Determinants of Some Socio-demographic Characteristics on Fertility in Ethiopia

Kidus Yosef

A thesis Submitted to

The Department of Statistics

Presented in Partial Fulfillment of the Requirements for the Degree of Master of Science in Statistics (Applied Statistics)

Addis Ababa University

Addis Ababa, Ethiopia

June 2012

Addis Ababa University

School of Graduate Studies

This is to certify that the thesis prepared by Kidus Yosef, entitled: ***Determinants of Some Socio-Demographic Characteristics on Fertility in Ethiopia*** and submitted in partial fulfillment of the requirements for the degree of Master of Science complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by Examining Committee:

Examiner	Signature	Date
Examiner	Signature	Date
Advisor	Signature	Date

Chair of Department or Graduate Program Coordinator

ABSTRACT

Determinants of some socio-demographic characteristics that affect fertility in Ethiopia

Kidus Yosef

Addis Ababa University, 2012

Ethiopia is the second most populous country in Africa next to Nigeria. According to Population and Housing Census (2007) the population Ethiopia was 77.1 million. The annual population growth rate was predicted at 2.7 percent showing that the country has a very high total fertility rate. Given this background the major objective of this study is to assess the effect of socio-demographic factors on fertility level in Ethiopia. In this study the data source is the Ethiopian Demographic and Health Survey conducted in 2005(EDHS 2005) by the Central Statistical Agency (CSA) with a total of 14,070 women of age 15-49 years. In this study 10,199 women who have at least one child are considered. Descriptive statistics and binary logistic regression are used for statistical analysis. The study found out that place of residence, level of education of a wife, religion, ever use of contraceptive, husband level of education, age 5-year group, age at first marriage and the work status of wife were significant determinants of total number of children ever-born. Socio-demographic characteristics of women such as poor educational status of wife, rural place of residence, age at first marriage below 18 years, women not working (not employed as wage/salary earner), low level of educational husband were found to have significant association with the risk of having five or more CEB.

Acknowledgement

First of all I would like to thank the almighty God. I want to express my sincere gratitude to my Advisor, Prof. Eshetu Wencheko, for his guidance, valuable advice and supervision for preparing my thesis.

Finally, I am grateful to those individuals who contribute to the completion of this study. My special thanks to my family and my friends who provide support and encourage me during this study

List of tables

Acronym

CSA	Central Statistical Agency
NSS	National Sample Survey
DS	Demographic Survey
EDHS	Ethiopian Demographic and Health Survey
NDHS	Nepal Demographic and Health Survey
CEB	Children Ever-Born
TFR	Total Fertility Rate
FFS	Family Fertility Survey
OR	Odds Ratio

Table of contents

CHAPTER ONE

INTRODUCTION

1.1 Background of the study

Fertility is one of the elements in population dynamics that has a significant contribution towards changing population size and structure over time. Fertility may be defined as the actual reproductive performance of a woman. "Fertility rate" is the number of children born per couple, person or population. Fertility rate is highest in sub-sharan Africa than any parts of the world, mainly due to strong kinship networks and high economic and social values attached to children (Caldwell and Caldwell, 1987). According to Bongaarts (1999) the desired family size is more than four children in sub-Saharan African countries, where child mortality is high and poverty is rampant.

Serious concerns have been expressed regarding population growth and its impact on human welfare. It is estimated that, unless there is reduction in fertility rate, the world population will cross seven billion by the year 2020. In the last decade of the twentieth century, the countries of sub-Saharan Africa had the highest fertility rate and natural increase in the world. Data from census and survey based on a case study in southern Ethiopia indicate fertility rate mostly between six and seven children per women with a few exception on either side of this range (Samson and Mulugeta 2009).

Ethiopia is one of the most populous countries in Africa with a population of 77.1 million in mid-2007 and an annual population growth rate of 2.7% (2007 population and housing census). Evidence show that nearly 2 million people are added to the country population each year. According to a recent projection, the population will be doubled to the size of

135 million by the year 2030. The country is characterized by very high fertility, low life expectancy, high maternal and child mortality, poor nutritional status, high infant mortality and low per capital income.

Though the recent two nationwide surveys witnessed fertility decline, according to seven different surveys carried out in the last 40 years, there is an erratic trend of fertility in Ethiopia. Total fertility rate is a common measure of current fertility and defined as the total number of birth a woman would have by the end of her childbearing period if she were to pass through those years bearing children at the observed age-specific fertility rates. Age-specific fertility rate expressed as the number of birth per thousand women in a specified age group, are calculated by dividing the number of life birth to women in a specific age group by the number of woman years lived in that age group. According to national sample surveys (NSS) Total Fertility Rate (TFR) was 6.7 and 5.8 in 1967 and 1971. However, the 1981 demographic survey (DS) reported TFR of 8.8, the 1990 Family and Fertility survey (FFS) and the 1994 population census reported TFR of 6.6 and 6.9, respectively. However, the recent two demographic and health surveys DHS 2000 and DHS 2005, reported a decline in fertility, TFR of 5.9 and 5.4, respectively.

1.2 Statement of the problem

Fertility is an important component of population dynamic and it plays a major role in changing the size and structure of a given population. It is also a complex variable affected by multiple factors. Most studies in developing countries have put emphasis on the role of demographic, socio-economic, and cultural factors in determining difference in fertility and also the impact of this factor on fertility. The specific nature of the determinant of fertility is intricate and complex. While, fertility behavior influences population growth, which has consequences on resources, employment situation, health and other social facilities and saving and investment, in turn, such consequences have great bearing on socio-economic variables that affect fertility behavior. Socio-economic factors are in turn affected by demographic factors. These factors directly and indirectly affect the interaction process. Socio-economic variables may not directly influence fertility but they may influence fertility indirectly (Samson and Mulugeta 2009).

The mechanism of factors affecting fertility is that intermediate variables influence fertility directly, while socio-economic and environmental variables affect fertility indirectly through intermediate variables (Bongaarts 1978). Some of these factors could be literacy status, occupation, religion, child birth-death ratio, wealth status, place of residence, household headship, contraceptive use, region, reproductive life span, age at first marriage and desired number of children. This study attempts to examine the effects of some selected factors that influence fertility in the case of Ethiopia.

1.3 Objective of the study

General objective

The general objective of this study is to identify socio-demographic factors on fertility level in Ethiopia.

Specific objectives

- ✓ To find the relationship between number of children ever born and some selected socio-demographic variables.
- ✓ To identify the variable's that significantly affect fertility.

1.4 Significance of the study

It is hoped that the findings from this research could be useful in many ways. The findings are believed to be useful for policy making, monitoring and evaluation activities of the government and different concerned agencies.

Since the study is an attempt to identify socio-demographic factors that affect fertility, the end user governmental and non-governmental organizations could take intervention measures and set appropriate plans to tackle fertility problems.

This study is expected to contribute its part by filling the information gap concerning the impact of factors that affect fertility in the country. Finally, the study could be useful as a stepping stone for further studies.

1.5 Limitation of the study

Although many factors affecting fertility and the impact of socio-demographic factor that affect fertility as indicated by different studies in different countries. This study is undertaken to explore a few socio-demographic factor that affect fertility in Ethiopia. These is due to

- ✓ Data problem - Some important variables are not included because of missing values and non-responses.
- ✓ The data, used in this study are from the EDHS 2004/05. Thus, the results may not necessarily reflect the current situation of Ethiopia.

CHAPTER TWO

LITERATURE REVIEW

The various causal factors affecting fertility have been reviewed from various literatures and discussed below.

2.1 General

Fertility is the natural capability of giving life. Fertility is one of the three major components of population dynamics that determine the size and structure of the population of a country. Differential in fertility behavior and fertility levels in different areas and among population strata or characteristics have been among the most pervasive finding in demography (Ramesh 2010). It is also a complex variable affected by multiple factors. According to Bongaarts (1978), factors affecting fertility are broadly classified into proximate (direct) and distal (indirect) factors. The proximal factors are biobehavioral factors, like being sexually active, use of contraceptive, duration of postpartum infecundability, abortion and sterilizing which affect fertility directly, whereas, distal determinant are socio-cultural factors which affect fertility indirectly through affecting the bio-behavioral factors.

Human fertility is responsible for the biological replacement and maintenance of the human species. In fact, fertility is a major counteracting force to population attrition from mortality and therefore, has a significant expansionary force in population dynamics. However, the phase of actual reproductive performance is counted in terms of the physiological potential of a woman to conceive and bear children. This phase is termed as the fecund period, which has two extremes, viz., menarche and menopause. In demographic studies, the reproductive span i.e., the child-bearing period of women is usually taken as between 15 to 49 years of age. Thus, a fecund woman may or may not be fertile but a fertile woman must be fecund. The main events or phenomena associated with fertility are age at menarche and age at menopause.

It is found that a later age at marriage reduces fertility. Educational level, economic status, religious attitudes, women's work participation etc. are other factors affecting fertility in addition to contraception control practice and attitudes (Samson and Mulugeta 2009).

2.2 Review of variables that influence fertility

Age at first marriage

Marriage is one of the proximate determinants of fertility (Bongaarts 1978). It affects fertility via frequent and regular exposure to sexual relations and the age at entering in to martial life. Given the fact that fertility often takes place within marriage, there is an inverse relationship between age at first marriage and fertility. This is because age at first marriage determines the length of exposure to the risk of becoming pregnant and the actual commencement of the process of child bearing (Islam2009).

Socio-economic development, particularly improvement in women's education, provision of increased employment opportunities for women and raising their status can contribute to higher age at marriage. World fertility survey data show, however, that postponement of marriage by a few years in many developing societies where age at marriage is very low has little reduction effect on fertility. It appears that in most countries only where age at marriage is raised substantially above the level of around age 21 is associated with lower fertility. Even above that age the impact of higher age at marriage is not always clear. Women who marry late tend to be better educated, have a higher social status and practice contraception to a great extent than women marrying younger (Abdul Hakim 1994).

