34

ADDIS ABABA UNIVERSITY

OFFICE OF GRADUATE PROGRAMS

FACULTY OF SCIENCE

DEPARTMENT OF STATISTICS

CORRELATES OF HARMFUL TRADATIONAL PRACTICES: THE CASE OF WOMEN IN ETHIOPIA

By:
Zemene Yohannes Ayalew

A Thesis submitted to the Office of Graduate Programs of Addis Ababa University in Partial fulfillment of the requirement for the Degree of Master of Science in Statistics

JUNE 2009

ADDIS ABABA UNIVERSITY

OFFICE OF GRADUATE PROGRAMS

FACULTY OF SCIENCE

DEPARTMENT OF STATISTICS

CORRELATES OF HARMFUL TRADATIONAL PRACTICES: THE CASE OF WOMEN IN ETHIOPIA

By:
Zemene Yohannes Ayalew

Approve by the Board of Examiners :

Sileshi Fanta	
Department Head	Signature
Sileshi Fanta	
Examiner	Signature
M.K. Sharma	
Examiner	Signature

Tel. 251-1-239465 **P.O.Box1176 Addis Ababa, Ethiopia**

TABLE OF CONTENTS

ACKNWOLEDGEMENT i

ACRONMY...ii

ABSTRACT..iii

CHAPTER Pages

1. INTRODUCTION...1

1.1 General Background of the Study...1

1.2 Objectives of the Study...4

1.3 Significance of the Study...4

1.4 Statement of the Problem..5

1.5 Limitations of the Study..5

1.6 Desired Outcome...6

1.7 Organization of the Study...6

2.LITRATURE REVIEW..7

3. **DATA AND** METHODOLOGY...**17**

3.1 Source of Data..17

3.2 Description and Importance of Variables Included in the Study..................18

3.2.1 Dependent Variable..18

3.2.2 Independent Variable..19

3.3 Methodology...20

3.3.1 Introduction..20

3.3.2 Logistic Regression Model...22

3.3.2.1 Parameter Estimation for Logistic Regression..............................28
3.3.2.2 Likelihood Function for Logistic Regression..............................29
3.3.2.3 Maximum Likelihood Estimation for Logistic Regression..............30
3.3.2.4 Parameter Inference for Logistic Regression..............................31
3.3.2.5 Selection of Predictor Variables in Logistic Regression.................33
3.3.2.6 Assessing the Goodness of Fit for Logistic Regression..................33
3.3.2.6.1 Likelihood Ratio and Pearson Chi-square Goodness of Fit Test 34
3.3.2.6.2 Deviance Goodness of Fit Test...36
3.3.2.6.3 Hosemer-Lemeshow Goodness-of-Fit Test.............................37
3.3.2.6.4 Classification Table..37
3.3.3 Probit Regression Model 38
3.3.3.1 Parameter Estimation for Probit Regression Model......................39
3.3.3.2 Parameter Inference for Probit Regression Model........................40
3.3.3.3 Selection of Predictor Variables for Probit Regression Model..........40
3.3.3.4 Goodness of fit for Probit Regression Model.............................40
3.3.3.4.1 The Pseudo R-Square for Probit Regression 41
3.3.3.4.2 Likelihood Ratio Goodness of Fit Test for Probit regression.........41
3.3.4 Relationship of Probit and Logistic regression.............................42
3.3.5 Model Adequacy...43
3.3.5.1 Influential Diagnostics...43
3.3.5.2 Detection of Outlier(s) 43
4. STATISTICAL DATA ANALYSIS..44
4.1 Introduction...44
4.2 Summery Statistics 44

4.2.1 Harmful Traditional Practices by Some Scio Economic and Demographic Variables..45

4.3 Logistic Regression..47

4.3.1 Variable Selection for Logistic Regression...............................47

4.3.1. 1 Bivariate Findings..48

4.3.1.2 Analysis of Multiple Logistic Regression 50

4.3.2 Model Checking and Diagnostics for Logistic Regression................51

4.3.2.1 The Likelihood Ratio Test..51

4.3.2.2 Classification Table..52

4.3.2.3 The Hosmer_Lemeshow Test..53

4.3.3 Influential Diagnostic for Logistic Regression...............................54

4.4 Probit Regression..55

4.4.1 Variable selection for Probit Regression......................................55

4.4.2 Bivariate Findings for Probit Regression......................................55

4.4.3 Analysis of Multiple Probit Regression 57

4.4.4 Goodness of Fit for Probit Regression..59

4.4.4.1 The Likelihood Ratio Test..59

4.4.4.2 Classification Table for Probit Regression.................................59

4.5 Probit and Logistic Regression 60

4.6 Interpretation of the Final Model..60

5.CONCLUSION AND RECOMMENDATION.................................62

5.1 Conclusion...62

5.2 Recommendation 64

REFERENCES.. ...66

APPENEDIX 69

ACKNOWLEDGEMENTS

I am very grateful to my advisor Dr. Fentaw Abegaz, for his continuous advice and useful comments during the preparation and finalization of this manuscript.

Many thanks are due to my institution the Central Statistical Agency (CSA) which allowed me to join MSc. program and to utilize computer and other facilities with out limitation.

I am indebted to my mother Sihirte Alamrew who initialized and helped me to continue the graduate study in Statistics. I also extend my gratitude to the entire staff of the Department of Statistics for their assistance in various ways. Finally, I would like to thank my friends in the Department of Statistics for their consistent support.

ACRONYMS

AIC:	Akaike Information Criterion
AIDS:	Acquired Immunodeficiency Syndrome
CSA:	Central Statistical Agency
EA:	Enumeration Area
EGLDA:	Ethiopia Goji Lemadawi Dirgitoch Aswegaji Mahiber
FGM:	Female Genital Mutilation
HIV:	Human Immunodeficiency Virus
ICPD:	International Conference on Population and Development
LRT:	Likely Ratio Test
OLS:	Ordinary Least Square
TV:	Tele Vision
SNNP:	South Nation Nationalities and People
STD:	Sexually Transmitted Diseases

ABSTRACT

This study attempts to identify the correlates of harmful traditional practices on Ethiopian women and to assess the predictive validity of some selected variables such as age, region, place of residence, educational level, ethnicity, religion, frequency of reading news paper or magazine, frequency of watching TV, frequency of listening radio, wealth index, knowledge and attitude on harmful traditional practices. In this study, the data collected by Central Statistical Agency (CSA) for the 2005 Demographic and Health Survey (DHS) were used. The study consists of 14,070 randomly selected women in Ethiopia. Preliminary analysis is carried out through bivariate analysis. Then combined effect of the variables is analyzed using logistic and probit regression models. The result shows that age, region, educational level, religion, place of residence, wealth index, ethnicity, and attitude on harmful traditional practices are significant correlates of harmful traditional practices. In order to decrease or discontinue harmful traditional practices like female circumcision and marriage by abduction, strategies should be designed in a such way that would focus on and address the most important correlates identified in this study.

CHAPTER ONE

INTRODUCTION

1.1 GENERAL BACK GROUND OF THE STUDY

Ethiopia has a great ethnic, religious and cultural diversity. In the country, attitudes towards women's rights are relatively homogenous in rural societies, where harmful traditional practices are the norm. There is an encouraging and a growing international awareness regarding harmful traditional practices as root causes for discrimination and violence against women.

Ethiopian Women constitute about half of the population and play a central role in the socio-economic development of the country. In rural Ethiopia, women work more than 12 hours per day. Most Ethiopian women start life from a position of disadvantage and deprivation. Although there are recent attempts to improve their life, they are still deprived of education, training, job opportunity and decision making in their family or national issues.

Lack of education, inadequate or inexistent health services, and inadequate utilization of reproductive health facilities and lack of basic health concepts, make pregnancy a dangerous process. High maternal death rates in Ethiopia are the unquestionable evidence of lack of education and poor health of women. Ethiopia is one of the developing countries with high child and maternal mortality. Like most societies in developing countries the ethnic cultures in Ethiopia are also affected by the myths, superstitions and conceptions of man. In almost all ethnic groups of the country there are harmful traditional practices which adversely affect the health of women and children. The most common harmful traditional practices on women in Ethiopia are female circumcision, marriage on abduction and early marriage.

Female circumcision persists primarily in Africa and among communities in the Middle East and Asia. According to the 2005 Ethiopia Demography and Health Survey; conducted by Central Statistical Agency it is estimated that 74.3% of women undergone female circumcision. The Survey also indicted that 31.4 % of women believes that practices should be continued among the society. Preliminary analysis of the survey data also indicated that in rural households there were a high percentage of harmful traditional practices than urban households and urban households being three times more likely than rural households to have accessed information on the harmful effects of harmful traditional practices.

Some immediate possible consequences and complications that may follow from female circumcision are hemorrhage, sever pain and in the worst cases death.Long-term complications include chronic pelvic infections, dysmenorrhoea and urinary tract infections. Because pelvic and urinary tract infections are known to cause pathological infertility, it is hypothesized that female circumcision would result infertility. It also contributes to the high maternal mortality rate.

Marriage by Abduction is another harmful traditional practice that is known to exist in Ethiopia. The Ethiopia Demography and Health Survey conducted in 2005 by Central Statistical agency indicated that eight percent of women reported that they had been married by abduction. This is most commonly reported that women in the age group15-19, rural women and women residing in SNNP and Oromiya were highly exposed to marriage by abduction. It is also indicated that this practice is the least common among the wealthiest group of women.

