) is the risk set at time t_i. The corresponding log-likelihood function is given by

$$LogL(\beta)=\sum_{i=1}^{n}\delta_i\left\{\beta' X_i-\log\sum_{l \varepsilon R(t_{(i)})}\exp(\beta' X_l)\right\}$$

The proportional hazards model for survival data assumes that the hazard function is continuous, and under this assumption, tied survival times are not possible. Of course, survival times are usually recorded to the nearest day, month or year, and so tied survival times can arise as a result of this rounding process.

In order to accommodate tied observations, the likelihood function has to be modified in some way. The appropriate likelihood function in the presence of tied observations is given in Kalbfleish and Prentice (1980).

3.3.4 Variable selection procedures

In a modeling approach to the analysis of survival data, a model is developed for the dependence of the hazard function on one or more explanatory variables. In this process, models that contain different sets of terms are fitted, and comparisons are made between them.

In order to compare alternative models fitted to an observed set of survival data, a statistic, which measures the extent to which the data are fitted by a particular model, is required.

Since the likelihood function summarizes the information that the data contain about the unknown parameters in a given model, a suitable summary statistic is the value of the likelihood function when the parameters are replaced by their maximum likelihood estimates.

Hence, for a given set of data, the larger the value of the maximized likelihood, the better is the agreement between the model and the observed data. Comparisons between a number of possible models can be made on the basis of the statistic

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

where q is the number of unknown β parameters in the model and α is a predetermined constant. The value of α is usually taken to be between 2 and 6. This statistic is known as Akaike's information criterion (AIC). $\hat{L}$ is the maximized likelihood. The smaller the value of this statistic, the better the model.

The choice $\alpha = 3$ is roughly equivalent to using a 5% significance level in judging the difference between the values of -2 log $\hat{L}$ for two models which is recommended for general use Collet (1994).

These automatic routines have a number of disadvantages. Typically, they lead to the identification of one particular subset, rather than a set of equally good ones. The subsets found by these routines often depend on the variable selection process that has been used, that is, whether it is forward selection, backward elimination or the stepwise procedure, and generally tend not to take any account of the hierarchic principle. They also depend on the stopping rule that is used to determine whether a term should be included in or excluded from a model.

Thus, instead of using automatic variable selection procedures, the following general strategy for model selection is recommended Collet (1994).

1. The first step is to fit models that contain each of the variables one at a time. The values of – 2 $\log \hat{L}$ for these models are then compared with that for the null model. The null model is a model where there is no any explanatory variable regressed on the response variable. Then this model is used to determine which variables on their own significantly reduce the value of this statistic.

2. The variables that appear to be important from step 1 are then fitted. In the presence of certain variables others may cease to be important. Consequently, those variables, which do not significantly increase the value of -2 $\log \hat{L}$ when they are omitted from the model, can now be discarded. We therefore compute the change in the value of -2 log $\hat{L}$ when each variable on its own is omitted from the set. Only those that lead to a significant increase in the value of -2 $\log \hat{L}$ are retained in the model. Once a variable has been dropped, the effect of omitting each of the remaining variables in turn should be examined.

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

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

When using this selection procedure, rigid application of a particular significance level should be avoided. In order to guide decisions on whether to include or omit a term, the significance level should not be too small. A level of around 10% is recommended.

3.3.5 Model checking

The adequacy of the model needs to be assessed after the model has been fitted to observed survival data. Many model-checking procedures are based on quantities known as residuals. These are values, which can be calculated for each individual in the study, and have the feature that the behavior is known when the fitted model is satisfactory.

When the Cox regression model has been fitted to the survival times, and that the linear component of the model contains p explanatory variables, X_1, X_2, …, X_p, then the Cox-Snell residual for the i^{th} individual, i=1, 2, ... , n is given by

$$r_{Ci} = \exp (b'x_i)H_o(t_i),$$

where $H_o(t_i)$ is the estimated cumulative base line hazard function at time t_i, the observed survival time of that individual. The Cox-Snell residuals r_{Ci} have properties that are quiet dissimilar to those of residuals used in linear regression analysis. They will not be symmetrically distributed above zero and they cannot be negative. Furthermore, when an appropriate model has been fitted, they have a highly skewed distribution.

Censored observations lead to residuals that cannot be regarded on the same footing as residuals derived from uncensored observations. Therefore to take an account for censoring, the modified Cox-Snell (martingale residual) for the i'th individual is given as

$$r_{Mi} = \delta_i - r_{Ci}$$

where $\delta_i = 0$ for censored observations and $\delta_i = 1$ for uncensored observations.

In large samples, the martingale residuals are uncorrelated with one another and have an expected value of zero. In this respect, they have properties similar to those possessed by residuals encountered in linear regression analysis. However, the martingale residuals are not symmetrically distributed about zero.

Plot of residuals against explanatory variables can be used to indicate whether any particular variable needs to be transformed prior to incorporating it in the model. When the residuals are plotted against explanatory variables that are not in the fitted Cox regression model they do not show an obvious relationship, then it can be suggested that the variable is not needed in the model. Hence in the plots of the martingale residuals against the values of the explanatory variables, if most of the points fall horizontally about zero then the fitted model is taken as satisfactory.

CHAPTER 4

RESULTS AND DICUSSIONS

Analysis of the data and the results are discussed in this chapter. Comparison between working and non-working mothers; professional/technical/clerical workers and agricultural/manual workers are also made. Assessing missing values and non- responses was accomplished before analyzing the data because their presence or absence could significantly change the analyis. No missing values and non-responses are observed in the data. Therefore weighting the data is found to be not advisable. So the analysis is done on the un-weighted cases. The total interviewed women age 15-49 at national level was found to be 15,367, of which 10,143 are women who gave birth to a child.

Here some caveats about the data need to be noted. The data on work status of the mother are based on self-reported work, which generally under-estimates the true extent of women's work. Thus the relationships found in this analysis are really between work that women themselves perceive as work and mortality of infants. Also, the work status refer to current work of women, while the demographic event of interest, that is, births and deaths, could have occurred any time in the past. The analysis is restricted to birth and death occurred within two years from the date of the interview. This is based on the assumption that the events occurred years back have no relation with current work status. The effective sample then reduces to 5,207 mothers with complete information in the two years before the survey.

Results of the analysis are given below and figures showing several intermediate results are also given in the Appendix.

4.1 Univariate examination of infant mortality

The probability of survival of infants is obtained using the Kaplan-Meier estimation technique. It was found to be 93.24% with standard error of 0.0060. The corresponding estimates for the months 0 to 11 are presented in Table-2 below.

Table – 2 Kaplan-Meier estimates of survival of infants

Time	Cumulative Survival	Standard Error
0	.9697	.0039
1	.9645	.0042
2	.9619	.0043
3	.9566	.0046
4	.9544	.0048
5	.9527	.0048
6	.9473	.0051
7	.9423	.0054
8	.9410	.0055
9	.9382	.0056
10	.9331	.0059

11	.9324	.0060

Besides, among the 15.3% children who died before 1st birth day, half of them died at an age of one month. Hence the median age is found to be one month. Moreover from the result it is also observed that the first and second quartile ages are zero and one month respectively. The third quartile age is five months. The semi inter-quartile range was found to be 2.5 months. This result also shows that there is a moderate variation in infant mortality.

4.2 Bivariate examination of infant mortality

Controls for a variety of biological factors known to influence infant mortality are incorporated into analysis including women's age, parity and previous birth interval. The association between both maternal age and parity with infant mortality is well known, with elevated risk among very young women experiencing first births, and among older women and higher order births. A control for previous birth interval is also incorporated due to its known influence on infant and child survival (Koenig, Phillips, Campbell, & D'Souza, 1990).

A categorical variable is developed to measure the type of education received by women in the sample. Mother's education has been found to be positively associated with an increase in survival of children (Bicego and Boerma 1991; Caldwell 1979, 1986; Cleland 1990; Hobcraft 1996).

In addition to these explanatory variables education of the husband, marital status and residence of the mother are also included in the analysis.

Table 3 given below presents bivariate statistics, which examine the relationship between infant mortality and each of the independent variables of interest. All the explanatory variables in the analysis are treated as time constant covariates.

Table 3. **<u>Test Statistics for Equality of Survival distributions</u>**

Variable	Log-rank statistics	df	Significance
Work	13.34	2	.0015
Ageres	8.57	2	.0137
Edures	.74	1	.3908
Eduhus	.90	1	.3429
Residenc	.10	1	.7467
Marital	27.27	1	.0000
Pbinterv	2.56	1	.1097
Birordr	27.47	2	.0000

Results of the Log-rank test above indicate that for the variables work, age, marital status and birth order the hypothesis that the survival functions are equal among the groups was rejected at the 5% significance level, which means the hazard of infant mortality from group to group varies significantly. However as it can be seen from the same output education and residence seems not to have a significant association with infant mortality. This does not however mean that these variables are not important and should be considered in a further analysis. But since these variables may have an indirect effect when they are seen together with other explanatory variables it is necessary to consider them in the multivariate case.

Moreover, in order to see how significantly each category of a given variable (for variables that have more than two categories) differ from the other category the pairwise Log-rank test was also applied. Like the above case this result also shows that all comparisons of education and residence are insignificant. On the other hand all category comparison of work status of the mother, age, marital status and birth order are significant at 5% level. The result found when each variable was independently regressed using the Cox-regression model is given in Table 4 below.