Education

There are numerous channels through which education can affect fertility. It is contended that education provides individuals with a new vision and normative orientation, better health care, better employment opportunities outside home, better knowledge of and access to family planning method these in turn, may produce a depressing effect on fertility (Cochrane 1979). Educational attainment alters parent's perceptions of the advantages of small families brings changes in the status of women, changes the social and economic aspirations, and affect attitude towards contraception and ability to understand and make use of particular methods. The empirical findings of a number of studies have supported such an inverse relationship between education and fertility.

Cochrane (1979) has shown that the aggregate level of education in the community affects the level of fertility, and she also concluded that an inverse relation is much more consistent for female than male education. Caldwell (1980; 227-249) considered that the greatest impact of education is not direct but through the restructuring of family relationships and family economies and the effect of these on the direction of net wealth flows between generations. Fundamentally, schooling attacks the traditional family economic structure by weakening the authority of the old over the young and of the male over the female. Elaborating on this theme, Caldwell (1982) maintains that intergenerational cost and benefit shifts occur when parents begin to invest more in the education of their children than children contribute to the family through child labor and care for aged parents. This reversal in the intergenerational flow of wealth from that of children to parents to that of parents to children encourages couples to opt for smaller families.

Of all the social factors that have been widely studied for their impact on fertility, women's educational attainment is the one that has proved most consistently and strongly related to fertility. Education provides women with some of the basic skills and confidence to begin their lives. It gives opportunities for participation in the outside world. It influences attitudes and perceptions. It gives openness to ideas. Education gives rise to liberal attitudes towards the sex of the child (Sharma, 1998).

Work status of women

Several theoretical frameworks explain the relationship between women's employment status and fertility behavior. Among them, the most influential are the women's incompatible-role hypothesis and the women's opportunity cost hypothesis. They predict that women's improved employment status will lead them to desire and have fewer children. Thus, the actual and desired number of children of women with formal employment should be lower than those of women with no work or with only informal work. The rationale for these two theoretical frame works are as follows (Mason and Palan 1981).

The women's incompatible role hypothesis argues that women play many important roles within the family and within the society. Two of these, mother (or wife) and income earner are important, and are usually incompatible. These are two kinds of incompatibility in women's roles; "social norm" incompatibility and "time" incompatibility. In some cultures and societies, social customs and norms encourage women to stay at home and to take care of their husbands and children. There is also a conflict between women's work and family responsibilities. The extent of incompatibility is closely related to the type of women's employment. This framework predicts that, because of these incompatible roles, women who work will usually choose a smaller family size. The more incompatible the roles are the stronger is this relationship (Mason and Palan 1981).

The women's opportunity cost hypothesis originates from a rational choice model. The two key variables in the rational choice model are preference, which is described by the households utility function, and opportunities, which are described by the households resources (mainly income and time) the price of market goods. There are two main kinds of rational choice models; Becker's "new home economics model "where household preference and Easterlin's "relative income model", where preference is indigenous (Easterlin, 197). In the rational choice model, the household makes the best choice based on both the constraint line and the utility function. So, any changes in the price of market goods, time and household income, or household preference will alter the household's decisions about production, consumption and childbearing. The rational choice model predicts that, because of their increasing time value or opportunity cost, women with a higher wage rate is likely to have fewer children.

Using these two frameworks, many empirical studies have been carried out to investigate whether there exists a negative relationship between women's employment status and fertility. However, the results of these studies have not been consistent. In developed countries many studies have found a consistent negative relationship between women's employment and fertility. While some studies in developing countries have found no relationship or a positive relationship (Stycos and Weller, 1967), other have a negative relationship between women's employment and fertility, particularly in the formal work sector (Okpala,1989).

Place of residence

The urban-rural differential in fertility is one of the most widely studied areas in fertility. It is generally expected that, in societies undergoing socio-economic, cultural, and demographic changes, fertility will be lower in urban than in rural areas due to the impact of a host of factors. The lower fertility in urban areas may be viewed from two perspectives. Firstly, at the aggregate (macro) level of analysis, urban fertility may be lower because the urban population has a larger proportion of white collar jobs and who are thus likely to marry late and have smaller families. Secondly, at the individual (micro) level of analysis, one may hypothesize that the same level of education and income will produce lower fertility in urban setting. Moreover, the urban population contains a higher proportion of those who are expected to adopt new ideas and new life styles first and of those for whom a large family does not represent an economic benefit but rather a burden (Abdul Hakim 1994).

Ethiopia, being one of the developing countries where subsistence agriculture is the major economic activity, families often prefers larger number of children since they are considered as economic asset rather than liabilities. In rural areas parents want to have a large number of children to get assistance in farming activities and emotional as well as economic support during old ages. According to Caldwell (1982), the economic importance of children is over lifetime. African children do not only provide support during childhood and adolescent ages but also beyond these ages. More psycho-social and economic support is expected when parents are getting older. Old-age security is one of the major motivational forces for having as many children as possible in Africa. In traditional societies, children are also expected to strengthen the extent of kin relations, which implies not only economic benefits but also physical protection. Getting larger in number is tantamount to strength in physical security. Like many countries in sub-Saharan Africa, traditional norms and values in Ethiopia are in favor of high fertility. Having many children is considered as a virtue and respect of God in a number of Ethiopian rural communities (Desta and Seyoum, 1998).

Contraception use

Family planning plays a pivotal role in checking the global population crises. According to United Nations estimates, family planning accounted for two-thirds of the decline in total fertility rate (TFR) in the developing world from 1960-65 to 1980-85.

The DHS 2000 report indicates that there is a high total fertility (5.9) and population growth rate (2.9%) in Ethiopia. In 2005 the Ethiopian population is estimated to be over 73 million. Many studies have been conducted in order to uncover obstacles of family planning acceptance and its continued use. Historically, however, women were typically the respondents of demographic research and most of this study recurrently documented various socio-demographic, fertility and other cognitive and affective attributes of women as determinants of contraception use since a long time ago. However, the socially defined gender roles of men and women gauge the power balance between the two sexes (Amaha Haile and Fikre Enqueslassie 2006).

In developing countries most communities afford inferior positions to women. In effect women are either under collective decision-making with their partners or completely rely on the male partner's decision on issues that affect their reproductive live. Many of the studies conducted in this area also shown spousal influence on each other's fertility preference and reproductive decision-making. Despite this fact, until recent years men's role in couple's fertility decision-making was ignored. However since the past few years demographic studies examined the role of men in family planning and many of them showed the importance of involving husbands for couples' family planning adoption (Amaha Haile and Fikre Enqueslassie 2006).

Besides men's involvement, improving the status of women was underlined in the past few years as one of the key strategies to ensure the sexual and reproductive health of men and women. However, except women's education only a few studies explored the implication of other proxy indicators of women's status, i.e women's socio-economic and domestic decision-making position etc. on couples' contraceptive behavior. In Ethiopia only a few studies explored the role of men and implication of women's socio-economic and decision-making autonomy on couple's family planning practices (Amaha Haile and Fikre Enqueslassie 2006).

Religion

Demographers have historically been interested in the relationship between fertility and religion in sub-Saharan Africa because of the region's high fertility and dynamic and influential religious environment. For example, John Caldwell and Pat Caldwell began their 1987 article with the statement: ―Sub-Saharan Africa may well offer greater resistance to fertility decline than any other world region. The reasons are cultural and have much to do with a religious belief system that operates directly to sustain high fertility but that also has molded a society in such a way as to bring rewards for high fertility" (Caldwell and Caldwell 1987:409). They were not alone in this sentiment: religion was and to an extent still is seen largely as a barrier to fertility decline and to family planning adoption in the region.

McQuillan (2004) proposed three preconditions necessary for religion to influence fertility. He argues that religion will affect fertility behavior when it: (1) articulates norms relevant to fertility; (2) can communicate these values and promote compliance; and (3) is central to the social identity of its followers. In isolated communities, such as the rural villages of sub-Saharan Africa, norms are articulated by religious leaders and fostered within local congregations around which social identity develops.

Economic status

Using data from Demographic and Health Surveys conducted in the 1990s, the World Bank has produced a series of demographic and health indicators by wealth quintiles for 22 Sub-Saharan African Countries (Gwatkin et al., 2000). These results indicate that, in most countries, fertility clearly decreases with increasing economic status. Using the Côte d'Ivoire LSMS survey, Ainsworth (1989), had found a slightly negative relationship between economic status and fertility among uneducated women.

Household economic status of women also plays an important role in fertility levels. Women of upper class household economic status are likely to possess maximum and expensive household goods and have exposure to modern amenities and also likely to be better educated. For maintaining their standard they may prefer to have fewer children compared to women of lower economic class. Now under the economic pressure the overall perception of the people is changing as it is becoming difficult to afford large families of the past and it is truer for those who want modern amenities (Hakim and Miller, 1996).

2.3 Methodology review

Mokshed (2000) examines the effects of selected socio-demographic characteristics on desire for additional children among couples in Bangladesh. The study was based on the Bangladesh Demographic and Health Survey, 1996-1997. The dependent variable is desire for additional children. The explanatory variables of this study are respondent's residence, religion, education and occupation of husband and wife, media exposure, membership of social organization, sex composition of existing children and number of children. Multiple logistic regression has been employed to predict relationship among the dependent and independent variables.

Toefiqua (2009) examines the effects of some selected socio-economic and demographic variables on fertility using a well known multivariate technique called path model analysis (BDHS). The study argues that for both cohorts (i.e, aged 15-30 years and aged 30+ years) women's education, age at first marriage and length of breast feeding are found to have significant direct negative effects, while the place of residence and number of dead children have significant direct postive effects on the number of CEB. Fetal loss appears to have a significant direct postive effect on fertility in Bangladesh.

Ramesh (2010) investigates demographic, socio-economic and cultural factors affecting fertility differential in Nepal using NDHS 2006 data. He used both bivariate and multivariate analyses to describe the fertility differentials for ever married women of reproductive age (8,644). The bivariate analysis (one-way ANOVA) was applied to examine the association between children ever-born and women's demographic, socio-economic, and cultural characteristics. Besides bivariate analysis, the net effect of each independent variable on the dependent variable after controlling for the effect of other predictors has also been measured through multivariate analysis (multiple linear regression).The result indicates the mean numbers of children ever born (CEB) among married Nepali women of reproductive age under 40 and among women aged 40-49 were three and five children, respectively. Age at first marriage, perceived ideal number of children, place of residence, literacy status, religion, mass media exposure, use of family planning methods, household headship, and experience of child death were used to explain the variance in fertility in the study.