Marriage by Abduction is gender-based violence which carry with them pain, disability and shame. Marriage by Abduction has sever psychological, emotional, medical, financial, legal consequences, and the victims are more susceptible to sexually transmitted diseases (STDs) including HIV/AIDS. Victims tend to be isolated from their peers and friends. This deprives them of their right to education, as well as limits any possibility of economic independence from their spouse, making it more difficult to escape from an unwanted marriage.

According to Demography and Health Survey conducted by CSA in 2005, many different factors associated with the practice of harmful tradition practices by women are investigated. Therefore, it becomes important to assess the major factors that cause the practice of harmful traditional practices by women. That is why this study has taken to work on it. The result of the study will be useful in providing information on harmful traditional practices for researchers and policy makers. Also, the reduction of harmful traditional practices implies women empowerment, which is important for national or regional development. It becomes easier to design effective strategies and polices if factors associated with the practice of harmful traditional practices by women are well understood.

1.2 OBJECTIVES OF THE STUDY

The general objective of this study is to investigate the most important correlates of harmful traditional practices on women in Ethiopia.

More specifically, the study attempts:

- To examine the relationship between demographic, socio-economic and health related factors on attitude and exercising harmful traditional practices.
- To identify the most important variables that describes harmful traditional practices.
- To fit logistic and probit regression models in the relationship between harmful traditional practices and the explanatory variables.

1.3 SIGNIFICANCE OF THE STUDY

Harmful traditional practices is a serious problem in our country. In order to design and implement effective prevention strategies, it is better to understand the nature of this problems. Thus, the aim of the study is to meet this need, that is, to provide valuable information, recommendations and suggestions for educational planners and policy makers so as to reduce and alleviate the problem. It is hoped that these suggestions could help educational planners and policy makers to reevaluate their approach towards the planning of strategies to reduce or discontinue harmful traditional practices at the national level. It is also hoped that the study will stimulate researchers to further investigation of related problems.

1.4 STATEMENT OF THE PROBLEM

Even though the degree varies from country to country, harmful traditional practices is a common phenomenon in all countries of the world. A recent study conducted by Ye Ethiopia Goji Limadawi Dirgitoch Aswegaj Mahiber (EGLDAM) reported that developing countries suffer from high rate of harmful traditional practices. Harmful traditional practices have a considerable effect on individuals and society at large. Female circumcision is a deep-rooted traditional practice that has serious health related problems to girls and women. Individual with the practice of female circumcision is subjected to life-long physical and emotional problems. Another traditional practice with serious health risk for girls and women is marriage by abduction.

Because many females are suffering from different harmful traditional practices, this problem is considered to be a serious issue in our country, Ethiopia. These harmful traditional practices in the country poses series difficulties on the country's efforts to achieve social and economic developments. It is this awareness of the adverse effect of harmful traditional practices that initiated us to study correlates of these practices.

1. 5 LIMITATIONS OF THE STUDY

Since an objective of this study was to assess correlates of harmful traditional practices on Ethiopian women, it was originally planned to collect information on many of the harmful traditional practices than those actually included in this study. For the difficulty of getting the required information on harmful traditional practices in the country, the scope of this study is limited to examining those harmful traditional practices for which information was available from 2005 Ethiopian Demographic and health Survey.

1.6 DESIRED OUTCOME

The desired outcome is to provide necessary empirical evidence regarding those most powerful variables selected as correlates of harmful traditional practices, there by enabling policy makers to give alternative information. It is expected that the suggestions and recommendations given in the study will motivate other researchers to further findings of such problems.

1.7 ORGANATIZATION OF THE STUDY

Chapter II deals with review of related literature. In this chapter, summaries of works cited in the past either in or outside of Ethiopia on harmful traditional practices of women are described.

In chapter III, methods and procedures of sampling and data collection are presented. The variables to be included in the study with their coding and designations, and methods of data analysis are also described in this chapter.

In chapter IV, the relationship of each explanatory variable on harmful traditional practices will be examined. And then the combined effects of the explanatory variables on harmful traditional practices is investigated using logistic and probit regression analysis in this chapter.

Finally, in chapter V results of the whole work are discussed. Moreover; some valuable suggestions, recommendations and their implications are provided.

CHAPTER TWO

LITRATURE REVIEW

A large body of empirical research has identified a wide range of factors that are associated with harmful traditional practices. In most of the studies logistic regression analysis (using maximum likelihood method of estimation) was applied to examine the effects of socio-demographic characteristics on harmful traditional practices. In this respect, an attempt is made to summarize findings of past studies and to give an overall view regarding the impact of selected characteristics on harmful traditional practices.

Esthero et al. (2008) examined the level of practice of female circumcision and the impact of a health education intervention in Shao community (Nigeria). In this study, factors that significantly associated with the practice of female circumcision were age, gender and educational status of respondents ($p < 0.05$). The age at which female circumcision is usually performed was put at under one year old by 60.3% of respondents. Most (88.0%) of the female respondents were excise. Attitudinal changes were also investigated under pre-intervention and post-intervention stages of the study. A greater proportion of men than women did not want the practice of female circumcision to be stopped in the pre-intervention stage; however, there was a statistically significant decrease in the proportion of males who did not want the practice of female circumcision to be stopped in the post-intervention stage. In addition, there was a significant increase in the proportion of respondents who had no intention to excise female children in the future during the post-intervention stage ($p < 0.05$). Legislation, female literacy and empowerment, educating men and provision of alternative vocation for excise were among the suggested ways by researchers to bring an end for such practice. The authors recommended that, health education intervention had a positive impact on the attitude of respondents towards female circumcision. However, for sustainable behavioral changes that will lead to elimination of female circumcision practice, they recommended that placing female circumcision elimination efforts within a comprehensive development strategy and the larger context of reproductive health and gender education in Nigeria.

A study conducted by Gage and Rossen (1999) focused on socio economic correlates and gender differences in attitudinal support for the discontinuation of female circumcision in Guinea. Data from structured interviews of men aged 15–59 and women aged 15–49 years in the 1999 Demographic and Health Survey were taken and multiple logistic regression methods were used to examine the relationship of socioeconomic factors and gender to attitudinal support for the discontinuation of female circumcision. The study showed that; more than 90% of women had undergone female circumcision. Attitudinal support for female circumcision discontinuation was more prevalent among men than women. The odds of supporting the discontinuation of female circumcision were negatively related to beliefs in social approval and religious support for female circumcision and its enhancement of women's marriageability. They concluded that community education, improvements in women's socioeconomic status and traditional and religious leader involvement would be critical for female circumcision eradication.

The study on the practice, knowledge and attitude towards circumcision was done in two private schools in Khartoum on female students who have passed the usual age for circumcision by Bashir (1993). The results obtained revealed relatively high percentage (78%) of female students was not circumcised. Most of the reasons given for non-practicing were that circumcision was harmful to health and that it was a useless old custom. In contrast, families who practice claimed a religious basis for circumcision and held on it as a Sudanese tradition of value. Therefore, conceptions, knowledge and attitude of families rather than the socioeconomic conditions had the main effect in making decision.

A study on prevalence and epidemiology of female circumcision in Sudan was conducted by El-Dareer (1982). The sample was made up of 10,000 females and 5000 males. Multistage random sampling was done with the household as a unit. Most respondents were 15-24 years (37.5%) and 25-34 years (36%), 50% lived in rural and 50% in urban areas, 71% of the subjects were uneducated; 17% were of elementary level; all mothers were illiterate and Muslim. The average age for circumcision was 7 years with the youngest 2 years and the eldest 11 years. 81% of cases were operated upon by midwives and the main instrument was a razor. Indigenous midwives did not use anesthesia. The method of approximation in 38% of the cases was binding the lower limbs together, and in 21% by applying a mixture of egg and sugar. The decircumcision in itself may cause bleeding and infection and needs to be closed again. 85% of women interviewed had been afraid of the first sexual intercourse and 92% of them were due to circumcision. Of the 348 married females, 80% underwent recircumcision mainly after delivery though not for each delivery. Most operations were carried out by midwives at home; only less than 1% were done in the hospital. 81% agreed with circumcision giving their main reason as religious requirements. Of the 195 who rejected circumcision, almost all suggested health education for stopping this practice.

The study conducted by Almroth et al. (2001) focused on male complications and attitudes with regard to female circumcision. A village in the Gezira Scheme along the Blue Nile in Sudan constituted the basis of the study. Interviews were carried out according to a pretested questionnaire, using structured questions with open-answer possibilities. Married men of the youngest parental generation and grandfathers were randomly selected from up-to-date election lists. A total of 59 men were interviewed, 29 young men and 30 grandfathers. The younger respondents were preferred to accept a woman without circumcision to become their daughter-in-law. A majority of the young men were preferred to marry a woman without female circumcision. This proportion was significantly higher than among the grandfathers. Female circumcision can no longer be considered to be only an issue for women. The acknowledged male complications and attitudes described may open new possibilities to counteract the practice of female circumcision.

Similarly, Islam and Uddin (2001) had conducted their research on women's attitudes towards female circumcision in Sudan. Information about attitudes toward the practice, the reasons why women support it and the social and demographic predictors associated with support for it were taken from approximately 1,000 ever-married women. A multivariate analysis was performed to find factors associated with the eradication of female circumcision. Education and economic empowerment of women would help to lower the support for the practice. More highly educated and economically better women were less likely to support continuation of the practice.

On the other hand, the study by Dirie, MA and Lindmark,G (1991) that involved two hundred and ninety Somalian women. The majority of these women justified the practice of female circumcision with religious reasons and all were willing to circumcise their daughters. Fifty-two percent of the respondents had been operated on by medically untrained persons, usually traditional birth attendants where the majority operated at home.