Table 4 Un-adjusted Cox regression model

Variable	B	SE	Wald	df	Sig.	Exp(B)	-2 Log likelihood
Work(not working)			12.830	2	.002		5490.170
Profess/technical/clerical	-.405	.167	5.888	1	.015	.667	
Agri/manual	.174	.122	2.040	1	.053	1.190	
Ageres(15-24)			8.346	2	.015		5513.565
25-34	-.343	.124	7.643	1	.006	.710	
35-45	-.294	.152	3.728	1	.054	.745	
Edures(illiterate)	-.121	.143	.723	1	.395	.886	5520.944

Eduhus(illiterate)	-.109	.116	.883	1	.347	.896	5520.793
Residenc(urban)	-.048	.149	.102	1	.749	.953	5521.584
Marital(married)	.770	.152	25.505	1	.000	2.160	5500.357
Pbinterv(2 years & below)	-.219	.139	2.503	1	.114	.803	5519.291
Biordr(1^{st} child) 2-4 children more than 5	-.565	.138	16.733	1	.000	.568	5497.619

* The reference categories are those indicated in brackets.

4.3 Multivariate examination of infant mortality

When there are a number of explanatory variables of possible relevance, the effect of each term cannot be studied independently of the others. The effect of any given term therefore depends on the other terms currently included in the model. Hence in the bivariate analysis technique the relations which are obtained for one factor do not take into account the other factors. So the multivariate analysis is used to know the most important factors in this analysis.

An initial step in the model selection process is to identify sets of explanatory variables that have the potential for being included in the linear component of a proportional hazards model. This set will contain those factors which have been recorded for each individual, but additionally terms corresponding to interactions between factors may also be required.

Thus, in order to analyze the effect of work status of a mother on infant mortality together with other explanatory variables the multivariate analysis was undertaken. To see the effect of work on infant mortality before analyzing the final model, two models are estimated whose results are shown in Table 5.

In the first model, work status of the mother is entered as a dummy variable that takes a value of 1 if the woman is working and 0 otherwise. In the second model, work is entered as a three-category variable capturing the type of work: professional/technical/clerical workers in one category; agricultural/manual workers in another; and not working at all.

Table 5 hazard ratio of infant mortality on mother's work status.

Explanatory variable	Mortality of infants			
	B	SE	Sig.	Exp(B)
Mother's work status				
Model 1: work status				
Not working				
Working	-.002	.114	.985	.998
Model 2: work status by type				
Not working				
Professional/technical/clerical	-.405	.167	.015	.667
Agricultural/manual	.174	.122	.153	1.190

The probability of an infant dying before its first birth day does not vary significantly by mother's work status (Table 5, model 1). However when mother's work is disaggregated by type it is found to be significantly affected. For infants whose mother's work is professional/technical/clerical, the hazard ratio of death are 33 percent lower than the hazard ratio for infants whose mother's are not working (Table 5, model 2). On the other hand the hazard ratio for infants whose mothers are working in the agricultural/manual sector is 19 percent higher than those not working.

Multivariate analysis assesses the strength of apparent association between work status of the mother and infant mortality by controlling other characteristics likely to influence the outcomes.

TABLE 6 Summary of the parameter estimate of the final model

	B	SE	Wald	df	Sig.	Exp(B)
WORK(not working)			11.350	2	.003	
Profess/technical/clerical	-.393	.168	5.464	1	.019	.675
Agri/manual	.157	.123	1.647	1	.199	1.170
EDURES(illiterate)	-.255	.121	4.467	1	.035	.775
MARITAL(married)	.640	.157	16.560	1	.000	1.896
PBINTERV(< 24 months)	-.451	.148	9.323	1	.002	.637
BIORDR(1st child)			26.410	2	.000	
2-4 children	-.713	.145	23.992	1	.000	.490
more than 5	-.622	.151	16.978	1	.000	.537

Table 6 presents Cox regression results as hazard ratio of the determinants of infant mortality. Estimates greater (less) than one represent higher (lower) risk of mortality relative to the reference category. In Table 6 we can observe that the relative risk of infant mortality is significantly higher among agricultural/manual workers than among non-working mothers, but lower among professional/technical/clerical workers than among non-working mothers. After controlling for numerous socio-economic and demographic factors such as age, education, birth order and previous birth interval, that are typically related to infant survival, the risk relative to that of agricultural/manual workers is 1.170 times higher than in the case of non-workers. Here the difference is only marginally significant. The risk of death is also significantly lower among professional/technical/clerical workers (32 percent) than in non-working mothers.

The relative risks associated with several other variables are statistically significant and in the expected direction. Among these factors the length of the preceding birth interval for infants exerted an expected beneficial effect on the hazard ratio of infants. The relative risk is 36 percent lower in those who have more than two years of birth interval than in those who have less than two years of birth interval.

There is also evidence of a strong detrimental effect of the birth order on the hazard ratio of infant death. As expected, being first born has a significantly high probability of dying of infants. Infant at birth order 2-4 have 51 percent lower risk of mortality than infants of birth order 1. Mothers with a birth order of 1 have a higher risk (46 percent) than mothers at birth order 5 or above.

The risk of infant death is 1.896 times higher in those who are not married than married, suggesting a critical role of fathers in providing financial assistance and emotional support for the mother. Among other variables higher mother's education will result in lower probability of infant death. For a literate mother the hazard ratio of an infant decreases by 22 percent.

The overall survival curves of infants for all factors in the final model are given in Figure1.

4.4 The effects of other predictor variables on infant mortality and their interaction with mother's work status

Here we evaluate the effects on infant mortality of the interaction of mother's work status with the following three variables: area of residence (urban, rural), age of the mother, education of the mother, birth order and previous birth interval. The interaction of work with each of these variables gives the effect of work on mortality separately for each of the different levels of the other variables in the interaction.

Five models are run with the relevant controls described above. The hazard ratios for the interactions are given in Table 7.

Table 7 Hazard ratios of infant mortality on mother's work status interacting with other variables.

	B	S.E	Wald	df	Sig.	Exp(B)
Interaction of work and area of residence						
Rural, mothers working	.245	.275	.790	1	.374	1.277
Urban, mothers working	-.051	.129	.162	1	.687	.951
Interaction of work and age.						
15-24, mothers working	.420	.194	4.663	1	.031	1.522
25-34, mothers working	-.083	.176	.223	1	.637	.920
35-49, mothers working	-.544	.247	4.861	1	.027	.581
Interaction of work and previous birth interval.						
<24 months, mothers working	.018	.250	.005	1	.943	1.018
1st birth or >24 months, mothers working	.005	.129	.002	1	.967	1.005

Interaction of work and birth order						
1 child, mothers working	.151	.207	.532	1	.466	1.163
2-4 children, mothers working	.150	.197	.583	1	.445	1.162
5 or more children, mothers working	-.249	.194	1.648	1	.199	.780
Interaction of work and education						
Illiterate, mothers working	-.011	.127	.008	1	.928	.989
Literate, mothers working	.012	.262	.002	1	.962	1.012

* **Not Working** is the reference category for all interactions.

While work status is not a significant determinant of the hazard of infants overall (model 1 table 5), it is an important factor in the age group 15-24 and 35-49 (Table 7). Indeed, the hazard ratio if the mother is aged 15-24 and working is 52 percent higher than non-working mother in the same age group. Also in the last age group mother's work status has an effect on infant mortality that is 42 percent lower risk. But the other category has no significant effect. The rest of the variables are seen to have no significant interaction with mother's work in influencing infant mortality.

4.5 Model adequacy

After a model has been fitted to an observed set of survival data the adequacy of the fitted model needs to be assessed. Indeed, the use of diagnostic procedures for model checking is an essential part of the model in process.

By comparing the value -2 Log L for the final proportional hazards model with that of a model in which all the coefficients are zero we found that the difference is significant. And also using the overall chi-square value (70.327), which is significant, we can reject the null hypothesis that all the parameters in the final proportional hazards model are zero.

Therefore we can say that there is no enough evidence to conclude that the final model does not show the best combination of the covariates.

The Cox regression model is applied using the assumption of proportional hazards for two or more groups. For checking this assumption the log-minus log plot in the final model for each variable is shown in Figure 4. From these plots for each variable the lines are parallel which indicates the validity of the assumptions. Plots of the martingale residuals against the value of the explanatory variables are given in Figure 5. For the assessment of the influential observations the martingale residuals versus the linear predictor (X' β) are also given in Figure 6.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

In the present study, using the proportional hazards model, it was determined whether there exists a relation between work status of mothers and infant mortality. The analysis is based on information on the last child born two years before the survey was carried out. In general, this study finds that working and nonworking mothers do not differ in terms of mortality of infants. But the type of work that a mother is engaged in has a significant effect on infant mortality.

Agricultural/manual work increases infant mortality while professional/technical/clerical reduces infant mortality. The multivariate analysis also supports this conclusion. For agricultural/manual workers the mortality value is 17 percent higher than non-workers. The risk for professional/technical/clerical is 32 percent lower than those of non-workers and this indicates that low-status work is associated with higher infant mortality.

Among other explanatory variables considered birth order and previous interval have the highest association with infant mortality. Here the relative risk for first birth is higher than the other birth orders.