Tewodros et.al, (2010) studied the determinants of adolescent fertility in Ethiopia using the Ethiopian Demographic Health Survey 2005 (EDHS-2005) data. Multiple logistic regression model was used to identify socio-demographic and economic determinants and Bongaarts model was used to determine proximate determinants of fertility. The result indicates that the major factors associated with adolescent fertility are age, educational status, place of residence, employment, marriage, contraceptive use and postpartum infecundability.

A cross-sectional, descriptive study with internal comparison was conducted among 1376 women of reproductive age with the objective of assessing the level and determinants of fertility in Awassa town, Ethiopia by Samson and Mulugeta (2009). The study shows socio-demographic characteristics of mothers like low educational status, low or no income, rural place of birth, early marriage, and other variables like history of child death, negative attitude of husbands towards contraceptive use, poor educational status of husbands, need for additional children, were found to have significant association with high fertility.

CHAPTER THREE

DATAAND METHODOLOGY

3.1 Source of data

This research utilizes the 2005 Ethiopian Demographic and Health Survey data. The survey is the second comprehensive and nationally representative population and health survey. An important feature of the data set is that it avails in-depth information on demographic and health aspects of households. The survey questionnaire included, among others respondent's knowledge about reproduction, contraception, pregnancy, postnatal care and breastfeeding, immunization and health, marriage, fertility preference, women's work status, maternal mortality, female circumcision, AIDS and other STDS, sexual activity. The 2005 Ethiopian Demographic and Health Survey is designed to provide estimates for the health and demographic variables of interest for the following domains; Ethiopia as a whole; urban and rural areas of Ethiopia (each as a separate domain) and 11geographic areas (9 regions and 2 city administrations).

In general, the DHS sample is stratified, clustered, and selected in two stages. In the 2005 EDHS a representative sample of approximately 14,500 households from 540 clusters was selected. The sample was selected in two stages. In the first stage, 540 clusters (145 urban and 395 rural) were selected from the list of Enumeration Areas (EA) from the 1994 Population and Housing Census sample frame. In the second stage, a complete listing of households was carried out in each selected cluster. Between 27 and 32 households from each cluster were then systematically selected for participation in the survey. All women age 15-49 who were either permanent residents of the households in the 2005 EDHS sample or visitors present in the household on the night before the survey were eligible to be interviewed. In addition, in a sub-sample of half of all the households selected for the survey, all men age 15-59 were eligible to be interviewed if they were either permanent residents or visitors present in the household on the night before the survey.

Thus, the analysis presented in this study on the impact of socio-demographic factor on fertility is based on 10,199 women who have at least one child and whose age ranges from 15 to 49 years.

3.2 Variables included in the study

As demonstrated in the literature review socio-demographic characteristics are considered as the most important factors that affect fertility.

Children ever-born (CEB) comprise information on the number of all children born alive (life time fertility) up to the survey date. The mean number of children ever-born to women represents the child bearing experience of real age cohort and reflects current and past fertility behavior. CEB does allow for the generalization of data and understanding that can provide the basis for further analysis. The choice of this variable is based on its closeness to the concept of individual fertility. In most fertility survey this information is obtained through a direct question on total number of live births. Therefore, in these study children ever born is the dependent variable. The “control” category (coded as 0) represents women with fewer than five CEB and the “case” (coded as 1) represents women with five or more children.

Explanatory variables included for the study are:

1. Place of residence.

2. Age at first marriage; that is, the age at which a woman enters first union.

3. Ever-use of contraception method.

4. Wife’s level of education: This measure the highest level of school attended. Consequently, the standard classification covered five categories (illiterate, primary, junior secondary, senior secondary and higher level). However, in this study the four categories (illiterate, primary, secondary and higher) are used because in the EDHS data the variable is categorized this way.

5. Wife’s work status (as employee): The work status of a woman is classified in to two categories which are “working” (as wage/salary earner) and “not working” (not employed as wage/salary earner).

6. Husband’s level of education. This is defined in the same way as “women’s level of education”.

7. Religion.

8. Age (5-year groups).

These explanatory variables are selected by considering partly the evident strength of their relationship to fertility level observed in previous comparable studies and partly on practical grounds of data availability.

Table 3.1 Description of variable included in the analysis

Response variable

Variable	Representation of Variable	**Categories**
Total children ever-born (CEB)	Y	0= control 1=case

Predictor/explanatory variables

Variables	Representation of variable	Categories
Place of residence (Residence)	X_1	0=Urban 1=Rural
Age at first marriage of wife(Amarriage)	X_2	0=15-17 years 1=≥18 years
Ever- use of contraception method (Contraceptive)	X_3	0=Never used 1=Ever used
Wife level of education(Wedu)	X_4	0=Illiterate 1=Primary 2=Secondary 3=Higher
Wife work status(Work)	X_5	0=Not working (working without

		wage/salary) 1=Working (working as earner of wage/salary)
Husband level of education (Hedu)	X_6	0=Illiterate 1=Primary 2=Secondary 3=Higher
Religion	X_7	0=Coptic Orthodox 1=Muslim 2=Protestant 3=Others
Age (Age of wife in 5-year groups)	X_8	1=15-19 2=20-24 3=25-29 4=30-34 5=35-39 6=40-44 7=45-49

3.3 Methodology

3.3.1 Logistic regression analysis

Logistic regression is widely used to model the outcomes of a categorical dependent variable. For categorical variables it is inappropriate to use linear regression because the response values are not measured on a ratio scale and the error terms are not normally distributed. In addition, the linear regression model can generate as predicted values any real number ranging from negative to positive infinity, whereas a categorical variable can only take on a limited number of discrete values within a specified range.

Logistic regression has proven to be one of the most versatile techniques in the class of generalized linear models. Whereas linear regression models equate the expected value of the dependent variable to a linear combination of independent variables and their corresponding parameters, generalized linear models equate the linear component to some function of the probability of a given outcome on the dependent variable. In logistic regression, that function is the logit transform: the natural logarithm of the odds that some event will occur. In linear regression, parameters are estimated using the method of least squares by minimizing the sum of squared deviations of predicted values from observed values. This involves solving a system of N linear equations each having N unknown variables, which is usually an algebraically straightforward task. For logistic regression, least squares estimation is not capable of producing minimum variance unbiased estimators for the actual parameters. In its place, maximum likelihood estimation is used to solve for the parameters that best fit the data (Draper and Smith 1998).

Often the outcome variable in social data is in general not continuous instead a binary one. In such a case, binary logistic regression is a useful way of describing the relationship between one or more independent variables and a binary outcome variable, expressed as a probability scale that has only two possible values. Indeed, a generalized linear model is used for binary logistic regression. The most attractive feature of a logistic regression model is that it neither assumes the linearity in the relationship between the covariates and the outcome variable, nor does it require normally distributed variables. It also does not assume homoscedasticity and in general has less stringent requirements than linear regression models. Thus logistic regression is used in a wide range of applications leading to binary dependent data analysis (Hilbe 2009 and Agresti 2002)

3.3.2 Assumption for logistic regression

1. Logistic regression predicts the odds of an event occurring, which is based on the probability P, of that event occurring. Precisely, the odds of an event occurring is

$$odds = \frac{probability\ of\ event\ occuring}{probability\ of\ event\ not\ occuring} = \frac{p}{1-p}$$

2. In logistic regression residuals follow a binomial rather than a normal distribution. Normality of variables is not a stringent requirement.
3. Logistic regression does not assume homoscedasticity.
4. Logistic regression assume that there is little or no multicollinearity.

5. The error terms in logistic regression need to be independent. Logistic regression requires each observation to be independent. That is the data-points should not be from any dependent samples design.
6. The categories (groups) in logistic regression must be mutually exclusive and exhaustive; a case can only being one group and every case must be a member of one of the groups.
7. Logistic regression needs larger samples than for linear regression because maximum likelihood coefficients are large sample estimates.
8. In logistic regression the independent variables need not be interval, nor normally distributed, nor linearly related, nor of equal variance within each group (Pampel, 2000).

3.3.3 Model

Relationship between the probability model and the independent variables are usually nonlinear rather than linear. If we correctly specify the model as linear, the statistical properties derived under the linearity assumption will not, in general, hold. In this case it is proper to specify a non-linear probability model in place of the probability model (John and Forrest, 1984).

For a binary response Y_i and a quantitative explanatory variable X_{Ji} , j=1, 2, . . . , m and i=1, 2, . . . , n let π_i=P(X_{Ji}) denote the ―success probability" when X_{Ji} take the values x_{Ji}. The problem with linear model is that the probability model E(Y) (where β is the vector of parameters to be estimated) is used to approximate a probability value π_i=P(Y_i=1) within the interval 0 and 1, while $E(Y_i)$ is not to be constrained. Therefore, we apply the logit transformation where the transformed quantity $ln\,(\frac{\pi_i}{1-\pi_i})$ lies in the interval $(-\infty, +\infty)$ and it is modeled as

$$logit\,(\pi_i) = Ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 X_{1i} + \ ... + \beta_m X_{mi} (1)$$

This implies

$$\pi_i = \frac{exp(\beta_0 + \beta_1 X_{1i} + \ ... + \beta_m X_{mi})}{1 + \ exp(\beta_0 + \beta_1 X_{1i} + \ ... + \beta_m X_{mi})} \qquad (2)$$

where the parameter β_J determines the rate of increase or decrease of X_{Ji} on the log of odds that $Y_i = 1$, controlling for other $x\ s$. Furthermore, $exp(\beta_J)$ is the multiplicative effect on the odds of unit increase in X_{Ji}, at fixed level of other $X\ s$ (John and Forrest, 1984).

3.3.4 Parameter estimation

Maximum likelihood estimation is the most popular technique for estimating the parameters of the logistic regression model. For a binary random variable Y assuming values either 0 or 1, the probability, $P(Y = 1)$ is given as

$$P(Y = 1) = p = \frac{1}{1 + e^{-x\,\beta}} \qquad (3)$$

$$logit\,(p) = log\left({}^{p}/_{1\,-\,p}\right) = X\,\beta$$

where β is a vector of coefficients and X is the vector of independent variables including a vector X_0 (which is a vector of 1’s) since the value of the outcome variable are not available, we can’t estimate the parameters directly, however, the likelihood provides a solution.