A study by Okemgbo(2002) focused on prevalence, patterns and correlates of domestic violence in selected Igbo communities of Imo State, Nigeria. Three hundred and eight Igbo women were randomly selected to respond to a number of questions on experiences, patterns and attitudes to female circumcision. Data was collected using structured questionnaires that were complemented with focus group discussions. The study revealed that the practice of female circumcision was this Igbo population, and 52.6% were of the view that it should be continued. Multiple logistic regression identified the correlates of these forms of violence such as age, place of residence, age at first marriage, type of marital union, level of income of women, and level of education of husband against women in Imo State. They recommend education of women and integration of services in reproductive health care service delivery as appropriate measures to eradicate these practice.

Snow et al. (2002) studied the prevalence and social determinants of female circumcision. The sample included 1, 709 women from three family planning clinics in south west Nigeria. They mentioned that ethnicity was the most significant social predictors of female circumcision, followed by age, religious affiliation and education. In total, 45.9% had undergone some form of cutting. There was evidence of a steady and steep secular decline in the prevalence of female circumcision in this region over the past 25 years, with age-specific prevalence rates of 75.4% among women aged 45-49 years, 48.6% among 30-34-year olds, and 14.5% among girls aged 15-19. Despite wide disparities in female circumcision prevalence across ethnic, religious and educational groups, the secular decline was evident among all social subgroups.

Using data from a national representative survey of adolescents, Omaima et al .(2002) investigated the prevalence and social correlates of circumcision among girls aged 10–19, the circumstances surrounding the procedure, and the attitudes of adolescents towards it. A life table analysis indicated that girls were 10 percentage points less likely to undergo female circumcision than their mothers. Circumcision may had begun to decline prior to the time when the current cohort of girls were at risk; however, the data hint at a temporal association between the decline and the time when the campaign against circumcision gained momentum as reflected in the 1994 International Conference on Population and Development (ICPD) in Cairo. Over half of circumcised girls reported that the procedure was performed by a physician or nurse rather than a traditional practitioner. A multivariate analysis indicated that girls who had been in school, who lived in urban areas, and who were older are more likely to believe that circumcision is not obligatory. When the analysis included boys as well as uncircumcised girls, a large gender gap emerges, with boys considerably more supportive of the practice than were their female counterparts on the decline of female circumcision in Egypt.

The prevalence and major correlates of Female Genital Mutilation (FGM) in Ghana's Kassena-Nankana district were assessed in three surveys conducted in

1995-96 by Mbacke et al. (1998). The surveys (a demographic panel survey, a study of pregnant women and a school-based survey) were part of a research program initiated by the Navrongo Health Research Center to provide a basis for the development, implementation, and monitoring of an intervention to eradicate the practice in the study area. 77.1% of the 5285 women 15-49 years of age included in the panel survey reported they had undergone female genital mutilation, generally at 15-19 years of age. Multiple logistic analysis indicated significantly lower rates of female genital mutilation among Christians and Muslims compared to those who practice traditional African religions, ethnic groups that contain mainly immigrants, single women, and women with an education beyond the primary school level. The relatively late age at female circumcision in Kassena-Nankana offers a window of opportunity in which to reach adolescent girls and discuss the practice before they undergo it.

Another study on the perceptions, degree of knowledge, attitudes and practices of the primary healthcare professionals in relation to female circumcision was done by Adriana et al.(2009). A transversal, descriptive study was performed with a self-administered questionnaire to family physicians, pediatricians, nurses, midwives and gynecologists. Trends towards changes in the two periods studied (2001 and 2004) were analyzed. The result showed that a total of 225 (80%) professionals answered the questionnaire in 2001 and 184 (62%) in 2004. Less than 40% correctly identified the typology, while less than 30% knew the countries in which the practice is carried out and 82% normally attended patients from these countries. They recommended preventive measures should be taken in primary healthcare medical offices with pediatricians and gynecologists designed to promote stands against these practices.

Suzuki and Meekers (2006) examined the effect of exposure to communication messages on support for female circumcision in Egypt. Data from the 2005 Egypt Demographic and Health Survey were analyzed using logistic regression on 19,106 ever-married women aged 15-49. The analysis revealed that high levels of exposure to female circumcision-related media messages were essential in reducing support for female circumcision. Women exposed to two or more female circumcision media messages were 1.6 times more likely than unexposed women to support discontinuing female circumcision. Moreover, women's belief that men want the practice discontinued and their belief that female circumcision can cause fatal complications, were both positively associated with women's support for discontinuing female circumcision. By contrast, women's belief on female circumcision and their belief that female circumcision prevents adultery were negatively associated with women's support for discontinuing female circumcision.

In addition, the above study shown that almost identical factors found to affect women's intention to circumcise their daughters. High exposure to female circumcision-related messages, and women's belief that men want the practice discontinued, were negatively associated with the outcome. Women's belief that a husband prefers a circumcised wife was not associated with women's intention to circumcise their daughters after controlling for all other variables in the model. Given that high level of exposure to female circumcision-related messages was key in reducing support for female circumcision, communication campaigns should be continued and intensified. Campaigns should also aim to change men's perception and support for the practice.

A community based cross sectional study was conducted in a rural district of North West Ethiopia between February and April 1997 to determine the magnitude of marriage through abduction ('Telefa') and identify problems associated with it. Randomly selected and currently married 1,168 women were interviewed. The prevalence of marriage through abduction was 6.2%. All the abductions reported were only once in lifetime during the first marriage. The median age at first marriage of abducted women was 13 years with a range of 7 to 20 years. About two third (66.7%) of abducted women had been married more than once in their lifetime. The recognition of the magnitude and the associated health problems of marriage through abduction (Telefa) were important. Appropriate strategies that address the health needs of abducted women must be designed. Enforcing the judiciary system to discourage this harmful practice and empowerment of young girls and rural women was needed.

A study on Knowledge, Attitudes and Practices on traditional harmful health in Dembia District, northwest Ethiopia was done by Getu and Melkie (2001.) Data were collected from 1181 households using a pre-tested questionnaire. Educational status and religion were found to be significantly associated with the attitude of respondents towards practicing the prevailing traditional malpractices of the study area ($P < .001$ for each factor). An integrated health activity which includes the issue of traditional harmful health practice and its associated risks should be given due attention at grass roots level. The practice of the most serious traditional harmful health practices like female circumcision should be condemned.

CHAPTER THREE

DATA AND METHODOLOGY

3.1 Source of data

The data on which this research used are the 2005 Demographic and Health Survey (DHS) conducted by the Central Statistical Agency of Ethiopia (CSA). This survey was designed to collect information for the purpose of policy decisions, planning, monitoring and evaluating of programs on health in general and reproductive health in particular at both national and regional levels. Moreover, this is the second survey of its kind in the country to provide national and regional estimates on population and health and comparable to similar surveys in other developing countries. This survey collected the demographic and health information from a nationally representative sample of women and men in the reproductive age groups 15-49 and 15-59, respectively. The survey was conducted in selected rural and urban areas throughout the country.

The sample for the survey is based on a two-stage stratified systematic sample of households. At the first stage of sampling 540 enumeration areas (EAs), 145 in the urban areas and 395 in the rural areas, were selected using systematic sampling with probability proportional to size.

A complete household listing operation was carried out in the selected EAs to provide a sampling frame for second-stage selection of households. At the second-stage of sampling, a systematic sample of 28 households per EA was selected in all the regions to obtain reliable estimates of key demographic and health variables.

The survey was designed to obtain completed questionnaire of interviews to 14,717 women of age 15-49. In addition, all males of age 15-59 in every second household were interviewed, to obtain a target of 6, 778 men interviews. In order to take non-response into account, a total of 14, 645 households from all over the country were selected. Of the 14,645 households selected, interviews were completed for 13,721 households, 14,070 women of age 15-49, and 6,033 men of age 15-59.

This research used data from this survey for the analysis of correlates of harmful traditional practices on Ethiopian women.

3.2 Description and Importance of Variables Included in the Study

Describing the study variables is very important to understand the application of the methodology and to make a proper usage of results. Here the dependent variable is characterized as dichotomous type (assuming zero and one) and the explanatory variables are qualitative and quantitative.

3.2.1 Dependent variable

The harmful traditional practices considered by this study are female circumcision and marriage by abduction. In order to analyze the effect of explanatory variables on the probability of women who have faced harmful traditional practices such as female circumcision and marriage by abduction, logit and probit models are adopted.

The dependent variable used in this study is a binary variable such that:

$Y_i = 1$; if the i^{th} women faced harmful traditional practices,

$Y_i = 0$; otherwise.