Infant at birth order 2-4 have 51 percent lower risk of mortality than infants of birth order 1. Also birth order 1 has a 46 percent higher risk than birth order 5 or above. As expected, the relative risk for those who have more than two years of birth interval is 36 percent lower than those with birth interval less than two years. For literate women the relative risk is lower, so also married women.

The observation that high frequency of infant death occurred when mothers are engaged in agricultural/manual work does not in any way imply that this type of work should be discouraged. The higher mortality of infant for mothers who work in such low-status work reflects the fact that work for women is an additional duty and burden to their traditionally ascribed responsibilities.

The possible solutions for alleviating these maternal burden are expansion of good-quality child care facilities which require massive public action and greater involvement by others, including men, in child care and house work.

REFERENCES

Basu, A. M. and K. Basu (1991): Women's economic roles and child survival: The case of India. *Health Transition Review* 1(1): 83-103.

Bicego G. T. and J. T. Boerma (1991): Maternal education and child survival: A comparative analysis of DHS data. In *proceeding of the Demographic and Health Surveys World Conference,* pp. 177-294. IRD/Macro International, Inc. Washington, DC.

Caldwell, J. C. (1979): Education as a factor in mortality decline: An examination of Nigerian data. *Population studies* 3(3): 395-413.

Caldwell, J. C. (1986): Routes to low mortality in poor countries. *Population and Development Review* 12(2): 171-220.

Cleland, J. G. (1990): Maternal education and child survival: *Health Transition Series* 1(2): 400-19. Canberra, Australia: The Australian National University, Health Transition Center.

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

CSA (2001): 2000 *Ethiopia Demographic and Health Survey*. Addis Ababa.

Farah, A. and S. Preston. (1982): "Child mortality differentials in Sudan," in *Demographic Transition in Metropolitan Sudan.* Australian National University Press. Canberra.

Hobcraft, J. (1996): Women's education, child welfare, and child survival. In *population and women*. Proceedings of the United Nations Expert Group Meeting on Population and Women Gaborone, Botswana, 1992. New York: United Nations.

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

Kishor, S. (1992): *All 'Devis' but all not wanted: A district-level analysis of female discrimination in India*, 1961-81. Paper presented at the 1992 annual meeting of the Population Association of America, Denver, Colorado.

Kishor, S. (1993): 'May God give sons to all': Gender and child mortality in India. *American Sociological Review* 58(2): 247-65.

Koenig, M. A., Phillips, J. F., Campbell, O., and D'Souza, S. (1990): Birth intervals and childhood mortality in rural Bangladesh. *Demography,* 27(2), 251-265.

Kumar, S. (1977): *Role of the household economy in determining child nutrition at low income levels: A case-study of Kerala*. Occasional Paper No. 95. Ithaca, New York: Department of Agricultural Economics, Cornell University.

Manderson, L. (1982): Infant feeding practice, market expansion and the patterning of choice in South East Asia. *New Doctor* 26: 27-32.

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

Mencher, J. P. (1988): Women's work and poverty: Women's contribution to household maintenance in south India. In Daisy Dwyer and Judith Bruce, eds. *A home divided: Women and income in the Third World*. Stanford University Press. Stanford, California.

Murthi, M. A. C. Guio and J. Dreze. (1995): Mortality, fertility and gender bias in India: A district-level analysis. *Population and Development Review* 21(4): 745-82.

Rogers, B. (1980): *The Domestication of women: Discrimination in Developing Societies*. St. Martin's Press. New York.