Each observation can be considered as an outcome of a Bernoulli trial, and hence for the i^{th} observation

$$P(Y = y_i) = p_i^{y_i}(1 - p_i)^{1-y_i} \tag{4}$$

Assuming the n observations are independent, the likelihood function is

$$L = \prod_{i=1}^{n} p_i^{y_i}(1 - p_i)^{1-y_i} = \prod_{i=1}^{n} \left({}^{1}\!/_{1 + e^{-X\beta}} \right)^{y_i} \left({}^{e^{-X\beta}}\!/_{1 + e^{-X\beta}} \right)^{1-y_i} \tag{5}$$

And the log likelihood function given as

$$Log\ L = \sum_{i=1}^{n} y_i log\left(\frac{1}{1 + e^{-X\beta}}\right) + \sum_{i=1}^{n} (1 - y_i) log\left(\frac{e^{-X\beta}}{1 + e^{-X\beta}}\right) \tag{6}$$

Hence, by maximizing (6) we can theoretically estimate the parameter vector β. However, the resulting equation obtained by talking the first order derivatives do not have an analytical solution. Therefore, β will be obtained by maximizing (6) using numerical iterative techniques.

3.3.5 Goodness of fit for logistic regression

The goodness of fit of a model measures how well the model describes the response variable. Assessing goodness of fit involves investigating how close values predicted by the model are to the observed values. The appropriateness of the fitted logistic regression model needs to be examined before it is accepted for use as in the case of all regression models.

In practice, several different measures exist for determining the significance or goodness of fit of a logistic regression model. These are Pearson, Hosmer-Lemeshow, Deviance goodness of fit, likelihood ratio test and the classification table. In theoretical sense, all measures are equivalent. To be more precise, as the number of observation goes to infinity, all measures converge to the same estimate of the model significances. The test can detect major departures from a logistic response function. Stating the null hypothesis one can test the significance of the effect of X on the binary response (Bewick, Cheek and Ball, 2005).

$$H_0: \beta = 0 \ (the\ probability\ of\ presence\ is\ independent\ of\ x)$$
$$H_1: \beta \neq 0 \ (the\ probability\ of\ presence\ is\ not\ independent\ of\ x)$$

3.3.5.1 Hosmer and Lemeshow goodness-of–fit test

The Hosmer–Lemeshow test is a commonly used test for assessing the goodness of fit of a model and allows for any number of explanatory variables, which may be continuous or categorical. The test is similar to a x^2 goodness of fit test and has the advantage of partitioning the observations into groups of approximately equal size, and therefore there are less likely to be groups with very low observed and expected frequencies. The observations are grouped into deciles based on the predicted probabilities (Bewick, Cheek and Ball, 2005)

The test statistic is obtained by applying a chi-square test on a $2 \times g$ contingency table. The contingency table is constructed by cross-classifying the dichotomous dependent variable with a grouping variable (with g groups) in which groups are formed by partitioning the predicted probabilities using the percentiles of the predicted event probability. In the calculation, approximately 10 groups are used ($g = 10$). The corresponding groups are often referred to as the ―deciles of risk" (Hosmer and Lemeshow, 2001).

If the values of independent variables for observation i and i' are the same, observation i and i' are said to be in the same block. When one or more blocks occur within the same decile, the blocks are assigned to this same group. Moreover, observations in the same block are not divided when they are placed into groups. This strategy may result in fewer than 10 groups (that is, $g \leq 10$) andconsequently, fewer degrees of freedom.

Suppose that there are Q block, and q^{th} blocks has m_q observations, $q = 1, \dots, Q$. Moreover, suppose that the k^{th} group $(k = 1, 2, \dots, g)$ is composed of $q_1^{th}, \dots, q_k^{th}$ blocks of observations. Then the total number of observations in the k^{th} group is $s_k = \sum_{q_1}^{q_k} m_j$. The total observed frequency of events (that is, $Y = 1$) in the k^{th} group, call it O_{1k}, is the total number of observations in the k^{th} group with $Y = 1$. Let E_{1k} be the total expected frequency of the event (that is, $Y = 1$) in the k^{th} group; then E_{1k} is given by $E_{1k} = S_k \xi_k$, where ξ_k is the average predicted event probability for the k^{th} group

$$\xi_k = \sum_{q_1}^{q_k} m_j \hat{\pi}_j / s_k$$

The Hosmer-Lemeshow goodness-of-fit statistic is computed as

$$\hat{C} = \sum_{k=1}^{g} \frac{(O_{1k} - E_{1k})^2}{E_{1k}(1 - \xi_k)}$$

The p value is given by $Pr(x^2 \geq \hat{C})$ where the chi-square statistic is distributed with degree of freedom (g-2) (Hosmer and Lemeshow, 2001).

3.3.5.2 Likelihood ratio test

The statistic used to determine the overall significance of a logistic model is called the likelihood ratio test. The likelihood ratio test compares the likelihood of the full model (i.e. with all the predictors included) with the likelihood of the null model (a model which contains only the intercept). It is analogous to the overall F-test, as used in linear regression. The formula for the likelihood ratio test statistic is

$$G^2 = 2\,log\frac{L}{L_0} = 2(log L - log\,L_0)$$

where L is the likelihood of the full model and L_0 is the likelihood of the null model. The likelihood ratio test statistic has an approximate x^2 distribution with k degrees of freedom (where k is the number of predictors in the full model). If significant, it suggests that, taken together, the predictors contribute significantly to the prediction of the outcome. Two conditions must be met when calculating the likelihood ratio test statistic:

- ✓ Both models must be fit using exactly the same observations. If a data set contains missing values for some predictors in the full model, then these would be omitted from the full model but included when the null model is computed. This must be avoided.
- ✓ The models must be nested. This means that the predictors in the simpler model must be a subset of those in the full model. This will not be a problem when the smaller model is the null model, but might be a problem in other situations (Stevenson, 2008).

3.3.5.3 R^2 for logistic regression

Most statistical packages provide statistics that maybe used to measure the usefulness of the model and that are similar to the coefficient of determination (R^2) in linear regression. The Cox & Snell and the Nagelkerke R^2 are two such statistics. The Cox and Snell's R^2 given by

$$R_{CS}^2 = 1 - \left(\frac{l(0)}{l(\hat{\beta})}\right)^{\frac{2}{w}}$$

where $w = \sum_i^n c_i$,c_i is the case weight,$l(\hat{\beta})$ the likelihood of the current model, and $l(0)$ is the likelihood of the initial model; that is $l(0) = w\,log(0.5)$ if the constant is not included in the model; $l(0) = w\left[\hat{}_0\, log\left\{\hat{}_0 \big/ 1 - \hat{}_0\right\} + log\left(1 - \hat{}_0\right)\right]$ if the constant is included in the model, where $\hat{}_0 = \sum_i^n c_i y_i \big/ w$. The Nagelkerke R^2 given by

$$R_N^2 = R_{CS}^2 \big/ max(R_{CS}^2)$$

where $max(R_{CS}^2) = 1 - \{l(0)\}^{2/w}$.

The maximum value that the Cox & Snell R^2attains is less than 1. The Nagelkerke R^2 is an adjusted version of the Cox & Snell R^2 and covers the fullrange from 0 to 1, and therefore it is often preferred. The R^2statistics do not measure the goodness of fit of the model but indicate how useful the explanatory variables are in predicting the response variable and can be referred to as measures of effect size (Cox and Snell, 1989; Nagelkerke, 1991).

3.3.6 Test of a single predictor

3.3.6.1 Wald test

An alternative approach to evaluating the significance of a single coefficient is to use a test that relates the coefficient to its standard error. A Wald test is the ratio of the coefficient estimate to its standard error and it follows (asymptotically) a standard normal distribution. This tests whether the coefficient is significantly different from zero. It is routinely computed by most computer programs and is the most widely used test of the significance of coefficients. However, the estimates of the coefficient and its standard error are only estimates and consequently, the normal approximation of its distribution might not be reliable (particularly if the sample size is small). Consequently, to evaluate the significance of variables with a p-value close to the rejection region, it is best to use a likelihood ratio test. Just as with multiple partial F-tests in linear regression, multiple parameters in a logistic model can be tested with a multiple Wald test. This is equivalent to a null hypothesis as follows$H_0: \beta_j = 0$. In this case, the test statistic is compared to a x^2 distribution with the degrees of freedom equal to the number of predictors being tested.The Wald statistic is:

$$W = \left(\hat{\beta} \middle/ se(\hat{\beta}) \right)^2.$$

For large sample size assuming normality of maximum likelihood estimates the Wald statistic has an approximate chi-squared distribution with one degree of freedom (Stevenson 2008).

3.3.7 Model selection criteria

The goal of logistic regression is to correctly predict the category of the outcome for individual cases using the most parsimonious model. To accomplish this goal, a model is created that includes all variables that are useful in predicting the response variable. The criteria for inclusion of a variable in a model may vary from one problem to the next and from on scientific discipline to another. For instance, a model with several predictors has the potential for multicollnearity. Strong correlations among predictors make it seem that no one variable is important when all the others are in the model. A variable may seem to have little effect simply because it ―overlaps" considerably with other predictor variables in the model (Agresti 2008).

As described by Hosmer and Lesmeshow (2001), there are certain steps that can be followed in the selection of variables for a logistic regression model. First, the selection process should begin with a careful univariate analysis of each the variables. For this purpose, the Pearson's chi-square and/or the likelihood ratio chi-square tests may be used. Up on completion of the univariate analysis we select variables for the multiple logistic regression analysis. Any variable whose univariate test has a p-value <0.05 should be considered as a candidate for multiple logistic regression analysis. Finally, following the fit of the multiple logistic regression model, the importance of each variable included in the model should be verified. This should include an examination of the Wald statistic for each of the variables and the comparison of each estimated coefficient with the coefficient from the univariate model containing only that variable. Variables that do not contribute to the model based on these criteria should be eliminated and a new model will be fitted. The new model should be compared to the old model through the likelihood ratio test. Also, the estimated coefficient for the remaining variables should be compared to those from the full model.