3.2.2 Independent variable

The variables that are assumed to be the correlates of harmful traditional practices are presented in table below:

Table 3.1: Independent Variables

No.	Description and Name	Categories
1	Region	(0) Tigray
		(1) Afar
		(2) Amhara
		(3) Oromia
		(4) Somali
		(5) Ben-Gumz
		(6) SNNP
		(7) Gambela
		(8) Harari
		(9) Addis Ababa
		(10) Dire Dawa
2	Age	continuous
3	Place of Residence	(0) Urban
		(1) Rural
4	Educational Level	(0) No education
		(1) Primary
		(2) Secondary
		(3) Higher
5	Religion	(0)Christian
		(1) Muslim
		(3) Others
6	Ethnicity	(0)Affar
		(1)Amara
		(2)Guragie
		(3)Oromo
		(4)Somalie

		(5)Tigraiy
		(6)Other Ethiopian Nationalities
7	Frequency of reading newspaper or magazine	(0)Not at all
		(1)Less than once a week
		(2)At least once a week
		(3)Almost Every day
8	Frequency of listening radio	(0)Not at all
		(1)Less than once a week
		(2)At least once a week
		(3)Almost Every day
9	Frequency of watching Tv	(0)Not at all
		(1)Less than once a week
		(2)At least once a week
		(3)Almost Every day
10	Wealth Index	(0)Poorest
		(1)Poorer
		(2)Middle
		(3)Richer
		(4)Richest
11	Knowledge on harmful traditional practices	(1) Have knowledge
		(2) Don't have knowledge
12	Attitude on the continuation of harmful traditional practices	(1) Continued
		(2) Discontinued
		(3) Indifferent

3.3 Methodology

3.3.1 Introduction

In the analysis of logistic and probit regression with a binary response data, it is important to predict whether an event will or will not occur and identifying the variables useful in the prediction.

There are different types of statistical models that can be used to predict a binary dependent variable from a set of independent variables. Amongst them multiple linear regression, discriminate analysis, the binary logistic regression and probit regression are related techniques that can be possibly used for analysis of binary response data. However, multiple linear regression and discriminate analysis techniques bring difficulties when the dependent variable can have only two values (an event occurring or not occurring, present or absent and success or failure). When the dependent variable can have only two values, the necessary hypothesis testing in regression is violated and it is unreasonable to assume that the error is normal. Since the assumption of normality for the error term is violated, multiple linear regressions cannot be used to predict a binary dependent variable from a set of independent variables.

Even if linear discriminate analysis does allow direct prediction of group membership, the assumption of multivariate normality of the independent variables as well as equal variance-covariance matrices in the two groups is required for the prediction rule to be optimal (Maddala G.S, 1992). Because it is usually unable to meet these assumptions, it is unreasonable to use discriminate analysis for binary response data. Thus, this study uses the binary logistic regression and probit regression models to predict a binary dependent variable from a set of independent variables.

Binary data

When the response variable is dichotomous, it is convenient to denote one of the outcomes as success and the other as failure. For example, if a patient is cured of a disease, the response is `success', if not, then the response is `failure'; if an item passes the quality control, the response is `success', if not, then the response is `failure'; if a mouse dies from toxic exposure, the response is `success', if not (*i.e.* if it survives) the response is `failure'. It is standard to let the response variable Y be a binary variable, which attains the value 1, if the outcome is `success', and 0 if the outcome is `failure'.

Odds

The odds of some event happening (*e.g.* the event that Y=1) is defined as the ratio of the probability that the event will occur divided by the probability that the event will not occur, i.e. the odds of the event E is given by:

$$\text{Odds }(E) = \frac{P(E)}{P(not\ \ E)} = \frac{P(E)}{1-P(E)}$$

3.3.2 Logistic Regression Model

Logistic regression estimates the probability of success over the probability of failure. The results of the analysis are in the form of an odds ratio. In this regression analysis, the dependant variables may be any type, but continuous variables are not used as dependent variables in logistic regression. In logistic regression, no assumptions are made about the distributions of the explanatory variables.

There are three types of logistic regression, namely Binomial (or binary), Multinomial, and ordinal logistic. Binomial (or binary) logistic regression is a form of regression which is used when the dependent is a dichotomy and the independents are of any type. Multinomial logistic regression handles the case of dependent variables with more than two classes of responses; it is also sometimes adopted for binary response data analysis. For multiple ranked classes of the dependent variable, ordinal logistic regression is preferred to multinomial logistic regression. This study uses a binomial logistic regression.

For binary data, we are interested in analyzing the relationship between the explanatory variables and the probability of the response being success, rather than analyzing the relationship between the value of the response variable and the explanatory variables. The binary logistic regression model describes the relationship between a dichotomous response variable Y, coded to take the values 1 or 0 being `success' and `failure'; respectively for k explanatory variables $x_1, x_2 ..., x_k$. The explanatory variables can be quantitative or indicator variables referring to the levels of categorical variables. Since Y is a binary variable, it has a Bernoulli distribution with parameter P=P(Y=1); i.e., P is the probability of success for given values $x_1, x_2 ..., x_k$ of the explanatory variables. Recall that, for a Bernoulli variable, the mean is given by E[Y] =P(Y=1) =P.

The multiple logistic regression model is defined as follows. Suppose that $Y_1, Y_2, ..., Y_n$ are independent Bernoulli variables, and let P_i denote the mean value of Y_i, that is, P_i= E[Y_i] =P(Y_i=1). The mean value P_i can be expressed in terms of the explanatory variables $x_{1i}, x_{2i}, ..., x_{ki}$ as

$$P(Y_i = 1) = E(Y_i) = \frac{\exp(\beta_0 + \beta_1 X_{1_i} + ... + \beta_k X_{k_i})}{1 + \exp(\beta_0 + \beta_1 X_{1_i} + ... + \beta_k X_{k_i})} \quad ....(1)$$

Again, the parameter β_j determines the effect of X_{ji} on the log of odds that Y_i=1, controlling for other X's. Furthermore, exp (β_j) is the multiplicative effect on the odds of unit increase in X_{ji}, at fixed levels of other X's. If we apply the logit-transformation to equation (1), we get a linear relationship between logit P_i

and the explanatory variables. In addition, the transformed quantity $\ln(\frac{P_i}{1-P_i})$ assumes values in the interval (-∞, ∞).

Thus,

$$Logit\ P_i = \ln(\frac{P_i}{1-P_i}) = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki} \quad \text{..................} \ (2)$$

Equation (2) is sometimes called the logit form of the model. Note that, logit P_i is the log odds (the logarithm of the odds) of success for the given values $x_{1i}, x_{2i}, ..., x_{ki}$ of the explanatory variables. The logistic transformation of the binomial probabilities is not the only transformation available, but it is the easiest to interpret.

Assumptions in Logistic Regression Model

Logistic regression is popular because it enables the researcher to overcome many of the restrictive assumptions of Ordinary Least Square (OLS) regression:

a. Logistic regression does not assume a linear relationship between the dependents and the independents. Because the logit link function on the left-hand side of the logistic regression equation is non-linear, it may handle nonlinear effects even when exponential and polynomial terms are not explicitly added as additional independents. However, it is also possible and permitted to add explicit interaction and power terms as variables on the right-hand side of the logistic equation as in OLS regression.

b. The dependent variable need not be normally distributed but does assume its distribution is within the range of the exponential family of distributions, such as Poisson, binomial, gamma. Solutions may be more stable if predictors have a multivariate normal distribution.
c. The dependent variable need not be homoscedastic for each level of the independents; that is, there is no homogeneity of variance assumption(variances need not be the same within categories).
d. Normally distributed error terms are not assumed.
e. Logistic regression does not require that the independents be interval.
f. Logistic regression does not require that the independents be unbounded

However, the following assumptions are still applied to logistic regression:

I. **Data level**: a dichotomous or polytomous dependent variable is assumed for binary or multinomial logistic regression, respectively. OLS regression is preferred to logistic regression for continuous dependent variables when OLS assumptions meet and superior to categorizing a continuous dependent for the purposes of running a logistic regression. Likewise, discriminant function analysis is more powerful than binary logistic regression for a binary dependent when the assumptions of the former met. If the categories of the dependent are ordinal, ordinal regression will be more powerful and is preferred. Independent variables may be interval or categorical if categorical, it is assumed that they are dummy or indicator coded variables.
II. **Meaningful coding:** logistic coefficients will be difficult to interpret if not coded meaningfully. The convention for binomial logistic regression is to

code the dependent class of greatest interest as 1 ("the event occurring") and the other class as 0 ("the event not occurring), and to code its expected correlates also as +1 to assure positive correlation. Logistic regression is predicting the log odds of being in the class of greatest interest. The "1" group or class of interest is sometimes called the target group or the response group.

III. **Proper specification:** the proper specification of the model is in particular crucial. Parameters may change magnitude and even direction when variables are added to or removed from the model.

   a. **Inclusion of all relevant variables in the regression model:** if relevant variables are omitted, the common variance they share with included variables may be wrongly attributed to those variables or the error term may be inflated.

   b. **Exclusion of all irrelevant variables:** if causally irrelevant variables are included in the model, the common variance they share with included variables may be wrongly attributed to the irrelevant variables. The more the correlation of the irrelevant variable(s) with other independents, the greater the standard errors will be in the regression coefficients for these independents.

IV. **Low error in the explanatory variables:** ideally assumes low measurement error and no missing cases.

V. **Linearity:** logistic regression does not require linear relationships between the independent factor or covariates and the dependent, as OLS regression does, but it assumes a linear relationship between the independents and the log odds (logit) of the dependent. When the assumption of linearity in the logits is violated, logistic regression will underestimate the degree of relationship of the independents to the dependent and will lack power (i.e., generating Type II errors, thinking there is no relationship when there actually is). One strategy for justifying lack of linearity in the logit of a continuous covariate is to divide it into categories and use it as a factor, thereby getting separate parameter estimates for various levels of the variable.

VI. **No outliers**: as in OLS regression, outliers can affect results significantly. We should analyze standardized residuals for outliers and consider removing them or modeling them separately. Standardized residuals greater than 2.58 are outliers at the .01 level, which is the customary level (standardized residuals greater than 1.96 are outliers at the less used .05 level).