Fig. 1 The overall Survival Function

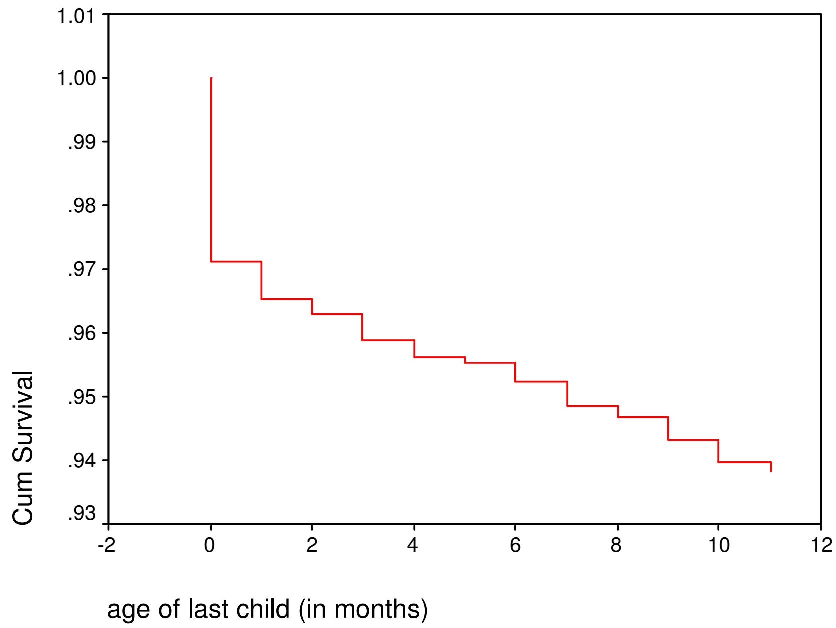


Fig. 2 The Hazard Function

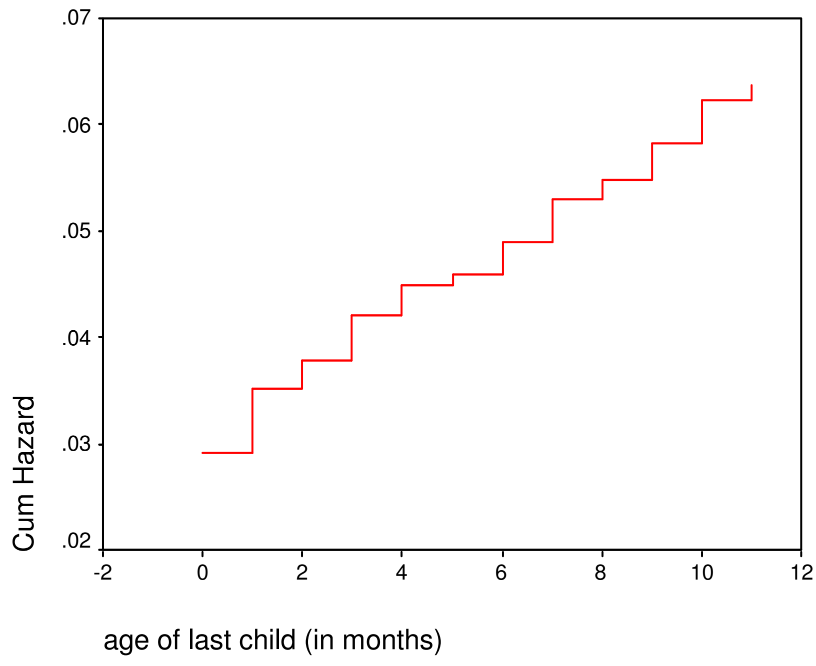
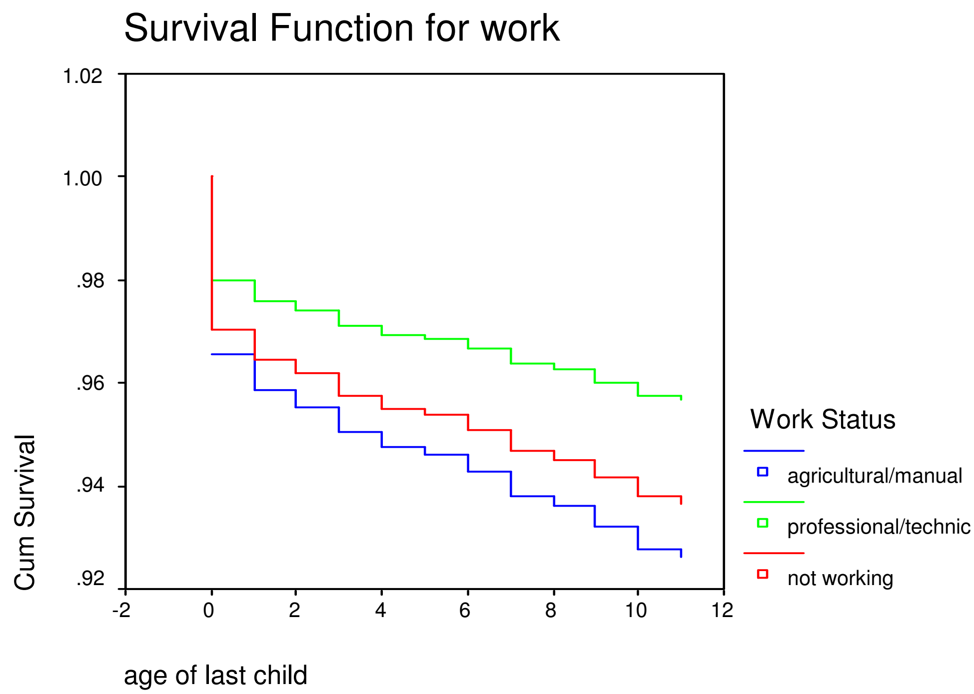
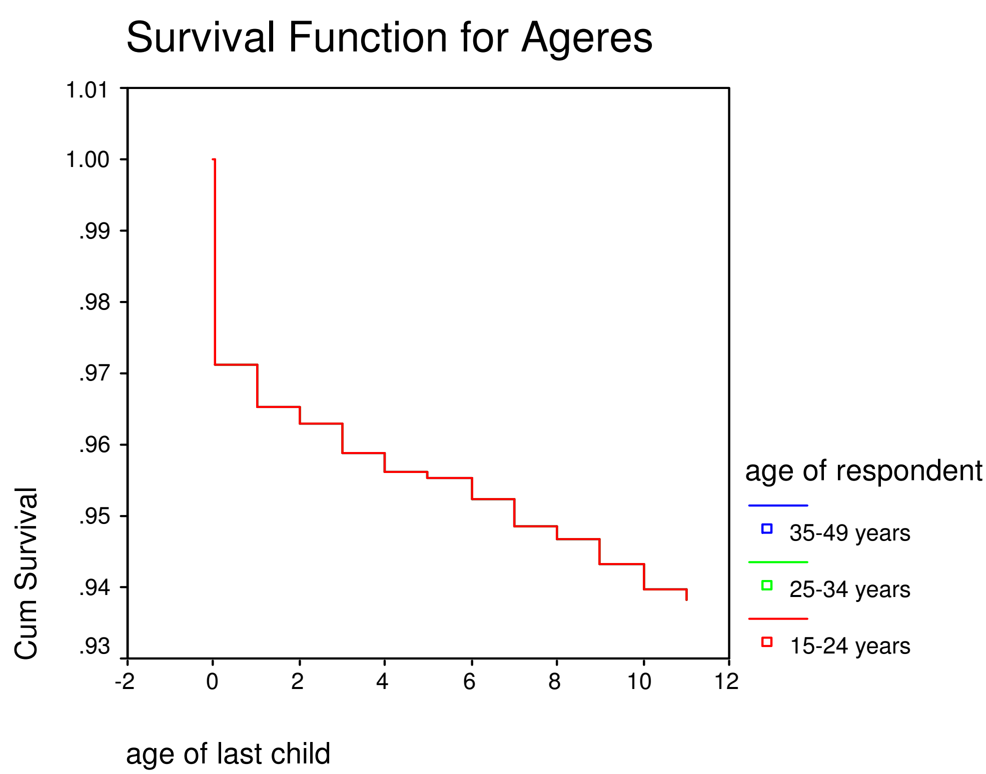
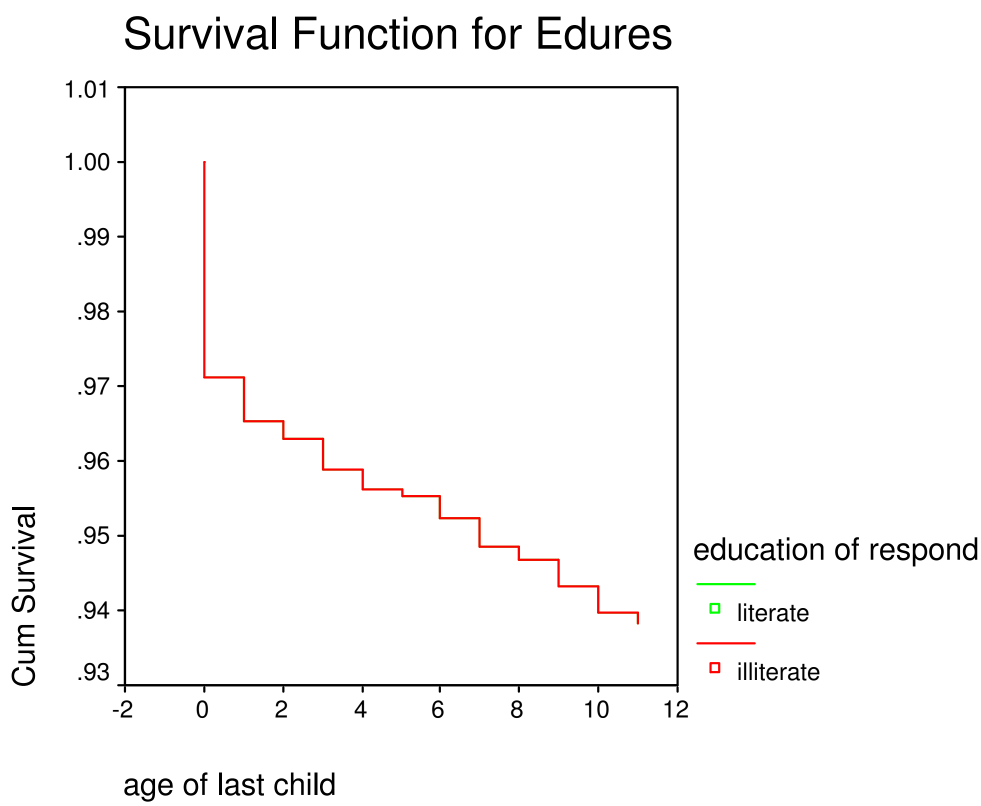
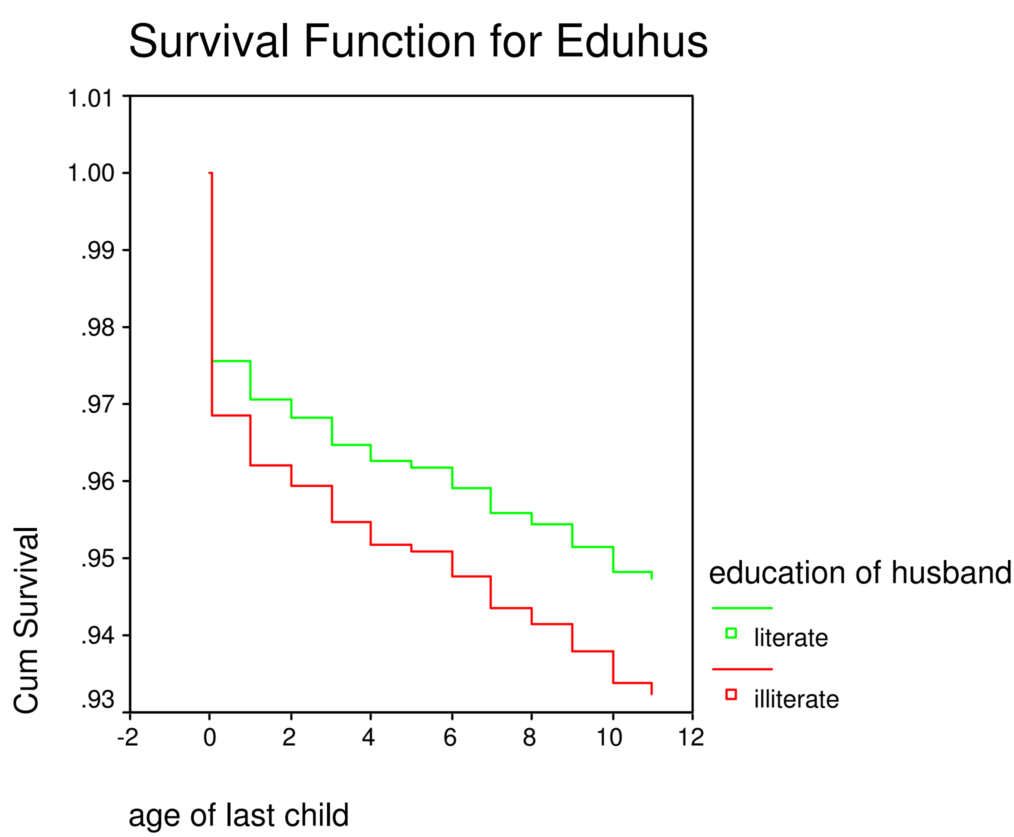