Another approach to variable selection is to use stepwise selection procedure. Stepwise selection of variables has been widely used in linear regression. In this method, variables are selected for either inclusion or exclusion from the logistic regression model in a sequential fashion based on statistical criterion that checks for the ―importance” of variables. The ―importance” of variables is defined in terms of a measure of the statistical significance of the coefficient for the variable. In stepwise selection procedure, backward selection and/or forward selection procedure are used (Hosmer and Lesmeshow 2001).

In backward elimination procedure, we start with a model that contains all the predictors and we systematically remove the largest non-significant p-value terms until we are left with a subset that consists of entirely statistically significant terms. Backward regression appears to be the preferred method of exploratory analysis, where the analysis begins with a full or saturated model and variables are eliminated from the model in an iterative process. The fit of the model is tested after the elimination of each variable to ensure that the model still adequately fit the data. When no more variables can be eliminated from the model, the analysis has been completed (Hosmer and Lesmeshow 2001).

The forward selection procedure starts with no predictors in the model and examines each term that could be possibly added and then add the most significant predictor, or the predictor with the smallest p-value. In the next stage, the procedure adds the next most significant term and checks to see if any previous terms are now non-significant and removes them if they are not significant. This procedure continues until there are no further significant terms to be added. Therefore, unlike backward elimination, this procedure builds the model by adding terms (Hosmer and Lesmeshow 2001).

3.3.8 Model checking

Fitting logistic regression models provides predicted probabilities that y=1. At each setting of the explanatory variables, one can multiply the predicted probability by the number of subjects to obtain a fitted count. The test of the null hypothesis that the model holds compares the fitted and observed counts using a Pearson'sx^2 or likelihood-ratio test statistics (G^2).x^2 and G^2 are given by

$$x^2 = \sum \left[{(obseved - fitted)^2}/{fitted} \right]$$

$$G^2 = 2\sum \left[(observed)\, log\left({observed}/{fitted} \right) \right] = -2\, ln\{{LLC}/{LLS}\}$$

For a fixed number of settings, when most fitted counts are about five, x^2 and G^2 have approximate chi-squared distributions. The degrees of freedom (the residual degrees of freedom) for the model, equal the number of sample logit (i.e. the number of settings of explanatory variables) minus the number of model parameters. Largex^2 or G^2 values indicate lack of fit of the model. A p-value is the right-tail probability above the observed value. When the fit is poor, residuals and other diagnostic measures describe the influence of individual observations on the fit and high light reasons for the inadequacy (Agresti, 2008).

A constant model is a model, which contains only the constant term, while a saturated model is a model which contains all factors selected by the stepwise procedure.

Goodness-of-fit statistics such as x^2 and G^2 are summary indictors of the overall quality of fit. Additional diagnostic analysis is necessary to describe the nature of any lack of fit. It is important to examine the adequacy of the resulting model in logistic regression. There are comparable diagnostics that should be used to identify data problems. The logistic regression provides a variety of such statistics.

There are three ways that an observation can be considered as unusual, namely outlier, influence and leverage. In logistic regression, a set of observations whose values deviates from the expected range and produce extremely large residuals and may indicate a sample peculiarity is called outliers. These outliers can unduly influence the results of the analysis and lead to incorrect inferences. An observation is said to be influential if removing the observation substantially changes the estimate of coefficients. Influence can be thought of as the product of leverage and outliers. An observation with an extreme value on a predictor variable is called a point with high leverage. Leverage is a measure of how far an independent variable deviates from its mean. In fact, the leverage indicates the geometric extremeness of an observation in the multi-dimensional covariate space. These leverage points can have an unusually large effect on the estimate of logistic regression coefficients (Cook, 1998).

The residual is the difference between the observed probability and the predicted probability of the event based on the model. It can be viewed as a component of $-2LL$ (-2 log-likelihood), which compares a model to the ―perfect ―model. For each case the deviance compares the predicted probability of being in the correct group based on the model to the prediction of one.

$$Deviance = \sqrt{-2\,log\left({}^{L_0}\!/_{L_1}\right)}$$

where L_1 is always one, since the likelihood of the correct prediction in a perfect model is one, and L_0 is the predicted probability of membership in the correct group. Large values for deviance indicate that the model does not fit the case well.

For large sample sizes the deviance is approximately normally distributed. The studentized residual for a case is the change in the model deviance if the case is excluded. Discrepancies between the deviance and the studentized residual may identify unusual cases. The logit residual is the residual for the model if it is predicted in the logit scale

$$Logitresidual = {}^{residual}\!/_{\pi_i(1-\pi_i)}$$

Leverage value is used for detecting observation that have a large impact on the predicted values. Unlike linear regression, the leverage values in logistic regression depend on the dependent variable scores and the design matrix. The leverage values vary between zero and one. A leverage value of one means, the model is being forced or levered to fit the corresponding case exactly. Thus the leverage can be used to detect influential outliers. The leverage of any given case may be compared to the average leverage equals ${}^{k}\!/_{n}$ where k are the number of estimated parameters, including the constant, and n is the sample size.Cases are influential having $\hat{h}_{ii} > 2\frac{k}{n}$. The leverage (h_{ii}) is the i^{th} diagonal element of the matrix

$$\hat{H} = \hat{V}^{1/2}X(X\,\hat{V}X)^{-1}X\,\hat{V}^{1/2}$$

where $\hat{H}$is then $n \times n$ estimated hat matrix and $\hat{V} = Diag\{\hat{\pi}_1(1-\hat{\pi}_1), \ldots, \hat{\pi}_n(1-\hat{\pi}_n)\}$ is $n \times n$ diagonal matrix(Cook, 1998).

An influence measure proposed by Cook is widely used. Cook's distance measures the difference between the regression coefficients obtained from the full data and the regression coefficients obtained by deleting the i^{th} observation, or equivalently the difference between the fitted values obtained from the full data and the fitted values obtained by deleting the i^{th} observation. Accordingly, Cook's distance measures the influence of the i^{th} observation by

$$D_i = \frac{\left(\hat{\beta} - \hat{\beta}_{(-i)}\right)\left(X\,\hat{V}X\right)\left(\hat{\beta} - \hat{\beta}_{(-i)}\right)}{p\hat{\sigma}^2}$$

it can be shown that D_i can be expressed as

$$D_i = \frac{r_i^2}{p}\left(\frac{\hat{h}_{ii}}{1-\hat{h}_{ii}}\right) \qquad i = 1, 2, \ldots, n$$

$where\ r_i$ is the i^{th} standardized person residual is given by

$$r_i = \frac{y_i - \hat{\pi}_i}{\sqrt{v_i(1-\hat{h}_{ii})}} \qquad i = 1, 2, \ldots, n$$

$\hat{h}_{ii}$ is the i^{th} leverage value, p regression coefficients including the constant and v_i is the i^{th} diagonal element of V. It has been suggested that points with D_i values greater than 1 as being influential (Chatterjee and Hadi 2006).

Another useful diagnostic measure is the change in logistic measure coefficients when the case is deleted from the model is DFBETA. For each parameter estimate, the procedure calculates a DFBETA diagnostic for each observation. The DFBETA diagnostic for an observation is the standardized difference in the parameter estimate due to deleting the observation, and it can be used to assess the effect of an individual observation on each estimated parameter of the fitted model and it is given by $DFBETAS_{j(i)}$, j = 0, 1, 2, . . . , m where each $DFBETAS_{j(i)}$ is the standardized change in $\hat{\beta}_j$ when the i^{th} observation is deleted from the analysis. Thus,

$$\mathrm{DFBETAS}_{j(i)} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{S_i\sqrt{c_{jj}}}$$

where $\hat{c}_{jj}$ is the j^{th} diagonal element from $\left(X\,\hat{V}^{-1}X\right)^{-1}$

In summarizing the diagnostic methods the residuals comparing observed and fitted counts are useful for this purpose. So the residuals (used to detect multicollinearity and other inadequacies), the leverage (used for detecting observation that have a large impact on the predicted values), Cook's distance (a measure of the influence a case on the residuals), and DFBETA (a measure of the change in the logit when a case is deleted from the model) are useful for model diagnostics.

CHAPTER FOUR
STATISTICAL DATA ANALYSIS

In this chapter, we are going to analyze socio-demographic factors that may have impact on total number of children ever-born by a woman in Ethiopia. The data used in this study for the analysis were obtained from the 2005 Ethiopian Demographic and Health Survey (EDHS) with reference to a total of 14,070 women in the age group 15-49 years. The dependent variable is a dichotomous random variable –case” (coded as 1) and –control” (coded as 0). Descriptive and binary logistic regression methods are used to measure the effects of socio-demographic factor that affect total CEB in Ethiopia. The descriptive part provides percentages of total CEB of Ethiopian women. The binary logistic regression analysis is employed to identify socio-economic factors that have effect on total CEB. The data are analyzed using the Statistical Package for Social Sciences (SPSS) version 13.

4.1 Summary of descriptive statistics

Table 4.1 Socio-demographic characteristics of total number of children ever-born

		Total number of children ever-born			
		Control		Case	
Predictor	category	Count	Row N %	Count	Row N %
Place of residence	Urban	1860	79%	509	21%
	Rural	4501	57%	3329	43%
Religion	Coptic Orthodox	3153	67%	1540	33%
	Muslim	2069	58%	1512	42%
	Protestant	958	60%	638	40%

	Others	181	55%	148	45%
Wife level of education	Illiterate	4131	56%	3302	44%
	Primary	1190	74%	410	26%
	Secondary	890	89%	107	11%
	Higher	150	89%	19	11%
Wife work status	Not working	4474	61%	2839	39%
	Working	1887	65%	999	35%
Age 5-year groups	15-19	865	100%	0	0%
	20-24	1692	99%	22	1%
	25-29	1823	83%	361	17%
	30-34	855	52%	791	48%
	35-39	562	36%	1002	64%
	40-44	328	28%	831	72%
	45-49	236	22%	831	78%
Age at first marriage	15-17 years	4058	58%	2961	42%
	eighteen years and above	2303	72%	877	28%
Husband level of education	Illiterate	3203	54%	2740	46%
	Primary	1557	68%	729	32%
	Secondary	1249	80%	309	20%
	Higher	352	85%	60	15%
Ever use of contraceptive	Never used	4424	59%	3013	41%
	Ever used	1937	70%	825	30%

The Table 4.1 reveals that total number of children ever-born differ by women's educational attainment. For instance, 4131(56%) of illiterate women are controls (have less than five CEB) and the remaining 3302(44%) are cases (have five or more CEB). On the other hand women who have higher educational level 150(89%) are controls and 19(11%) are cases.