VII. **Large samples**: logistic regression uses maximum likelihood estimation (MLE) rather than ordinary least squares (OLS) to derive parameters. MLE relies on large-sample that is the reliability of estimates declines when there are few cases for each observed combination of independent variables. That is, in small samples one may get high standard errors. In the extreme, if there are too few cases in relation to the number of variables, it may be impossible to converge on a solution. Very high parameter estimates (logistic coefficients) may signal inadequate sample size. As a rule of thumb, Peduzzi et al. (1996) recommend that the smaller of the classes of the dependent variable have at least 10 events per parameter in the model. Hosmer & Lemeshow (1989) recommend a minimum of 10 cases per independent variable. Pedhazur (1997) recommends sample size be at least 30 times the number of parameters being estimated.

Uses of Logistic Regression Model

Logistic regression can be used to:

1) To predict a dependent variable on the basis of continuous and/or categorical independents,
2) To provide knowledge of the relationships and strengths among the variables.

3.3.2.1 Parameter Estimation for Logistic Regression

Logistic regression applies maximum likelihood estimation after transforming the dependent into a logit variable (the natural log of the odds of the dependent occurring or not). In this way, logistic regression estimates the odds of a certain event occurring. That is, logistic regression calculates changes in the log odds of the dependent, not changes in the dependent itself as OLS regression does. Maximum likelihood estimation involves finding the value(s) of the parameter(s) that give rise to the maximum likelihood. The estimated logistic coefficients $\hat{\beta}_j$ is the change in the log-odds for every unit increase or decrease depending on the variable change in the jth explanatory variable (x_j) holding other predictors constant. The exponential of estimated logistic coefficients denoted by e^{β_i} where i=1,2,3,…,n (n is the sample size) and it represents the multiplicative factor by which the odds change for every unit change in x_j controlling for the other predictor variables.

3.3.2.2 Likelihood Function for Logistic Regression

Since each Y_i (the *i*th response variable) observation is an ordinary Bernoulli random variable, we can represent its probability distribution as

$$f_i(y_i) = P_i^{y_i}(1-P_i)^{1-y_i}, y_i = 0,1, i = 1,2,...,n \quad .... (3)$$

Where n is the sample size and p_i is the $\Pr(Y_i = 1)$

Since the observations are independent, their joint probability function is:

$$g(y_1, y_2, ..., y_n) = \prod_{i=1}^{n} f_i(y_i)$$

$$= \prod_{i=1}^{n} P_i^{y_i}(1-P_i)^{1-y_i} \quad .... (4)$$

Again, it will be easier to find the maximum likelihood estimates by working with the logarithm of the joint probability function:

$$\ln g(y_1, y_2, ..., y_n) = \ln \prod_{i=1}^{n} P_i^{y_i}(1-P_i)^{1-y_i}$$

$$= \sum_{i=1}^{n}\left[y_i \ln\left(\frac{P_i}{1-P_i}\right)\right] + \sum_{i=1}^{n} \ln(1-P_i) \quad .... (5)$$

Since E $(Y_j) = P_j$ for a binary variable, it follows that

$$1-P_i = [1+\exp(\beta' X_i)]^{-1}$$

Hence, the logarithm of the likelihood function can be expressed as:

$$\ln g(y_1, y_2, ..., y_n) = \sum_{i=1}^{n} y_j(\beta' X_j) - \sum_{i=1}^{n} \ln(1+\exp(\beta' X_i)) \quad .... (6)$$

given the sample observations.

3.3.2.3 Maximum Likelihood Estimation for Logistic Regression

Maximum Likelihood Estimation is the method used to calculate the logit coefficients. Maximum likelihood methods seek to maximize the log likelihood, LL, which reflects how likely the odds that the observed values of the dependents may be predicted from the observed values of the independents. OLS can be seen as a subtype of maximum likelihood for the special case of a linear model characterized by normally distributed disturbances around the regression line, where the regression coefficients computed maximize the likelihood of obtaining the least sum of squared disturbances. When error is not normally distributed or when the dependent variable is not normally distributed, maximum likelihood estimates are preferred to OLS estimates because they are unbiased beyond the special case handled by OLS.

The maximum likelihood estimates of β in logistic regression model are those values of β that maximize the log-likelihood function in equation (6). No closed form solution exists for the values of β in equation (6) that maximize the log-likelihood function. Computer intensive numerical search procedures are therefore required to find the maximum likelihood estimates of β. We shall rely on standard statistical software programs specifically designed for logistic regression to obtain the maximum likelihood estimates of β.

And an approximate $(1-\alpha)$ 100% confidence interval for βj is given by:

$$\hat{\beta}_j \pm Z_{(1-\alpha/2)} \times S.e.(\hat{\beta}_j)$$

3.3.2.4 Parameter Inference for Logistic Regression

The Wald statistic is an alternative test which is commonly used to test the significance of individual logistic regression coefficients for each independent variable (that is, to test the null hypothesis in logistic regression that a particular logit (effect) coefficient is zero). The Wald test corresponds to significance testing of β coefficients in OLS regression. We may well want to drop independents from the model when their effect is not significant by the Wald statistic. The effects of each predictor variable in explaining the outcome variable was separately made by testing the null hypothesis that $H_0 : \beta_j = 0$ against $H_1 : \beta_j \neq 0$.

The Wald test statistic is given by :

$$\chi^2_{wald} = \left[\frac{\hat{\beta}_j}{s.e(\hat{\beta}_j)} \right]^2 ,$$

we reject the null hypothesis if $\chi^2_{wald} > \chi^2_{\alpha}(1)$

Each Wald statistic is compared with a $\chi 2$ distribution with 1 degree of freedom. Wald statistics are easy to calculate but their reliability is questionable, particularly for small samples. For data that produce large estimates of the coefficient, the standard error is often inflated resulting in a lower Wald statistic, and therefore the explanatory variable may be incorrectly assumed unimportant in the model. Likelihood ratio tests are generally considered superior. The constant term in the model (β_0) has no simple practical interpretation but it is generally retained in the model irrespective of its significance.

We can test the above hypothesis using the likelihood ratio test, the test statistic is: $\chi^2 = -2\ln\left[\dfrac{L(R)}{L(F)}\right]$ which is distributed as chi-square with 1 degrees of freedom where L(R) is the restricted likelihood under the null hypothesis and L(F) is the full likelihood.

For each fitted model, the likelihood ratio test enables us to test whether some predictor variables can be dropped from the model. The hypothesis is:

$$H_0 : \beta_q = \beta_{q+1} = \ldots = \beta_{p-1} = 0$$

$$H_1 : \text{Not all of the } \beta_k \text{ in } H_0 \text{ equal to zero}$$

are tested by first fitting the full model

$$P(Y_i = 1) = E(Y_i) = \frac{\exp(\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki})}{1 + \exp(\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki})}$$

Obtaining the maximum likelihood estimates β_F, and evaluating the likelihood function at $\beta = \beta_F$. This likelihood value is denoted by $L(F)$:

$$L(F) = L(b_0, b_1, \ldots, b_{k-1}).$$

Next the reduced model

$$P(Y_i = 1) = E(Y_i) = \frac{\exp(\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki})}{1 + \exp(\beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki})}$$

is fitted. The maximum likelihood estimates are obtained and the likelihood function is evaluated at $\beta = \beta_R$. This likelihood value is denoted by $L(R)$:

$$L(R) = L(b_0, b_1, \ldots, b_{q-1})$$

The ratio of the two likelihoods is the likelihood ratio: $\dfrac{L(R)}{L(F)}$. If the null hypothesis is true and the sample size is sufficiently large, the quantity -$2\,Ln\left[\dfrac{L(R)}{L(F)}\right]$ has a chi- square distribution with k degrees of freedom, where k is the difference between the number of parameters in the full model and reduced model. We reject the null hypothesis if $\chi^2_{wald} > \chi^2_{\alpha}(k)$.

3.3.2.5 Selection of Predictor Variables in Logistic Regression

In logistic regression as in other multivariate statistical techniques, one may want to identify subsets of independent variables that are good predictors of the dependent variable. All the problems associated with variable selection algorithms in linear regression and discriminate analysis are found in logistic regression as well.

The logistic regression procedures have several methods available for variable selection. Variables can be entered into the model using stepwise selection, forward selection and backward elimination selection procedures. A better but computationally more intensive criterion for determining variables to be removed from the model is the likelihood ratio test.

There are two steps in the selection of variables for a logistic regression. The first step is the selection of variables from the bivariate analysis of each variable. The second step is the selection of variables for the multiple logistic regression analysis based on the results in the bivariate analysis along with all variables. Finally, the importance of each variable included in the multiple logistic regression model should be verified by different model assessment techniques.