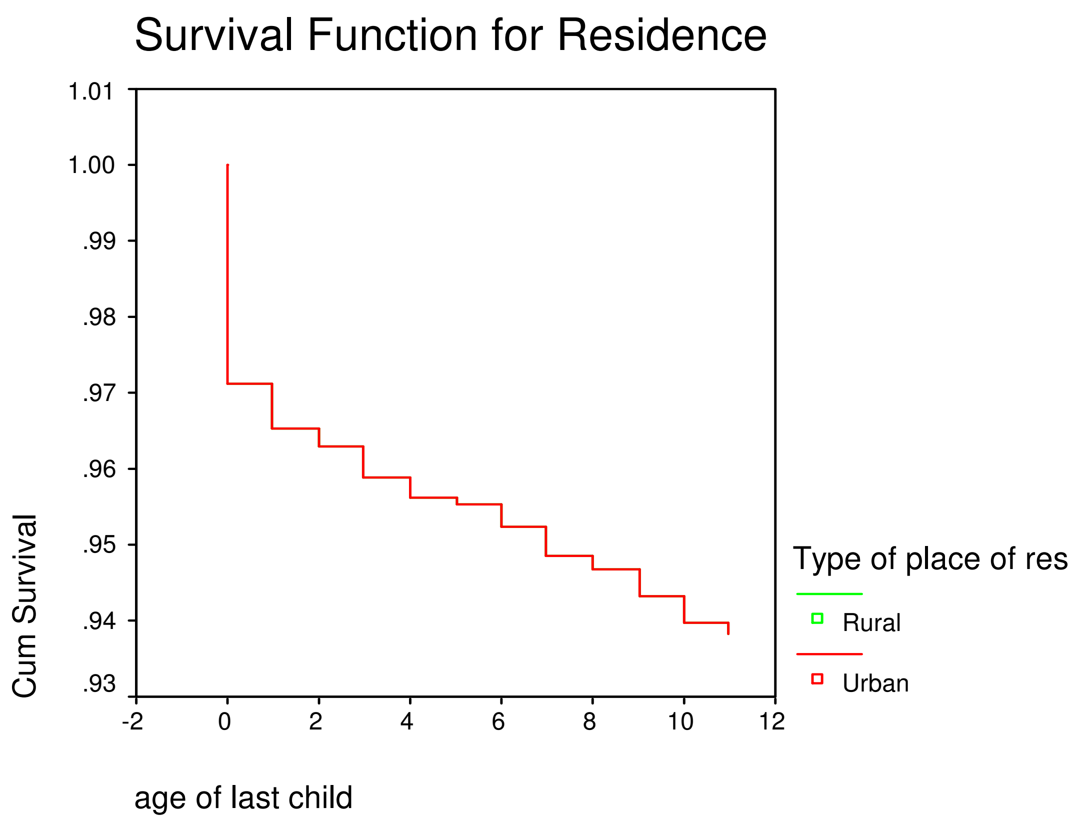
Fig. 3 Survival functions of covariates



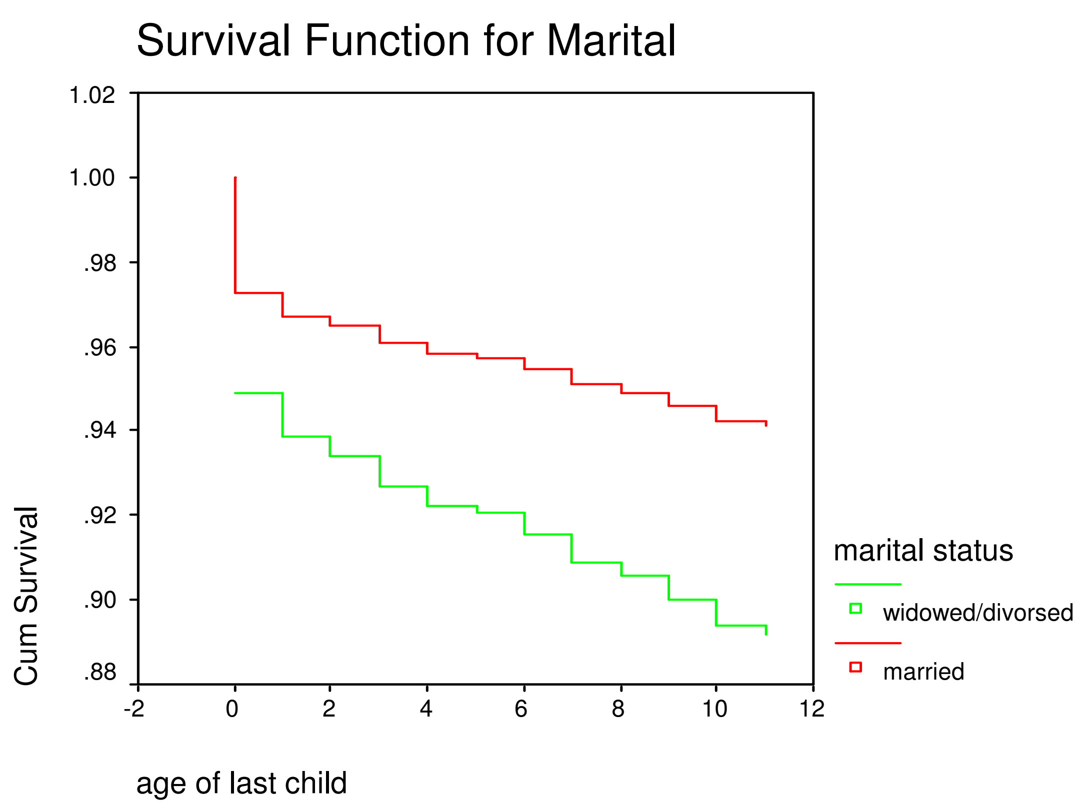



Survival Function for Edures
1.01
1.00
.99
.98
.97
.96
.95
.94
.93
Cum Survival
-2
0
2
4
6
8
10
12
age of last child
education of respond
literate
illiterate

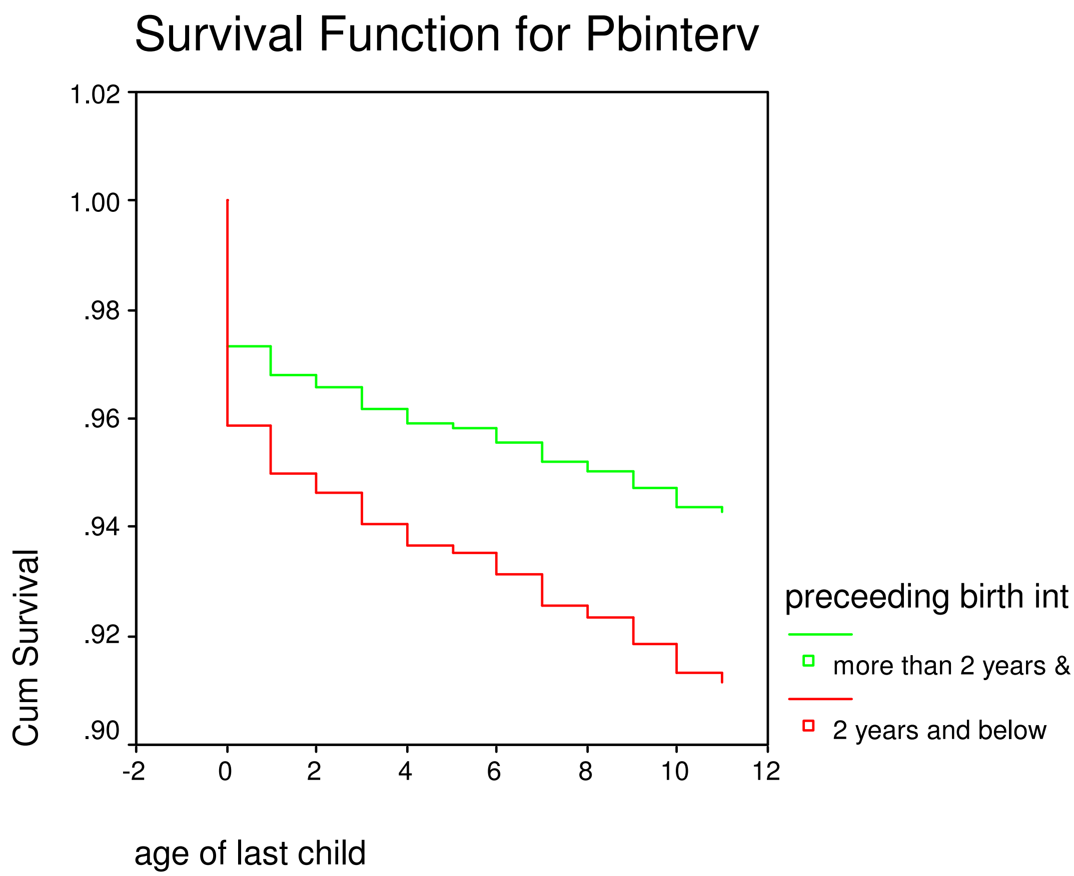

Survival Function for Eduhus
1.01
1.00
.99
.98
.97
.96
.95
.94
.93
Cum Survival
-2
0
2
4
6
8
10
12
age of last child
education of husband
literate
illiterate



Survival Function for Residence
Cum Survival
1.01
1.00
.99
.98
.97
.96
.95
.94
.93
-2
0
2
4
6
8
10
12
age of last child
Type of place of res
Rural
Urban



Survival Function for Marital
Cum Survival
1.02
1.00
.98
.96
.94
.92
.90
.88
-2
0
2
4
6
8
10
12
age of last child
marital status
widowed/divorsed
married


Survival Function for Pbinterv
Cum Survival
1.02
1.00
.98
.96
.94
.92
.90
-2
0
2
4
6
8
10
12
age of last child
preceeding birth int
more than 2 years &
2 years and below