The 4474 (61%) not working women are controls (have less than five CEB) and 2839(39%) are cases (have five or more CEB). Likewise, from working women 1887(65%) are controls and 999(35%) are cases.

Results on husband educational level show that from women whose husbands are illiterate 3203(54%) are controls and 2740(46%) are cases. Similarly, women whose husbands are higher educated 352(85%) are controls and 60(15%) are cases.

Women of the age group 15-19 are 865(100%) controls (meaning those having given less than five live births) because of their younger age (15-19) as we do not expect them to have more than five children. Women in the age group 45-49, 236 (22%) are controls and 831(78%) are cases.

From among Coptic Orthodox women 3153(67%) are controls and 1540(33%) are cases. Women who are Muslim 2069(58%) are controls and 1512(42%) are cases. Similarly, women who are followers of other religions 181(55%) are controls and 148(45%) are cases.

With regards to contraceptive use, women who never use any contraceptive method 4424(59%) are controls and 3013(41%) are cases. From women who ever use any contraceptive method 1937(70%) are controls and 825(30%) are cases.

The total CEB differ by place of residence. Among women who live in urban areas 1860(79%) are controls and 509(21%) are cases. Among rural women 4501(57%) are controls and 3329(43%) are cases.

Total CEB differ by age at first marriage. Among women whose age at first marriage was below eighteen years 4058(58%) are controls and 2961(42%) are cases. Women whose age at first marriage above eighteen 2303(72%) are controls and 877(28%) are cases.

4.2 Analysis of binary Logistic Regression

Table 4.2Case Processing Summary

Unweighted Cases(a)		N	Percent
Selected Cases	Included in Analysis	10199	100.0
	Missing Cases	0	.0
	Total	10199	100.0
Unselected Cases		0	.0
Total		10199	100.0

a. If weight is in effect, see classification table for the total number of cases.

Table 4.2 provides a basic descriptive window showing how many subjects were analyzed, and how many subjects were missing. As can be seen above, all 10,199 subjects in this analysis were included, with zero subjects missing, and none unselected.

Table 4.3 Response Variable Encoding

Original Value	Internal Value
Control	0
Case	1

Block 0: Beginning Block

Table 4.4 Iteration history for intercept only model

Iteration History[a,b,c]

Iteration		-2 Log likelihood	Coefficients Constant
Step 0	1	13508.420	-.495
	2	13508.157	-.505
	3	13508.157	-.505

a. Constant is included in the model.

b. Initial -2 Log Likelihood: 13508.157

c. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

Table 4.5 Classification Table

Classification Table[a,b]

	Observed		Predicted: Total number of children ever-born: Control	Predicted: Total number of children ever-born: Case	Predicted: Percentage Correct
Step 0	Total number of children ever-born	Control	6361	0	100.0
		Case	3838	0	.0
	Overall Percentage				62.4

a. Constant is included in the model.

b. The cut value is .500

The Beginning Block evaluates our model with only the constant in the equation (sometimes called the null model). The constant is analogous to the y-intercept in OLS regression. Table 4.5 is an SPSS classification table for the response variable based on how well the model does with only a constant term included. Notice that with the constant-only model, the predictive power of the logistic regression is perfect for those women who are ―control" or ―women who have less than five CEB," with 6361 subjects correctly classified, but not good for those "cases" or ―women who have five or more CEB". For those 3838 subjects, the constant-only logistic regression correctly classified 0% of them. The overall classification for the model is equal to 6361/10199 = 62.4%. What this means is that the model doesn't do a good job at classifying subjects, which actually is expected at this stage of the analysis since we only have the constant term included in the model. That is, we have yet to use more predictors to aid in classification and to sharpen our predictive power.

Table 4.6 Variables in the Equation

	$\hat{\beta}$	S.E.	Wald	Df	Sig.	Exp($\hat{\beta}$)
Step 0 Constant	-.505	.020	611.024	1	.000	.603

Table 4.6 shows the logistic coefficient ($\hat{\beta}$) associated with the intercept as it is included in the model. This table is similar to and contains analogous information as the coefficients table in a standard regression.The logistic coefficient for the constant is similar to the y-intercept term in standard regression. The Wald statistic is a chi-square 'type' of statistic and is used to test the significance of the variable in the model. The value of exp($\hat{\beta}$) is the estimated odds ratio attributed to a variable. In Table 4.6 for the intercept-only model is $\ln(\text{odds}) = -0.505$. If we exponentiate both sides of this

expression we find that our predicted odds [Exp($\hat{\beta}$)] = .603. That is, the predicted odd of having five or more CEB is .603. Since 6361of our subjects have less than five CEB and 3838 of them have five or more CEB, our observed odds are 3838/6361 = .603.

Block 1: Method = Enter

Table 4.7 Omnibus test of model coefficients

Omnibus Test of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	5836.495	19	.000
	Block	5836.495	19	.000
	Model	5836.495	19	.000

Table 4.7 gives results of "Omnibus Test." "Omnibus" means "overall," and so this output tells whether the model with all explanatory variables predicts the response better than the intercept only model. The results show that the model with explanatory variables does better at predicting the response variable, and is statistically significant at $p < .05$. The Omnibus Test gives a Chi-Square of 5836.495 with 19df, significant beyond .005. This is a test of the null hypothesis that adding explanatory variables to the model has not significantly increased our ability to predict the decisions made by our subjects. Since the omnibus test is significant we can conclude that adding the explanatory variables to the model has significantly increased our ability to predict total number of children ever-born made by our subjects.

Table 4.8 Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	7671.662(a)	.436	.594

a. Estimation terminated at iteration number 8 because parameters estimates changed by less than .001.

Table 4.8 gives result of "Model Summary," which are summary statistics for the model at "Step 1" which is the model with 8 predictors. Under Model Summary we see that the -2 Log Likelihood statistic is 7,671.662. This statistic measures how poorly the model predicts the decisions**.** The smaller the statistic the better the model. Although from iteration history (beginning block) we can see that the -2 log likelihood for the constant only model is 13,508.157. Adding the predictor variables reduced the -2 Log Likelihood statistic by 13,508.157 – 7,671.662 = 5,836.495 (the value of chi-squared above). The "Cox& Snell R Square" value is .436. This statistic is referred to as a "pseudo-R" statistic, in that it is designed to tell us something similar to what R-squared tells us in ordinary least-squares regression, that of the proportion of variance accounted for in the dependent variable based on the predictive power of the explanatory variables in the model(Meyer, Gamst, &Guarino 2006).

Table 4.9Hosmer and Lemeshow Test

Step	Chi-square	Df	Sig.
1	11.688	7	.166

Contingency Table for Hosmer and Lemeshow Test

		Total number of children ever-born = Control		Total number of children ever-born = Case		Total
		Observed	Expected	Observed	Expected	
Step 1	1	2039	2033.154	2	7.846	2041
	2	990	990.768	27	26.232	1017
	3	920	908.040	100	111.960	1020
	4	701	703.160	207	204.840	908
	5	653	663.811	369	358.189	1022
	6	427	440.404	595	581.596	1022
	7	287	304.728	731	713.272	1018
	8	209	199.118	821	830.882	1030
	9	135	117.816	986	1003.184	1121

The "Hosmer and Lemeshow Test" is a measure of fit which evaluates the goodness of fit between predicted and observed probabilities in classifying the response variable. Similar to the -2 log likelihood test, we want this chi-squared value ($x^2_{(7,0.05)} = 11.688$) to be low and non-statistically significant (p-value=.166) if the predicted and observed probabilities match up nicely. In this case we see that the test is statistically insignificant (p >.05), suggesting that the probabilities of predicted versus observed values of the response variable match up as nicely as we would like. Therefore, our fitted logistic regression model is good fit.

Table 4.10 Classification table for block one

Classification Table[a]

Observed			Predicted: Total number of children ever-born: Control	Predicted: Total number of children ever-born: Case	Percentage Correct
Step 1	Total number of children ever-born	Control	5397	964	84.8
		Case	801	3037	79.1
	Overall Percentage				82.7

a. The cut value is .500

A classification table is a 2×2 table in the logistic regression output for a dichotomous response and reports correct and incorrect estimates obtained by the logistic regression model.

Observed - This indicates the number of 0's and 1's that are the observed values of the response variable.

Predicted - These are the predicted values of the response variable based on the full logistic regression model. Table 4.10shows how many subjects are correctly predicted (5397 subjects are observed to be 0 and are correctly predicted to be 0; 3037 subjects are observed to be 1 and are correctly predicted to be 1), and how many subjects are not correctly predicted (964 subjects are observed to be 0 but are predicted to be 1; 801 subjects are observed to be 1 but are predicted to be 0).

Overall Percentage - As can be seen from the classification table, the overall percentage of those correctly classified is equal to 82.7%, a big increase from the constant-only model correctly classifying 62.4%. We can inspect the table for some specifics now: of those women who have less than five CEB, the model correctly classifies 5397 out of 6361subjects (the marginal total for "control"), or 84.8%. The model predicting women who have five or more CEB (cases) correctly classifies 3037 out of 3838, or 3037/3838 = 79.1% of cases. The "Overall Percentage" of 82.7 is calculated by summing up the diagonal values (5397 and 3037) and dividing over the total number of subjects. The "diagonal" consist of the sum of "Control-Control" and "Case-Case" frequencies. Therefore, the resulting percentage correctly classified is 8434/10199 = 82.7%. The note at the bottom of the table regarding the cut-off value being .500 simply states that when the classification table was produced by SPSS, it needed a way to group the resulting frequencies in the correct cells.