3.3.2.6 Assessing the 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 for all regression models.

In practice, several different measures exist for determining the significance or goodness of fit of a logistic regression model. These measures include the Pearson, Hosmer-Lemeshow, deviance goodness of fit, likelihood ratio test and the classification table. In a theoretical sense, all measures are equivalent. To be more precise, as the number of observation in the set predictor goes to infinity, all measures converge to the same estimate of model significance. However, for any practical regression problem with a finite number of observations in the set of predictors, each measure produces a different estimate. Commonly a regression model designer refers to more than one measure of goodness of fit. If any single measure indicates a low goodness of fit or if the measures differ greatly in their assessments of significance, the designer goes back and makes improvements to the regression model. The test can detect major departures from a logistic response function. The alternatives of interest are:

$$H_0: \; P(Y_i = 1) = E(Y_i) = \frac{\exp(\beta_0 + \beta_1 X_{1_i} + \ldots + \beta_k X_{k_i})}{1 + \exp(\beta_0 + \beta_1 X_{1_i} + \ldots + \beta_k X_{k_i})} \;\ldots$$

$$H_1: \; P(Y_i = 1) = E(Y_i) \neq \frac{\exp(\beta_0 + \beta_1 X_{1_i} + \ldots + \beta_k X_{k_i})}{1 + \exp(\beta_0 + \beta_1 X_{1_i} + \ldots + \beta_k X_{k_i})} \;\ldots\ldots\ldots\ldots\ldots\ldots 7$$

3.3.2.6.1 Likelihood Ratio and Pearson Chi-square Goodness of Fit Test

We can test of the above null hypothesis (equation (7)) using a Pearson chi-square goodness of fit test χ^2_{cal} or likelihood-ratio test (LRT). For a fixed number of settings of explanatory variables, when most fitted counts are about five (use of 5 to 10 is common), χ^2_{cal} and LRT have approximate chi- squared distribution. The degree of freedom (the residual df) for the model, equal the number of parameters in restricted model (i.e., the number of settings of explanatory variables, p) minus the number of model parameters, q. The large χ^2_{cal} or LRT values indicate lack of fit of the model, and 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 model fit.

The usual Pearson chi- square goodness of fit test statistic is given by:

$$\chi^2_{cal} = \sum\left[\frac{(Observed - Fitted)^2}{Fitted}\right]$$

The decision rule for testing the alternatives in equation (7), when controlling the level of significance at α is:

$$\text{If } \chi^2_{cal} \le \chi^2(1-\alpha, p-q)\text{, accept } H_0$$

$$\text{If } \chi^2_{cal} > \chi^2(1-\alpha, p-q)\text{, reject } H_0$$

The likelihood- ratio test statistics is given by

$$\text{LRT} = 2\sum\left[(Observed)Ln\left(\frac{(Observed)}{(Fitted)}\right)\right]$$

Hence, the decision rule for testing the alternative in equation (7) is

$$\text{If LRT} \le \chi^2(1-\alpha, p-q)\text{, accept } H_0$$

$$\text{If LRT} > \chi^2(1-\alpha, p-q)\text{, reject } H_0$$

Goodness-of-fit statistics, LRT and χ^2_{cal}, are summary indicators of the over all quality of fit. Additional diagnostic analysis is necessary to describe the nature of any lack of fit.

3.3.2.6.2 Deviance Goodness of Fit Test:

The deviance of a fitted model compares the log-likelihood of the fitted model to the log-likelihood of a model with n parameters that fits the n observations perfectly. Such a perfectly fitting model is called a saturated model. Consider again the log-likelihood function in equation (5) for n- independent Bernoulli (0, 1) variables. If we let P_i be unconstrained so that each Y_i observation takes on the value 1 with a different probability P_i without restricting P_i in any way, then we will have n parameters for the n observations and can obtain a perfect fit (i.e the residuals will all be zero).

The log-likelihood value for the saturated model will now be compared with the log-likelihood value for the fitted model. The log-likelihood value for the fitted model can never be larger than the log-likelihood value for the saturated model because the fitted model has fewer parameters. The deviance is based on the difference between the two log-likelihood values.

The smaller the difference in the two log-likelihood values, the smaller the deviance and the closer is the fitted model to the saturated model. Hence, the model deviance can be used as a goodness of fit criterion. The larger the model deviance, the poorer is the fit. For large sample size the deviance is approximately normally distributed. The strong difference between the deviance and Pearson chi-square may be a clue that the assumptions for asymptotic distribution of Pearson's chi-square and the deviance may be violated.

3.3.2.6.3 Hosemer-Lemeshow Goodness-of-Fit Test

The test divides subjects into deciles based on predicted probabilities and then computes a chi-square from observed and expected frequencies. Then a probability (*p*) value is computed from the chi-square distribution with 8 degrees of freedom to test the fit of the logistic model. The Hosmer-Lemeshow (H-L) test statistics is given by:

$$\hat{C} = \sum_{j=1}^{10} \frac{(O_k - E_k)^2}{V_k}$$

Where O_k is the observed number of events in the kth group and E_k is expected number of events in the kth group, and V_k is a variance correction factor for the kth group. If this test statistic p-value is greater than 0.05, we will accept the null hypothesis that there is no difference between observed and model-predicted values. As the sample size gets large, the H-L statistic can find smaller and smaller differences between observed and model-predicted values to be significant.

3.3.2.6.4 Classification Table

A classification table is used to assess how well the model fits by comparing predicted values to the observed outcomes. It consists of cases the over all percent of correctly classified and misclassified cases.

3.3.3 Probit Regression Model

The probit regression model for a binary response data describes the relationship between a dichotomous response variable Y, coded to take the values 1 or 0 for `success' and `failure', respectively, and k explanatory variables $x_1, x_2, \ldots x_k$. The explanatory variables can be quantitative or indicator variables referring to the levels of categorical variables. It analyzes the relationship between the probability of the response being success and the explanatory variables, rather than analyzing the relationship between the value of the response variable and the explanatory variables. This assumes that error terms are independent and normally distributed. The probit model for a dichotomous dependent variable is defined as follows:

$$P(y = 1) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k) = \int_{-\infty}^{\beta_0 + \beta_1 X_{1+} + \beta_k X_k} f(z)dz$$

Where,

- P(Y=1) is the probability of Y=1,
- f(z) represents the density function of standard normal random variable z,
- Φ the cumulative distribution function of the standard normal distribution

and $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$ is the probit score or index .

3.3.3.1 Parameter Estimation for Probit Regression Model

The parameter $\beta_0, \beta_1, ..., \beta_k$ can be estimated by the maximum likelihood method using the log-likelihood function. The likelihood of the i^{th} observed unit is:

$$L_i(x_i, \beta) = \Phi(x_i'\beta)^{y_i}\left(1 - \Phi(x_i'\beta)\right)^{1-y_i}$$

Where

- $\Phi(x_i'\beta)$ is the probability of choosing $y_i = 1$ and $1 - \Phi(x_i'\beta)$ is probability of choosing $y_i = 0$.

Taking the product over all units in the sample $i = 1, \ldots, n$ gives the likelihood function:

$$L(y|x, \beta) = \prod_i \Phi(x_i'\beta)^{y_i}[1 - \Phi(x_i'\beta)^{(1-y_i)}$$

$$= \prod_i \Phi_i(x_i'\beta)^{y_i}[1 - \Phi_i(x_i'\beta)]^{1-y_i}$$

The log likelihood function for probit is given by:

$$\ln L = \sum_i y_i \ln \Phi_i(x_i'\beta) + \sum_i (1 - y_i)\ln[1 - \Phi(x_i'\beta)]$$

The maximum likelihood estimates of β in probit regression model are those values of β that maximize the log-likelihood function. The interpretation of a probit coefficient β is that a one unit increase in the predictor leads to increasing the probit score by β standard deviations.

3.3.3.2 Parameter Inference for Probit Regression Model

The probit regression uses z-statistic to test whether a parameter is significantly different from zero. Where Z-value is equals to the estimated parameter divided by its standard error. And to test the null hypothesis all coefficients except that of the intercept are equal to zero, it uses the Likelihood ratio test. The Test statistic is:

$$\mathrm{LR}=\sum_{i=1}^{n} y_i \ln[\frac{\Phi(x_i^{'}\beta)}{\{1-\Phi(x_i^{'}\beta)\}}]+\sum_{i=1}^{n}\ln[1-\Phi(x_i^{'}\beta)]$$

LR follows a χ^2 distribution with K degrees of freedom, where K is the number of parametrs.

3.3.3.3 Selection of predictor variables for Probit Regression Model

For probit regression, the selection of the best variable is the same as logistic regression. There are two steps in the selection of variables for a probit regression. The first step is selection of variables from the bivariate analysis of each variable and the second step is selection of variables for the multivariate analysis based on the results in the bivariate analysis along with all variables. Finally, the importance of each variable included in the multivariate model should be verified by different model assessment techniques.

3.3.3.4 Goodness of fit for Probit Regression Model

Assessing goodness of fit involves investigating how close values predicted by the model are to the observed values. The Goodness of fit of probit regression model may be judged by Pseudo R^2, the classification table, and likelihood ratio test.

3.3.3.4.1 The Pseduo R-Square for Probit Regression

The Pseudo R^2 is defined as:

$$PseudoR^2 = 1 - \frac{\ln \hat{L}(M_{Full})}{\ln \hat{L}(M_{Intercept})} \qquad 0 \leq PseudoR^2 \leq 1$$

Where $\ln \hat{L}(M_{Full})$ is the likelihood of the full model, and

$\ln \hat{L}(M_{Intercept})$ is Likelihood with all coefficients except that of the intercept restricted to zero.

An increasing *P*seudo R^2 may indicate a better fit of the model, whereas here it will not be as such simple as the interpretation of R^2 applied in linear regression. A high value of *P*seudo R^2 does not necessarily indicate a good fit. Since *P*seudo R^2 increases with additional variables an adjustment should be taken to *P*seudo R^2. Therefore, an adjusted measure may be appropriate:

$$PseudoR^2_{adjusted} = 1 - \frac{\ln \hat{L}(M_{Full}) - K}{\ln \hat{L}(M_{Intercept})}$$

where K is the number of explanatory variables.