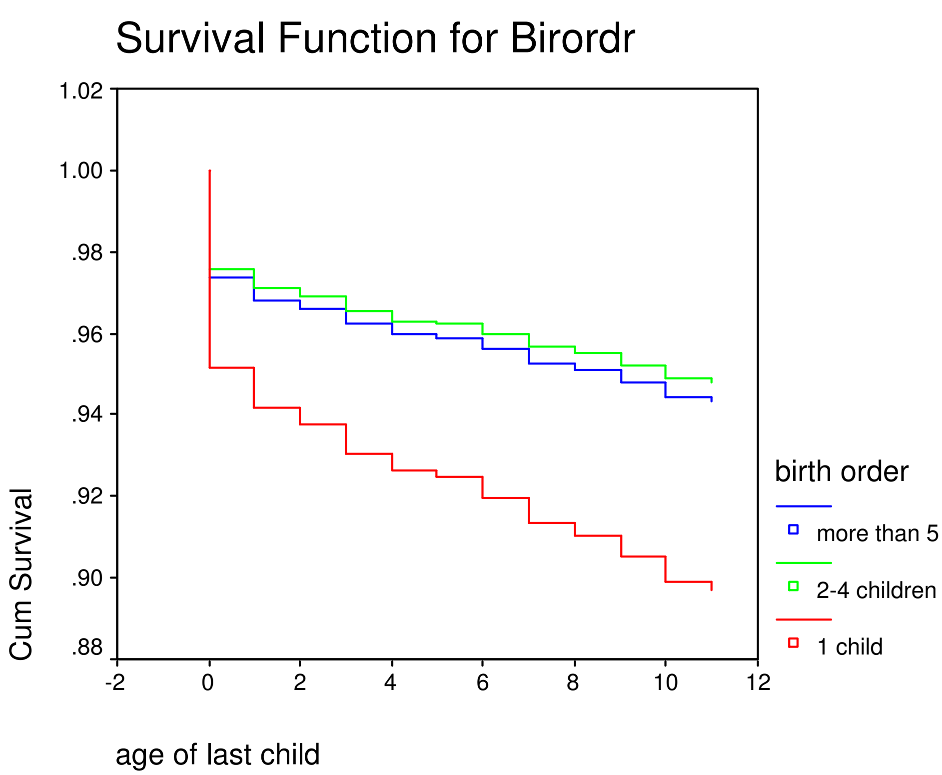
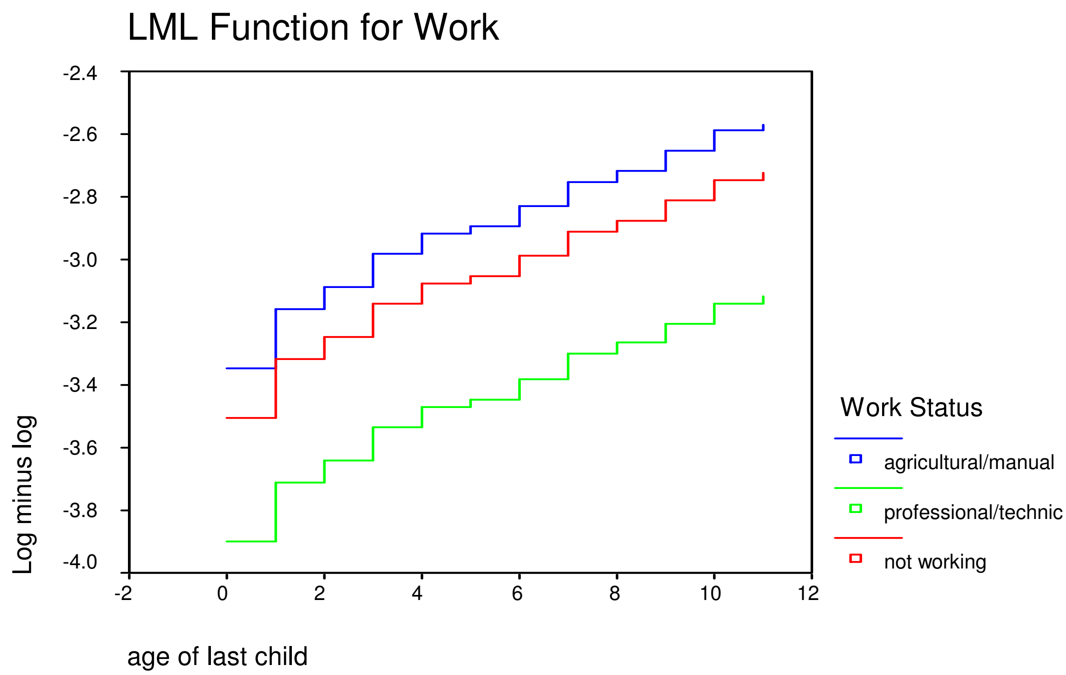
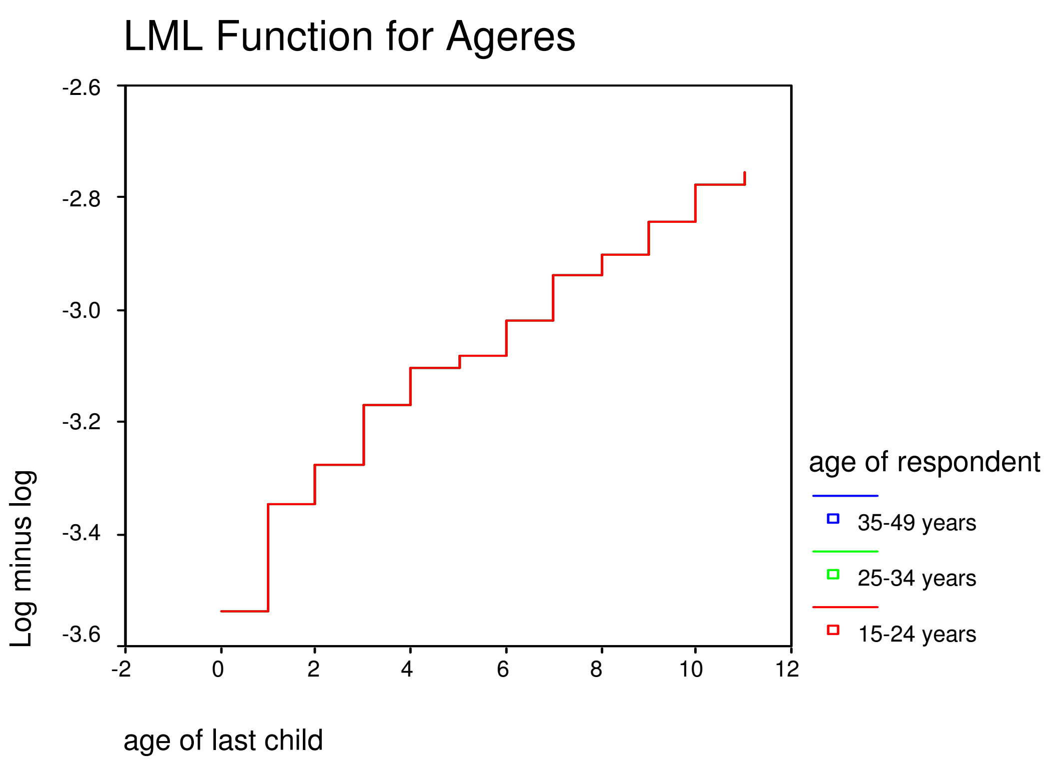
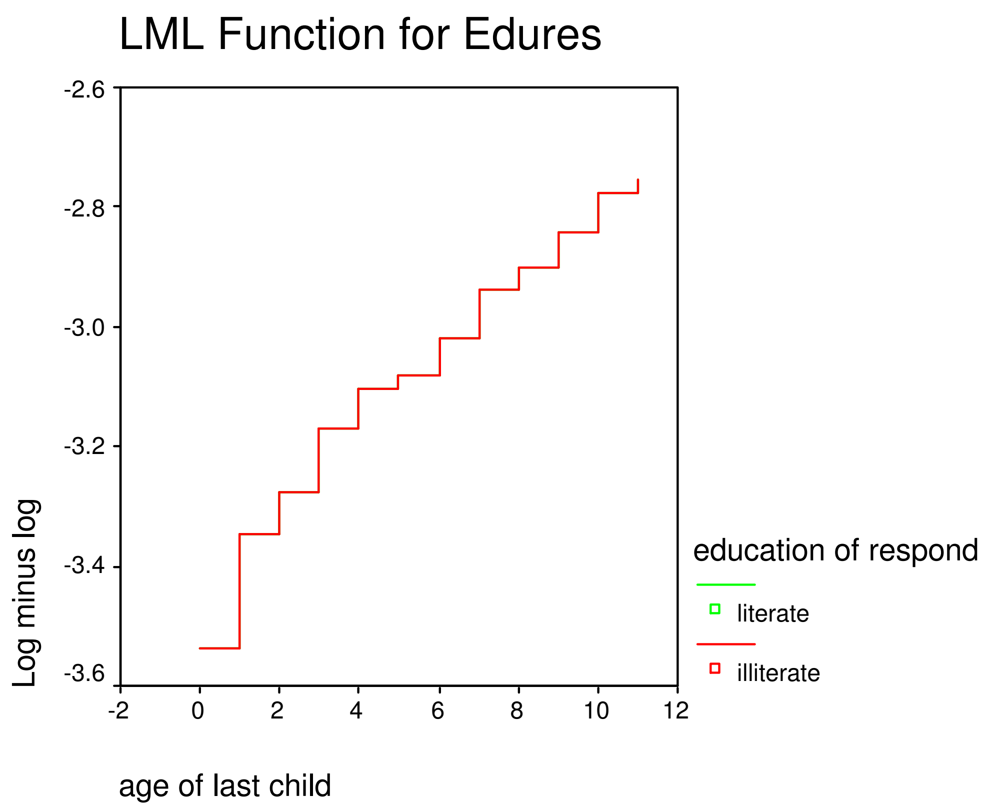

Survival Function for Birordr
Cum Survival
1.02
1.00
.98
.96
.94
.92
.90
.88
-2
0
2
4
6
8
10
12
age of last child
birth order
more than 5
2-4 children
1 child