4.3 Model diagnosis

Besides global examination of the model, it is also useful to investigate the characteristics of individual subjects in our data. That is, the adequacy of the fitted model is checked for possible presence and treatment of outliers and influential values. The diagnostic test results for detection of outliers and influential values are given in the following table.

Table 4.11 Results of diagnostic tests influential values

	Total N	Minimum	Maximum
Analog of Cook's influence statistics	10199	.00000	.02196
Leverage value	10199	.00003	.00651
Standard residual	10199	-2.85390	2.79926
Deviance value	10199	-2.85358	2.79893
DFBETA for constant	10199	-.00627	.01351
DFBETA for Residence	10199	-.00861	.00586
DFBETA for Religion	10199	-.00319	.00227
DFBETA for Wedu	10199	-.00482	.00649
DFBETA for Work	10199	-.00270	.00316
DFBETA for Age	10199	-.00170	.00042
DFBETA for Amarriage	10199	-.00210	.00370
DFBETA for Hedu	10199	-.00486	.00405
DFBETA for Contraceptive	10199	-.00438	.00375

The above table shows that the standardized residuals are within the interval of -3 and +3 which implies that no outlier is detected at the 0.05 significance level. The Cook's influence statistics (D_i) less than one and the DFBETA's for model parameters including the constant term are all less than 0.02. Therefore, we conclude that the model is adequate.

Table 4.12 Result of binary logistic regression

Estimates, standard errors, Wald, degree of freedom, p-values, estimated odds-ratio

Predictors	$\hat{\beta}$	S.E	Wald	df	Sig	$Exp(\hat{\beta})$= odds ratio	95% CI for $\widehat{OR}$	
							Lower	Upper
Residence								
Urban	-1.283	.091	197.395	1	.000	.277	.232	.332
Rural (ref)								
Wife level of education			26.005	3	.000			
Illiterate	.782	.306	6.527	1	.011	2.185	1.200	3.979
Primary	.719	.305	5.549	1	.018	2.052	1.128	3.732
Secondary	.053	.302	.030	1	.861	1.054	.583	1.905
Higher (ref)								
Religion			99.293	3	.000			
Coptic orthodox (ref)								
Muslim	.621	.068	84.645	1	.000	1.862	1.631	2.125
Protestant	.191	.086	48.268	1	.000	1.823	1.539	2.160
Others	.410	.156	6.900	1	.009	1.507	1.110	2.046
Ever use of contraceptive								
Never used								
Ever used (ref)	-.330	.074	19.761	1	.000	.719	.621	.831
Husband level of education			12.391	3	.006			
Illiterate								
Primary	.521	.202	6.623	1	.010	1.683	1.132	2.503
Secondary	.659	.203	10.511	1	.001	1.932	1.298	2.878
Higher (ref)	.432	.197	4.811	1	.028	1.541	1.047	2.267
Age 5-year group								
15-19			1903.897	6	.000			

20-24	-23.407	1329.176	.000	1	.986	.000	.000	
25-29	-6.368	.236	730.867	1	.000	.002	.001	.003
30-34	-3.571	.111	1035.232	1	.000	.028	.023	.035
35-39	-1.768	.104	288.791	1	.000	.171	.139	.209
40-44	-.876	.104	71.555	1	.000	.416	.340	.510
45-49 (ref)	-.442	.111	15.798	1	.000	.643	.517	.799
Age at first marriage								
15-17 years								
18 years and above(ref)	1.203	.064	351.899	1	.000	3.330	2.936	3.776
Wife work status								
Not working								
Working(ref)	.297	.064	21.346	1	.000	1.346	1.187	1.527
Constant	-.112	.330	.011	1	.041	.088		

Table 4.12 above contains the estimated coefficients (under the column heading $\hat{\beta}$) and estimated values of the logistic regression model that predict the total CEB. The standard error of the estimates (under the column heading S.E) will help in computing the Wald statistics. The Wald statistic, which is the square of the ratio of the coefficient to its standard error, has a chi-square distribution with one degree of freedom. The significance of the Wald statistic (under the column labeled Sig) tells the importance of the predictor variable in the model.

The column $\exp(\hat{\beta})$, is the factor by which the odds of total number of children ever-born change when the i^{th} independent variable increases by one unit. If $\hat{\beta}_i$ is positive, $\exp(\hat{\beta}_i)$

will be greater than one, which means the odds of total CEB increases. If $\hat{\beta}_i$ is negative, $\exp(\hat{\beta}_i)$ will be less than one, which means the odds of total CEB decreases.

4.4 Estimated odds ratios and 95% confidence interval for odds ratio

We can interpret the odds ratio of place of residence obtained in the above table using the reference category rural. Women who live in urban area are 82.3% [$\widehat{OR}$= 0.227, 95% CI 0.232-0.332] times less likely to have five or more children in comparison with women who live in rural area controlling for other variables in the model.

For variable religion, the reference category is ―Coptic Orthodox". Women who follow Muslim are about 86% [$\widehat{OR}$ 1.862, 95% CI 1.631-2.125] more likely to have five or more children in comparison with women who follow Coptic Orthodox controlling for other variables in the model. Protestant and others religion women followers are about 82% [$\widehat{OR}$=1.823, 95% CI 1.539-2.160] and nearly 51% [$\widehat{OR}$ 1.507, 95% CI 1.110-2.046] more likely to have five or more children, respectively, compared to the reference category (controlling for other variables).

For the variable ―wife level of education" the reference category is ―higher education". The odds of women to have five or more children is about twice [$\widehat{OR}$=2.185, 95% CI 1.200-3.979] more likely for illiterate women; for women with primary education it is also twice [$\widehat{OR}$=2.052, 95% CI 1.128—3.732] more likely. The odds of women with secondary education to have five or more children is almost the same as women with ―higher" education in the model.

For the variable age 5-year groups the reference category is 45-49. Women with the age group 20-24 and 25-29 were found to have less chance of having CEB greater or equal to five, with estimated odds ratio of 0.002[95% CI 0.001-0.003] and 0.028[95% CI 0.023-0.035] respectively. Those in the youngest group of 15-19 are naturally very unlikely to have five or more children ($\widehat{OR} = 0.000$). From the descriptive statistics we can see that women in the age group 15-19 all of them are control (have less than five CEB).

For the variable “age at first marriage”, the reference category is eighteen years and above. Women with age at first marriage below eighteen years are [$\widehat{OR}$=3.33, 95% CI 2.936-3.776] more than three times more likely to have five or more children in their life time in comparison with women with age at first marriage eighteen years and above.

For the variable “husband’s level of education”, the reference category is “higher education”. The likelihood of having five or more children is 68.3% higher [$\widehat{OR}$=1.683 95% CI 1.132-2.503] among women whose husbands are illiterate compared to women married to a husband higher level of education. Similarly, the odds of having five or more children is 93.2%higher [$\widehat{OR}$=1.932 95% CI 1.298-2.878] among women whose husband have primary education controlling for other variable in the model. For those with secondary level education the odds of having five or more children is about 54% higher [$\widehat{OR}$=1.541 95% CI 1.047-2.267].

For the variable “ever use of contraceptive”, the reference category is “ever used” any contraceptive method. The odds of women to have five or more children is about 28%

lower [$\widehat{OR}$=0.719, 95% CI 0.621-0.831] among women who never used any contraceptive method.

For the variable wife work status, the reference category is working women. The odds of women to have five or more children is 34.6% higher [$\widehat{OR}$=1.346, 95% CI 1.187-1.527] for women working not as employed as compared to those women who work controlling other variables in the model.

CHAPTER FIVE

DISCUSSION, CONCLUSION AND RECOMMENDATIONS

5.1 Discussion and Conclusion

This study is undertaken to determine effect of selected socio-demographic characteristics on fertility. Data from the Ethiopia Demographic and Health Survey (EDHS) 2004/2005 are used for analysis. A total of 10,199 women aged 15-49 who had at least one child were chosen from the main sample in the EDHS as the sample respondents of this study. For these purpose, descriptive statistics and binary logistic regression are used. The results obtained are discussed as follows.

The finding of this study shows that total CEB significantly is higher among women who are illiterate and with primary education. This result is similar to that of Ramesh (2010). Education exposes women to information, empowers women, makes them more likely to be employed outside their home environment, and makes them more aware of their own health and the health of their children. Educated women are more likely to postpone marriage, have smaller family size, and use contraception than are uneducated women (Ramesh Adhikari 2010).

The finding of this study shows that residential differences had a significant impact on total CEB. Women who lived in rural areas were more likely to have more children than urban women. This finding corresponds with the result of a study in Nepal by Sharma (1998) where they show that rural women desire to have more children than urban women. The difference in desire for additional children among rural and urban women could be due to the parents' perceived costs and benefits of children. Rural women perceive greater benefit from children and see lots of advantages in having large families. They also anticipate support from their children in old age.

A study by Abdul Hakim (1994) shows that fertility is lower in urban areas than in rural areas due to the impact of a host of factors. The lower fertility in urban areas may be viewed from two perspectives. Firstly, at the aggregate (macro) level of analysis, urban fertility may be lower because the urban population has a larger proportion of collar jobs and who are thus likely to marry late and have smaller families. Secondly, at the individual (micro) level of analysis, one may hypothesize that the same level of education and income will produce lower fertility in urban setting. Moreover, the urban population has a higher proportion of those who are expected to adopt new ideas and new life styles first and of those for whom a large family does not represent an economic benefit but rather a burden.

Age at first marriage is another important variable affecting total CEB. This study shows that women married before reaching 18 years of age are more likely to have more children in their life time in comparison with women married after eighteen years. Ramesh (2010) shows that older age at first marriage plays an important role in minimizing fertility. Higher age at first marriage has an adverse effect on high fertility. Early marriage not only marks a woman's entry into a sexual union and the beginning of exposure to childbearing but may also be an important gauge of women's status, since the higher the age of a women at first marriage, the greater the likelihood that she attends school or gets employed, and the greater her chances of having a more equal relationship with her husband.

The result of this study shows that women married to husbands with illiterate, primary and secondary education are likely to have five or more CEB than those married to husbands with higher education. A similar study by Mokshed (2000) also shows that husbands who had secondary or higher education were less likely to desire for additional children than those who had no education. Traditional attitude, beliefs and norms can overcome the education, which may lead the women to have birth until the desire sex compositions of children are met. If a husband would like to have a big family, then a woman has to follow his wish.