3.3.3.4.2 Likelihood Ratio Goodness of Fit Test for Probit regression

To test the null hypothesis that the model is a good fit against the alternative the model is not a good fit, the probit regression uses a likelihood ratio test(LRT) and LRT have approximate chi- squared distribution. The degree of freedom (the residual df) for the model, equal the number of parameters in restricted model (i.e the number of settings of explanatory variables, p) minus the number of model parameters, q. The likelihood- ratio test statistics is given by:

$$\text{LRT}= 2\sum\left[(Observed)Ln\left(\frac{(Observed)}{(Fitted)}\right)\right]$$

Hence, the decision rule for testing the above hypothesis is

$$\text{Accept } H_0 \quad \text{if LRT} \leq \chi^2(1-\alpha, p-q),$$

$$\text{Reject } H_0 \quad \text{if LRT} > \chi^2(1-\alpha, p-q),$$

There are also different measures of goodness of fit for probit regression some of them are R^2 of McKelvey and Zavoinas, Akaike Information Criterion (AIC), etc.

3.3.4 Relationship of Probit and Logistic regression

In binary response data there is no compelling reason to choose between logistic and probit regression unless different model diagnosis are made. In practice many researchers choose the logit model because of its comparative mathematical simplicity. Though the models are similar, one has to be careful in interpreting the coefficients estimated by the two models. We can not directly compare the logit and probit regression coefficients. The reason is that, although the standard logistic (the basis of logit) and the standard normal distributions (the basis of probit) both have a mean value of zero, their variances are different; 1 for the standard normal (as we already know) and $\pi^2/3$ for the logistic distribution, where $\pi \approx 22/7$. Therefore, if you multiply the probit coefficient by about 1.81 (which is approximately $= \pi/\sqrt{3}$), we will get approximately the logit coefficient. Alternatively, if we multiply a logit coefficient by 0.55 (= 1/1.81), one will get the probit coefficient. Maddala G.S(1992), however, suggests multiplying a logit estimate by 0.625 to get a better estimate of the corresponding probit estimate. Conversely, multiplying a probit coefficient by 1.6 (= 1/0.625) gives the corresponding logit coefficient.

3.3.5 Model Adequacy

3.3.5.1 Influential Diagnostics

An observation is said to be influential if the observations have a large impact on the predicted values. Influential values have impact on the analysis and conclusion to be drawn from the analysis. Leverage values are influential cases that are used for detecting observations that have a large impact on the predicted values. Unlike linear regression, the leverage values in logistic and probit regression depend on the dependent variable scores and the design matrix bounded by 0 and 1. Their average value is k/n where k is the number of estimated parameters, including the constant, and n is the sample size. Probit and logistic regressions have similar measures of influence.

We take up two measures of influence that are widely used in practice, each based on the omission of a single case to measure its influence. These are: Cook's distance measure and DEFBETAS. Detail discussion are found in Christensen (1997).

3.3.5.2 Detection of Outlier(s)

Observation that is surprisingly distant from the remaining observations in the sample is called outlying values or outliers. If there are outliers, the analysis will give a wrong result. There are different statistical methods for detecting the presence of outlier(s). Some of them are : Unstandardized residuals, $Y_i - \hat{P}_i$ and the Pearson or standardized residuals.

CHAPTER FOUR

STATISTICAL DATA ANALYSIS

4.1 Introduction

The purpose of this chapter is to quantify the effect of different socio-economic and demographic correlates of harmful traditional practices of women in Ethiopia. The study used the 2005 Demographic and Health Survey (DHS) data to provide the correlates of harmful traditional practices. In the data analysis first we begin with summary statistics then the bivariate analysis followed by the multiple logistic regression analysis, and finally different model diagnostic techniques is used for both logistic and probit regression models. The data are analyzed using the Statistical Package for Social Sciences (SPSS) version 13 and STATA 9.

4.2 Summary Statistics

In the survey women of age 15-49 were asked for whether they faced with harmful traditional practices or not. From the total of 14, 070 women 10, 160 were faced harmful traditional practices. According to the survey result at the country level 72.2 percent of women of age 15-49 were faced harmful traditional practices as shown in Table 4.1.

Table 4.1: Frequency and Percentage Distribution of Women Who Faced Harmful Traditional practices

Harmful traditional practices faced with	*Frequency*	*Percent*
No	*3, 910*	*27.8*
Yes	*10, 160*	*72.2*
Total	*14, 070*	*100.0*

4.2.1 Harmful Traditional Practices by Some Scio-Economic and Demographic Variables

In the data analysis, first we summarize information of the variables considered in the study. Among the total 10,160 women who faced harmful traditional practices, the majority (63.0 %) were illiterate. About 58.2% of the total women who faced harmful traditional practices were Christian followed by Muslim (40.4%). Of the total women who experience harmful traditional practices, 86.1% were those who do not have knowledge on harmful traditional practices. Moreover, the result showed that the frequency of reading newspapers or magazines, listening radio and watching TV brought differences to women to experience harmful traditional practices. That is, women who do not exposed to medias at all, and almost every day to each of those printed and electronic medias indicated above, taken the greatest and the least percentage, respectively. Of the total women who faced harmful traditional practices 69 percent were in rural areas. The percentage distributions of women who faced harmful traditional practices by some Scio-economic and demographic variables is given in Table 4.2 below.

Table 4.2 Percentage Distributions of Women Who Faced Harmful Traditional Practices by Some Scio-Economic and Demographic Variables

No.	Description and Name	Categories	Number	Percentage
1	Region	(0) Tigray	396	3.9
		(1) Afar	742	7.3
		(2) Amhara	1207	11.9
		(3) Oromia	2005	19.7
		(4) Somali	627	6.2
		(5) Ben-Gumz	506	5.0
		(6) SNNP	1730	17.0
		(7) Gambela	267	2.6
		(8) Harari	773	7.6
		(9) Addis Ababa	1290	12.7
		(10) Dire Dawa	617	6.1
2	Age	Contineous	27.9(mean)	9.445(s.d.)
3	Place of Residence	(0) Urban	3152	31
		(1) Rural	7008	69
4	Educational Level	(0) No education	6400	63
		(1) Primary	2044	20.1
		(2) Secondary	1506	14.8
		(3) Higher	210	2.1
5	Religion	(0)Christian	5914	58.2
		(1) Muslim	4109	40.4
		(3) Others	137	1.4
6	Ethnicity	(0)Affar	579	5.7
		(1)Amara	2906	28.6
		(2)Guragie	711	7.0
		(3)Oromo	3119	30.7
		(4)Somalie	671	6.6
		(5)Tigraiy	457	4.5
		(6)Other Ethiopian Nationalities	1717	16.9
7	Frequency of reading newspaper or magazine	(0)Not at all	8183	80.5
		(1)Less than once a	1641	16.2

		week		
		(2)At least once a week	177	1.7
		(3)Almost Every day	159	1.6
8	Frequency of listening radio	(0)Not at all	5345	52.6
		(1)Less than once a week	2805	27.6
		(2)At least once a week	359	3.5
		(3)Almost Every day	1651	16.3
9	Frequency of watching Tv	(0)Not at all	7337	72.2
		(1)Less than once a week	1321	13
		(2)At least once a week	256	2.5
		(3)Almost Every day	1246	12.3
10	Wealth Index	(0)Poorest	1923	18.9
		(1)Poorer	1452	14.3
		(2)Middle	1492	14.7
		(3)Richer	1486	14.6
		(4)Richest	3807	37.5
11	Knowledge on harmful traditional practices	(1) Don't Have knowledge	8747	86.1
		(2) Have knowledge	1413	13.9
12	Attitude on the continuation of harmful traditional practices	(1) Discontinued	3484	34.3
		(2) Continued	6118	60.2
		(3) Indifferent	558	5.5

4.3 Logistic Regression

4.3.1 Variable Selection for Logistic Regression

To determine the factors that are significantly correlated with harmful traditional practices of women in Ethiopia, first a univariate analysis was used and then a multiple logistic regression analysis were done using the significant variables in the univariate analysis. For both the univariate and the multiple logistic regression analysis, we used Wald test to identify the significant variable which are correlated with harmful traditional practices. Using scatter plot, linearity assumption for age is checked and the scatter plot in appendix shows that age has a linear relation to the dependent.

4.3.1. 1 Bivariate Findings

From the simple logistic regression of each predictor, we identify the significant variables using Wald test. The results in Table 4.3 shows that all the 12 explanatory variables Age, Region, Place of residence, Educational level, Religion, Ethnicity, Frequency of reading newspaper or magazine, Frequency of listening radio, Frequency of watching Television , Wealth Index, Knowledge and Attitude on harmful traditional practices have significant association with harmful traditional practices.