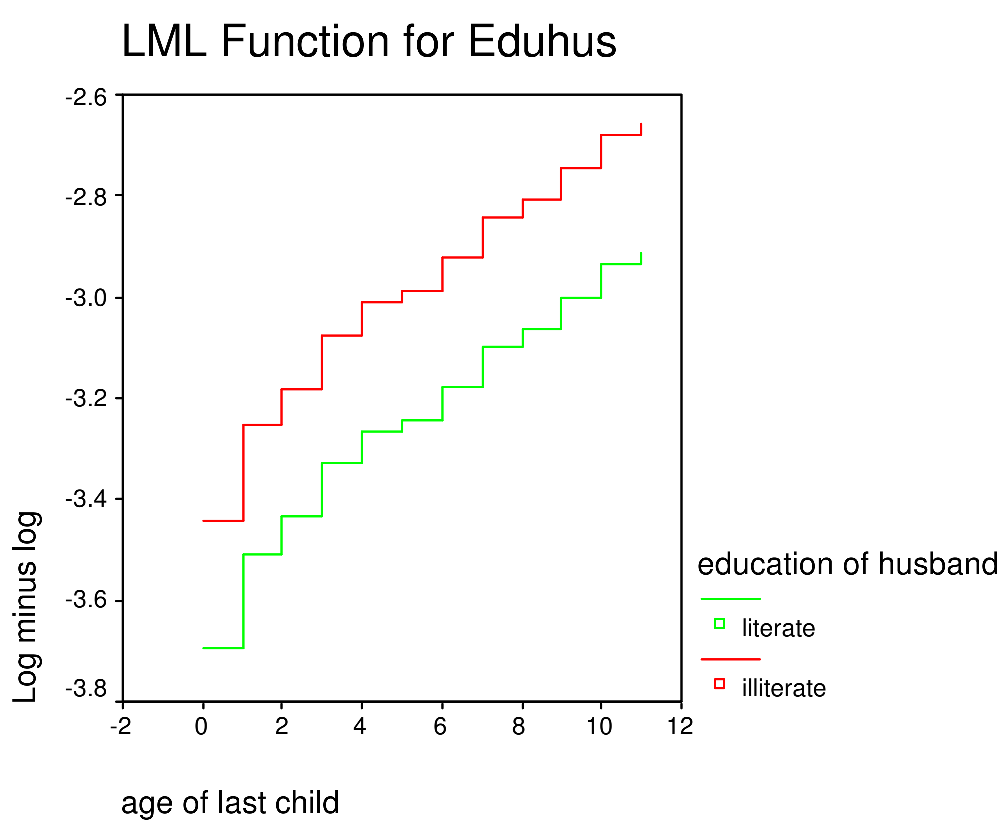
Fig. 4 LML functions of covariates



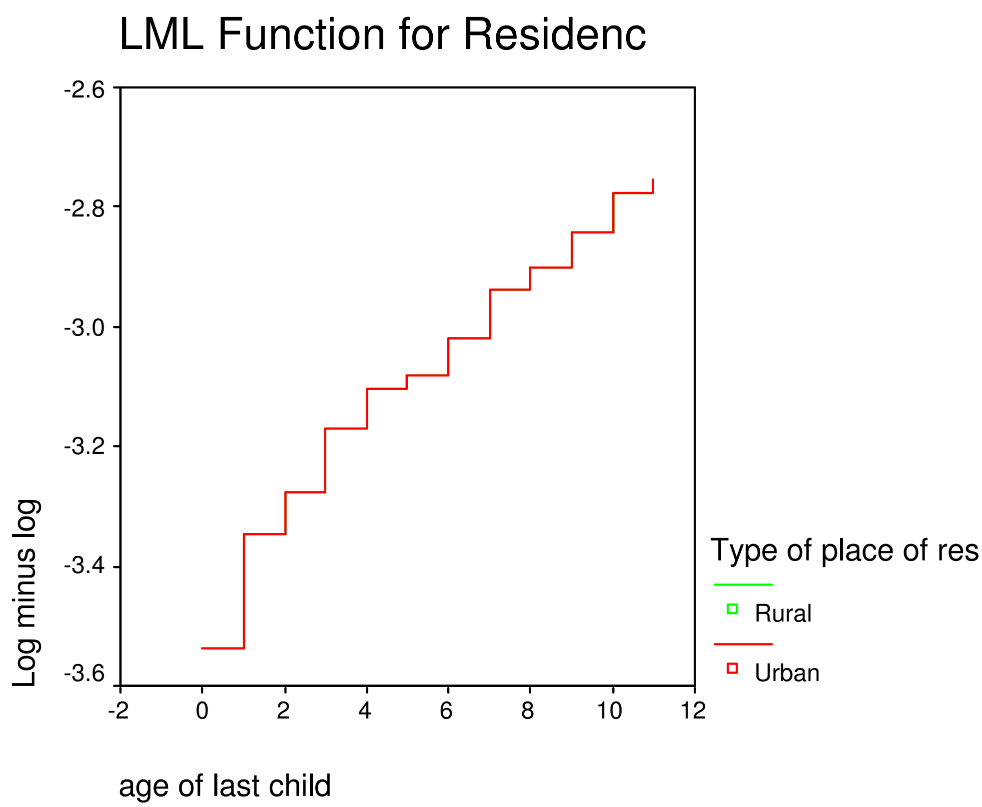


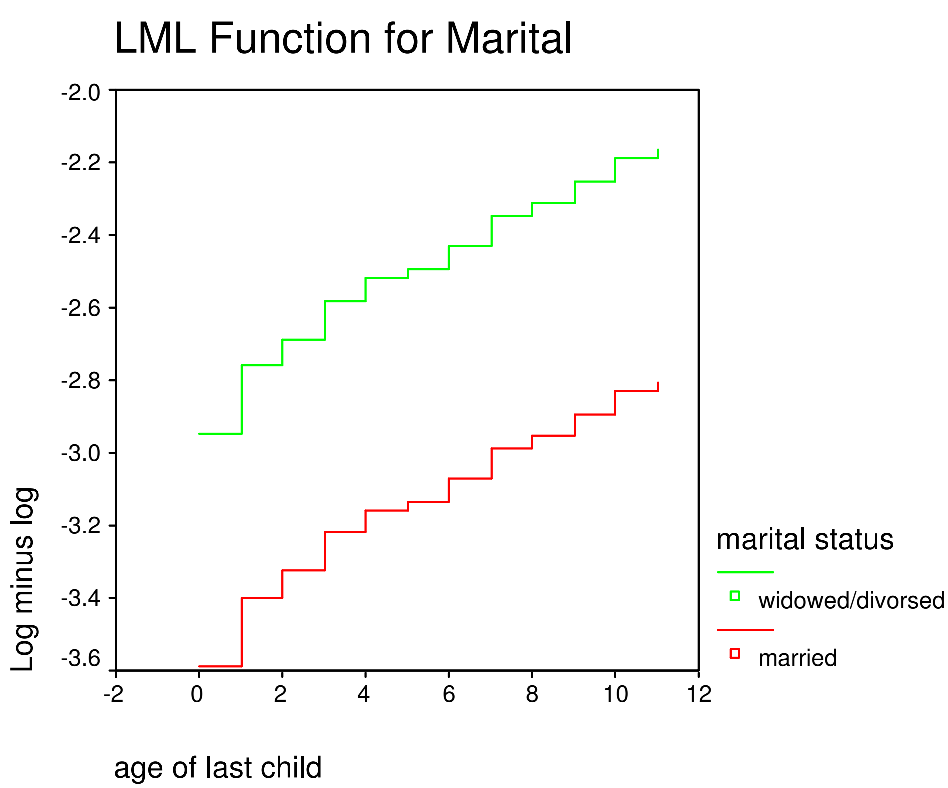


LML Function for Edures
-2.6
-2.8
-3.0
-3.2
-3.4
-3.6
Log minus log
-2
0
2
4
6
8
10
12
age of last child
education of respond
literate
illiterate

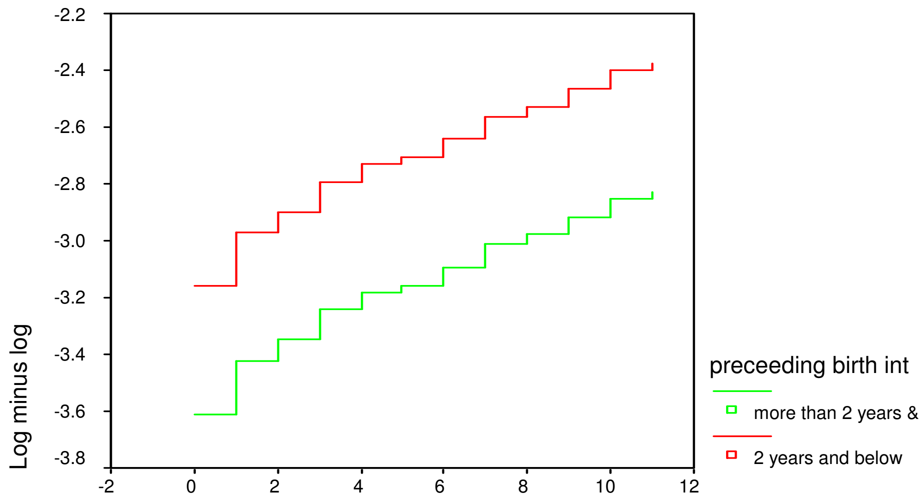


LML Function for Eduhus
-2.6
-2.8
-3.0
-3.2
-3.4
-3.6
-3.8
Log minus log
-2
0
2
4
6
8
10
12
age of last child
education of husband
literate
illiterate