According to our findings, religion has been found to have a significant effect on total CEB. Our findings show that women who follow Muslim, Protestant and religions other than Islam, Coptic orthodox and Protestant are more likely to have five or more children in comparison with women who follow the Coptic Orthodox religion.

The age of women is also an important variable that can affect total CEB. Women whose age is between 20-24 are less likely to have five or more children than women whose age is between40-45. A similar study also shows that teenagers whose age was between 18 and 19 years were about eight times more likely to be fertile than teenagers whose age between 15-17 years. As the age increases, the risk of exposure to pregnancy and childbearing also increases, because of higher probability of getting sexual relation and marriage (Tewodros Alemayehu et al 2010).

Women who use contraceptive are found to be a significant variable for total CEB. Women who ever used contraceptives are more likely to have five or more children than women who never used any contraceptives. The study by Ramesh (2010) shows that women who had ever used contraception had more children than those who did not use contraception.

The work status of a woman is another important variable affecting total CEB. The finding of this study shows that women who are working (as wage/salary earners) are less likely to have five or more CEB than women who are not employed as wage/salary earners.

5.2 Recommendations

Based on the findings of the current study, the following recommendations are put forward to bring about reduction in CEB.

1. The level of education of a woman was found to be inversely related with fertility showing the need to educate women. Measures to improve women education must be taken into account by government.
2. Programs should focus on creating awareness about the disadvantages of early marriage.
3. The government should seek ways to empower women economically by producing income-generating schemes and increasing employment opportunities.
4. Awareness creation must be done without consideration of boundaries of religions. It is something that must be done at national level.

REFERENCES

Abdul Hakim, 1994.Factors affecting fertility in Pakistan. The Pakistan Development Review 33(4): 685-709

Agresti, A. 2008. An introduction to categorical data analysis second edition. John Wiley and Son Inc, New York.

Agresti, A., 2002. Categorical data analysis. Second edition, John Wiley and Sons, Inc, publication, New Jersey.

Ainsworth, M. 1989. Socio-economic determinants of fertility in Coted'voire, LSMS working paper No 53.World Bank, Washington DC.

Amaha Haile and **Fikre Enqueselassie 2006**. Influence of women's autonomy on couple's contraception use in Jimma town, Ethiopia. Ethiopian Journal of Health Dev. 20(3): 145-151.

Bewick, V., Cheek, L, and **Ball, J, 2005.** Statistics review 14: Logistic regression. University of Brighton, Australia. Biomed central Ltd. Volume 9:112-118.

Bongaarts J (1978).A framework for analyzing the proximate determinants of fertility Population and Development Review 4 (1):105-132.

Bongaarts J (1999). Fertility Decline in the Developed World: Where Will It End? The American Economic Review, 89(2)

Caldwell J.C 1980. Mass Education as Determinant of the Timing of Fertility Decline. Population and Development Review 6(1) 225-255.

Caldwell J.C 1982. Theory of Fertility Decline. London: Academic Press.

Caldwell J.C and **Caldwell P. 1987**. The cultural context of high fertility in sub-sharan Africa. Population and Development Review 13(3): 409-437.

Chatterjee, S. and **Hadi, A.S 2006.** Regression analysis by example. Fourth edition, John Wiley and Sons, Inc., Hoboken, New Jersey.

Cochrane S.H. 1979. Fertility and Education. Baltimore: The John Hopkins University Press.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Applied multiple regression/correlation analysis for the behavioral sciences. Lawrence Erlbaum Associates: New Jersey.

Cook, R.D., 1998. Regression graphics: ideas for studying regression through graphics. Wiley and Sons, New York.

Cox, D.R. and **Snell, E.J. 1989.**The Analysis of Binary Data, Second Edition, London: Chapman and Hall.

Desta K and **Seyoum G 1998.** Family system in Ethiopia in: Hand book on population and family line education for secondary school teachers in Ethiopia. Edited by Seyoum Gebreselassie and Markos Ezra DTRC and ICDR. Addis Ababa.

Draper, N.R. and **Smith, H. 1998**. Applied Regression Analysis. 3rd ed. New York: John Wiley.

Easterlin, R 1975. An economic framework for fertility analysis. Studies in Family Planning 6(3): 54-63.

Gwatkin, D., S. Rutstein, K. Johnson, R. Pande and **A. Wagstaff, 2000.**Socio-economic differences in health, nutrition and population. Washington D.C. The World Bank

Hakim, A., and **Miller P.C. 1996.** Family Planning in Pakistan: have were ached a turning point? IUSSP Conference Proceedings, Rawalpindi.

Hilbe, J. M, 2009. Logistic regression model. Chapman and Hall, CRC press, New York, page 656.

Hosmer, D.W and **Lemeshow, S., 2001.** Applied logistic regression. 2nd Edition John Wiley and Sons, Inc; New York.

Islam, S. (2009). Differential determinants of birth spacing since marriage to first live birth in rural Bangladesh. Pertanika Journal of social science and humanities 17(1):1-6

John , H.A and **Forrest D.N 1984.** Linear probability, Logit and Probit models. SAGE publications, Inc. U.S.A.

Mason K.O. and **Palan V.T. 1981.**Female employment and fertility in peninsular Malaysia: The maternal role compatibility hypothesis reconsidered. Demography 18(4):549-575.

Maudlin W.P. and **Berelson B. 1978**. Conditions of fertility decline in developing countries, 1965-1975. Studies in Family Planning 9: 89-145.

Mcquillan, K. 2004. When does religion influence fertility? Population and Development Review 30(1):25-56.

Meyers, L.S., Gamst, G. & Guarino, A. J. 2006. Applied multivariate research: design and interpretation. Sage Publications: London

Mokshed, Ali 2000. The effect of selected socio-demographic characteristics on desire for additional children among couples in Bangladesh. Mahidol Universty.

Nag, M 1982. Modernization and its impact on fertility: the Indian Scene. The population council, New York

Nagelkerke, N. J. D. 1991. A note on a general definition of the coefficient of determination. Biometrika, 78, 691-692.

Okpala, A.O. 1989. Female employment and family size among urban Nigerian women. Journal of Developing Areas 23(3):439-456.

Pampel, F.C. 2000. Logistic Regression: A Primer. CA. Sage publications, Thousand Oaks.

Ramesh A. 2010. Demographic, socio-economic and cultural factors affecting fertility differential in Nepal. Geography and Population Department, Tribhuvan Universty, Kathmndul, Nepal.

Samson Gebremedhin and **Mulugeta Betre 2009.** Level and differentials of fertility in Awassa town, southern Ethiopia. African Journal of Reproductive Health 13 (1):93-112.

Sharma S. 1998. Relationship between gender of existing children and desire for additional children by Nepalese women. Master's Thesis, Institute for Population and Social Research, Mahidol University.

Stevenson, M 2008. An introduction to logistic regression. Epicenter, Massey University, Newzeland, page 1-28.

Stycos, J. M. and **Weller R. 1967.** Female working roles and fertility. Demography 4(1):210-217.

Tewodros Alemayehu, Jemal Haider and **DerejeHabte 2010**. Determinants of adolescent fertility in Ethiopia. Ethiop. J. Health Dev. 24(1):30-38.

Toefiqua Mahfuza 2009. Islam, influence of socio-demographic variables on fertility in Bangladesh. Application of path model analysis. Pakistan journal of social science 6(5);

313-320. Africa at the sub-national level. Institute of demography, university of Louvain Belgium.

APPENDIX

Categorical Variables Coding

		Frequency	Parameter coding (1)	(2)	(3)	(4)	(5)	(6)
Age 5-year groups	15-19	865	1.000	.000	.000	.000	.000	.000
	20-24	1714	.000	1.000	.000	.000	.000	.000
	25-29	2184	.000	.000	1.000	.000	.000	.000
	30-34	1646	.000	.000	.000	1.000	.000	.000
	35-39	1564	.000	.000	.000	.000	1.000	.000
	40-44	1159	.000	.000	.000	.000	.000	1.000
	45-49	1067	.000	.000	.000	.000	.000	.000
Husband level of education	Illiterate	5943	1.000	.000	.000			
	Primary	2286	.000	1.000	.000			
	Secondary	1558	.000	.000	1.000			
	Higher	412	.000	.000	.000			
Religion	Coptic Orthodox	4693	1.000	.000	.000			
	Muslim	3581	.000	1.000	.000			
	Protestant	1596	.000	.000	1.000			
	Others	329	.000	.000	.000			
Wife level of education	Illiterate	7433	1.000	.000	.000			
	Primary	1600	.000	1.000	.000			
	Secondary	997	.000	.000	1.000			
	Higher	169	.000	.000	.000			
Ever use of contraceptive	Never used	7437	1.000					
	Ever used	2762	.000					
Wife work status (as employee)	No	7313	1.000					
	Yes	2886	.000					
Age at first marriage	15-17 years	7019	1.000					
	eighteen years and above	3180	.000					
Place of residence	Urban	2369	1.000					
	Rural	7830	.000					

Variables not in the Equation

	Variables	Score	Df	Sig.
Step 0	Residence(1)	342.718	1	.000
	Religion	89.868	3	.000
	Religion(1)	85.918	1	.000
	Religion(2)	49.576	1	.000
	Religion(3)	4.429	1	.035
	Wedu	601.886	3	.000
	Wedu(1)	538.763	1	.000
	Wedu(2)	116.552	1	.000
	Wedu(3)	340.664	1	.000
	Work(1)	15.597	1	.000
	Age	3752.696	6	.000
	Age(1)	570.277	1	.000
	Age(2)	1159.719	1	.000
	Age(3)	527.269	1	.000
	Age(4)	90.884	1	.000
	Age(5)	550.031	1	.000
	Age(6)	646.645	1	.000
	Amarriage(1)	198.950	1	.000
	Hedu	517.608	3	.000
	Hedu(1)	435.687	1	.000
	Hedu(2)	41.382	1	.000
	Hedu(3)	248.193	1	.000
	Contraceptive(1)	97.220	1	.000
	Overall statistics	4577.933	19	.000