LML Function for Residenc
Log minus log
-2.6
-2.8
-3.0
-3.2
-3.4
-3.6
-2
0
2
4
6
8
10
12
age of last child
Type of place of res
Rural
Urban

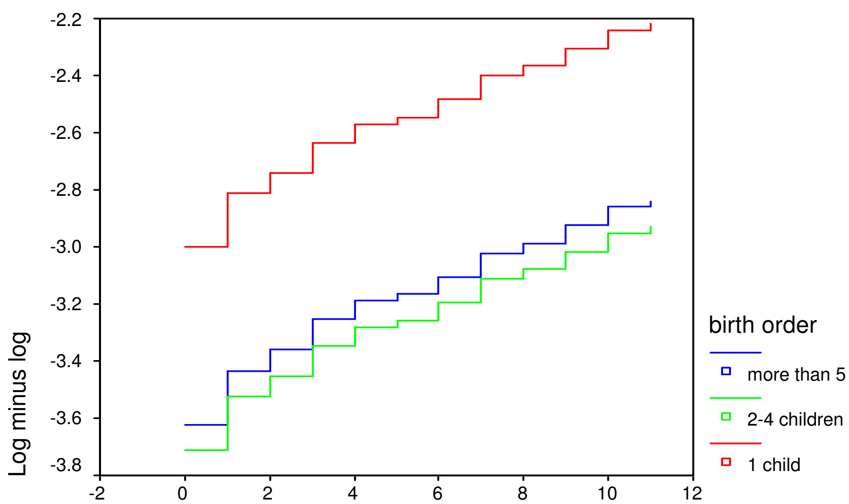

LML Function for Marital
Log minus log
-2.0
-2.2
-2.4
-2.6
-2.8
-3.0
-3.2
-3.4
-3.6
-2
0
2
4
6
8
10
12
age of last child
marital status
widowed/divorsed
married

LML Function for Pbinterv



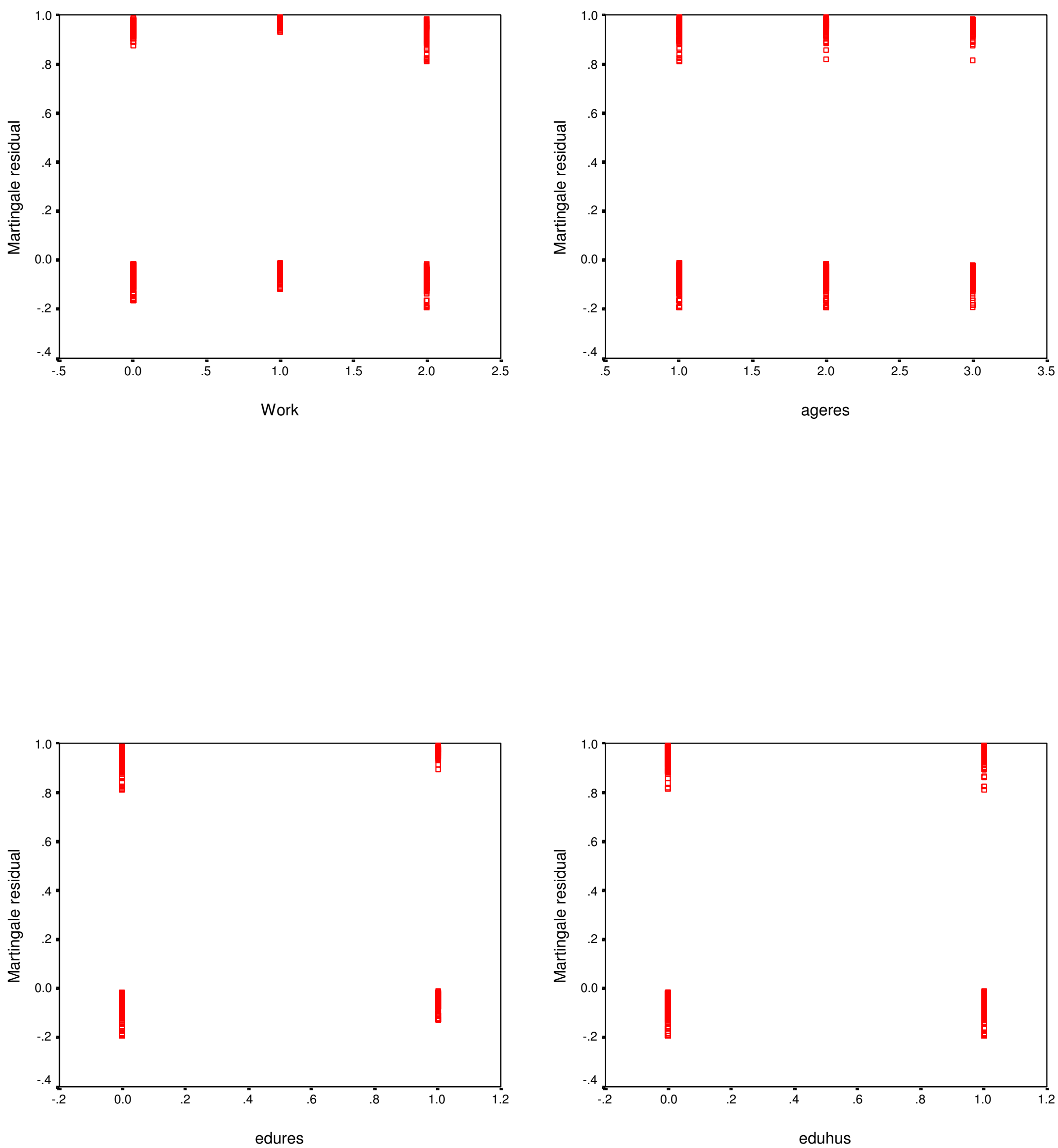
age of last child

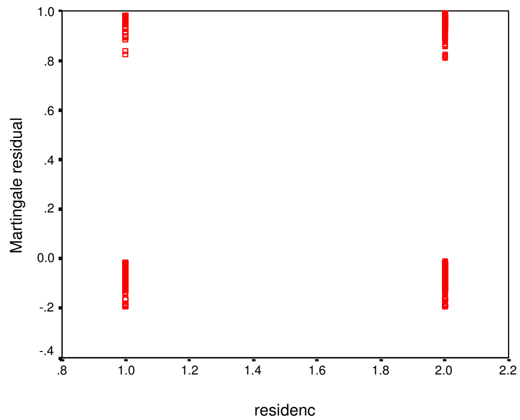
LML Function for Birordr

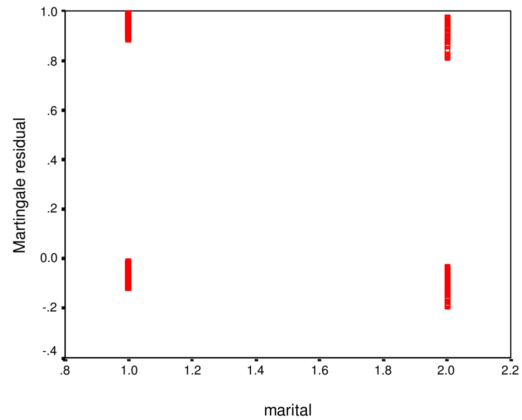


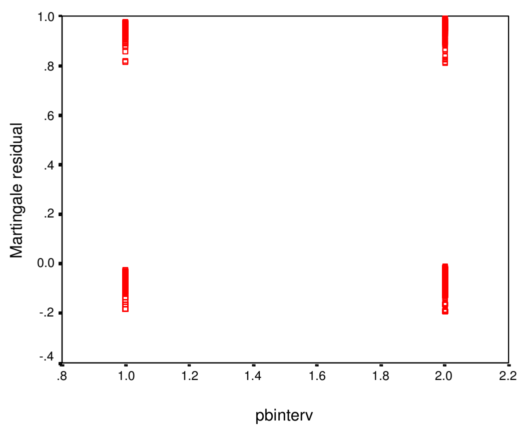
age of last child

Fig. 5. Martingale residuals against covariates




Martingale residual
1.0
.8
.6
.4
.2
0.0
-.2
-.4
.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
residenc


Martingale residual
1.0
.8
.6
.4
.2
0.0
-.2
-.4
.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
marital


Martingale residual
1.0
.8
.6
.4
.2
0.0
-.2
-.4
.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
pbinterv

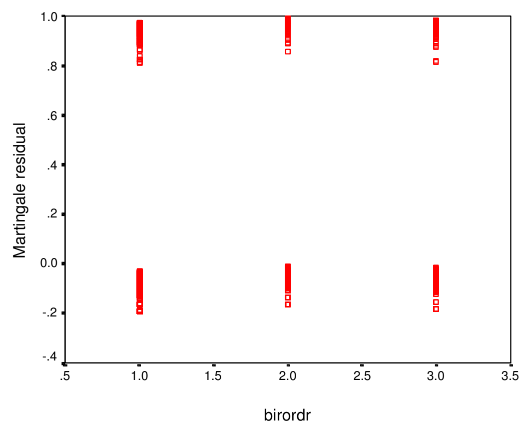

Martingale residual
1.0
.8
.6
.4
.2
0.0
-.2
-.4
.5
1.0
1.5
2.0
2.5
3.0
3.5
birordr

Fig. 6 plot of martingale residuals against X'Beta