Table 4.3 Bivariate Logistic Regression Result

Covariates	β	S.E.	Wald	df	sig.	Exp(β)
Region			1994.549	10	0.000	
Tigray	-1.479	0.078	357.733	1	0.000	0.228
Afar	1.832	0.144	161.983	1	0.000	6.246
Amhara	0.121	0.069	3.028	1	0.082	1.128
Oromia	1.355	0.082	269.967	1	0.000	3.876
Somali	2.963	0.251	139.922	1	0.000	19.362
Ben-Gumz	-0.017	0.088	0.037	1	0.847	0.983
SNNP	0.361	0.07	26.727	1	0.000	1.435
Gambela	-1.352	0.092	214.487	1	0.000	0.259
Harari	1.257	0.115	119.795	1	0.000	3.515
Dire Dawa	1.751	0.138	160.331	1	0.000	5.763
Addis Ababa(Ref.)						1.000
Age	0.033	0.002	237.327	1	0.000	1.033
Residence						
Rural	-0.069	0.04	2.899	1	0.089	0.934
Urban(Ref.)						1.000
Education			147.245	3	0.000	
No education	0.798	0.11	52.517	1	0.000	2.222

Primary	0.456	0.114	15.922	1	0.000	1.578
Secondary	0.316	0.116	7.421	1	0.006	1.371
Higher(Ref.)						1.000
Religion			1012.925	2	0.000	
Muslim	0.688	0.119	33.218	1	0.000	1.990
Others	2.412	0.128	353.567	1	0.000	11.154
Christian(Ref.)						1.000
Ethinicity			1957.508	6	0.000	
Affar	0.648	0.051	158.601	1	0.000	1.913
Guragie	3.04	0.216	197.862	1	0.000	20.911
Oromo	2.077	0.128	262.482	1	0.000	7.978
Somalie	2.012	0.069	843.652	1	0.000	7.480
Tigraiy	3.433	0.242	201.221	1	0.000	30.957
Other Ethiopian Nationalities	-0.919	0.069	176.989	1	0.000	0.399
Amhara(Ref.)						1.000
Frq. of lisning			7.516	3	0.057	
Not at all	0.106	0.052	4.15	1	0.042	1.112
Less than once a week	0.158	0.058	7.424	1	0.006	1.171
At least once a week	0.114	0.111	1.055	1	0.304	1.120
Almost Every day(Ref.						1.000
Frq. of reading			103.01	3	0.000	
Not at all	0.56	0.131	18.227	1	0.000	1.751
Less than once a week	0.143	0.136	1.111	1	0.292	1.154
At least once a week	0.007	0.18	0.001	1	0.970	1.007
Almost Every day(Ref.)						1.000
Frq. Of watching TV			22.458	3	0.000	
Not at all	0.235	0.055	18.388	1	0.000	1.264
Less than once a week	0.308	0.074	17.461	1	0.000	1.360
At least once a week	0.143	0.129	1.228	1	0.268	1.154
Almost Every day(Ref.)						1.000
Wealth index			25.549	4	0.000	
Porest	-0.116	0.052	4.98	1	0.026	0.890
Poorer	-0.138	0.057	5.894	1	0.015	0.871
Middle	0.015	0.059	0.066	1	0.797	1.015
Richer	0.172	0.061	7.903	1	0.005	1.188
Richest(Ref.)						1.000
Knowledge						
Don't have knowledge	-0.777	0.046	280.077	1	0.000	0.460

Have knowledge(Ref.)						1.000
Attitude			510.555	2	0.000	
Continued	1.374	0.113	146.928	1	0.000	3.953
Discontinued	-0.174	0.097	3.221	1	0.073	0.840
Indifferent(Ref)						1.000

4.3.1.2 Analysis of Multiple Logistic Regression

Further multiple logistic regression analysis is applied on significant variables in the bivriate analysis. Forward stepwise likelihood ratio method is used to select variables in the multiple logistic regression analysis. The stepwise selection is proceeded by entering variable Ethnicity on step 1; Attitude on step 2; Age onstep3; Region on step4; Educational level on step 5; Religion on step 6; Wealth index on step 7; and Place of Residence on Step 8.

The variables that are found to be significant in the multiple logistic regression analysis are Ethnicity, Religion, Attitude on harmful traditional practices, Age, Region, Educational level, Wealth index, and Place of residence. The estimated coefficient (β) in the final model with their corresponding standard errors, Wald statistics, d.f, probability of significance and exp(B) is given in Table 4.4.

Table 4.4 Variables in the Final Multiple Logistic Regression Model

Covariates	B	S.E.	Wald	df	Sig.	Exp(B)
Age	.043	.003	193.703	1	.000	1.044
Education			46.372	3	.000	
No Education	.762	.139	30.007	1	.000	2.143
Primary	.651	.137	22.510	1	.000	1.917
Secondary	.295	.131	5.094	1	.024	1.343
Higher education(Ref.)						1.000
Wealth index			14.454	4	.006	
Poorest	-.338	.124	7.464	1	.006	.713
Poorer	-.148	.119	1.546	1	.214	.862
Middle	.016	.116	.020	1	.888	1.016
Richer	.006	.114	.003	1	.958	1.006

Richest(Ref.)						1.000
Region			202.850	10	.000	
Tigray	-.656	.167	15.508	1	.000	.519
Affar	.760	.238	10.226	1	.001	2.139
Amhara	.041	.109	.141	1	.707	1.042
Oromia	.545	.116	22.149	1	.000	1.725
Somalie	1.808	.525	11.847	1	.001	6.099
Ben-Gumz	.520	.143	13.155	1	.000	1.683
SNNP	.684	.132	27.043	1	.000	1.982
Gambela	-.060	.154	.152	1	.697	.942
Harari	.880	.127	47.629	1	.000	2.410
Dire Dawa	1.388	.149	87.158	1	.000	4.006
Addis Ababa(Ref.)						1.000
Religion			37.134	2	.000	
Muslim	.479	.081	34.753	1	.000	1.614
Others	-.267	.253	1.114	1	.291	.766
Christian(Ref.)						1.000
Ethnicity			280.508	6	.000	
Affar	.641	.336	3.639	1	.056	1.898
Guragie	1.364	.145	88.840	1	.000	3.911
Oromo	.855	.094	82.584	1	.000	2.351
Somalie	1.010	.494	4.174	1	.041	2.745
Tigray	-.982	.141	48.278	1	.000	.375
Others	-.412	.105	15.541	1	.000	.662
Amhara(Ref.)						1.000
Attitude			235.974	2	.000	
Discontinued	-1.255	.082	234.710	1	.000	.285
Indifferent	-1.062	.131	65.631	1	.000	.346
Continued(Ref.)						
Rural	.223	.108	4.271	1	.039	1.250
Urban(Ref.)						1.000
Constant	.019	.174	.012	1	.914	1.019

4.3.2 Model Checking and Diagnostics for Logistic Regression

From different techniques of goodness of fit test discussed in chapter 3, we used likelihood ratio test, Hosmer-Lemeshow goodness of fit test, and the classification table.

4.3.2.1 The Likelihood Ratio Test

A good model is one that results a high likelihood of the observed results. The likelihood ratio test is the difference between -2loglikelihood (-2LL) for the final model and -2LL for the model which include the intercept only. The likelihood ratio chi-squared value is given in Table 4.5.

Table 4.5 Model Fitting Information for Logistic Regression

Model	-2LL	Chi-square value	d.f	Sign.
Intercept only	12543.724			
Final model	9550.805	2992.919	12	0.000

The results in Table 4.5 used to test the over all significance of the fitted model. Hence, the result at above table shows that the null hypothesis that the coefficients for all the terms in the fitted model, except the constant term, are zero is rejected. This implies that the fitted model is a good model.

4.3.2.2 Classification Table

Another method of assessing how best our model fit is to compare our predictions to the observed outcomes. A good model is the model which minimizes miss-classifications. As shown in Table 4.6 the fitted model has an over all predictive accuracy of 83.1% with 96.1% of the yes group, and 33.6 % of the no group being correctly classified and this indicates that the model classification power is high enough to consider the model reliability.

Table 4.6 Classification Table

Observed	Predicted		
	Women faced harmful traditional practices		Percentage correct
	No	Yes	
No	858	1692	33.6
Yes	380	9345	96.1
Overall percentage			83.1

4.3.2.3 The Hosmer-Lemeshow Test

Table 4.7 and table 4.8 are the values of Hosmer-Lemeshow statistic and its contingency table respectively.

Table 4.7 Hosmer and Lemeshow Test

	Chi-square	df	Sig.
1	14.576	8	.068

Table 4.8 Contingency Table for Hosmer and Lemeshow Test

	Harmful trad.practices = No		Harmful trad.practices = Yes		Total
	Observed	Expected	Observed	Expected	
1	631	643.266	631	618.734	1262
2	487	487.483	774	773.517	1261
3	417	398.907	845	863.093	1262
4	315	321.819	948	941.181	1263
5	243	250.275	1023	1015.725	1266
6	181	182.533	1082	1080.467	1263
7	161	129.292	1101	1132.708	1262
8	69	85.023	1193	1176.977	1262
9	52	53.913	1210	1208.087	1262
10	24	27.489	1231	1227.511	1255

The result of Hosmer-Lemeshow test shows that the model is good; that is, we do not reject the null hypothesis that the fitted model is good at 5 % level of significance.

4.3.3 Influential Diagnostic for Logistic Regression

After building a statistical model and applying the goodness of fit test, it is important to examine further model adequacy checking and further model diagnosis. The adequacy of the fitted model was checked for possible presence and treatment of outliers, influential values. Cook's distance and DFBETAS are used to measure influence and normalized residuals are used for outliers. The diagnostic test results for detection of outliers and influential values are presented in Appendix.

The minimum and maximum values of the diagnostic results are presented in Appendix. The table shows that the normalized residuals are within the interval of -3 and 3 implying that no outliers were detected at 0.05 level of significance. The absolute value of DFBETAs for model parameters including the constant term and Cook's influence statistics were both less than unity. This implies no impact of an observation on the coefficient of a particular predictor variable.

4.4 Probit Regression

4.4.1 Variable selection for Probit Regression

To determine factors which are significantly correlated with harmful traditional practices of women in Ethiopia, we follow the same procedure for probit regression as the logistic regression. For both the univariate and the multiple logistic regression analysis, we used Z-test to identify the significant variables that are correlated with harmful traditional practices.

4.4.2 Bivariate Findings for Probit Regression

From the bivariate probit regression of each predictor, we identify the significant variables using z-test. The bivariate probit regression results in Table 4.9 show that Age, Region, Educational level, Place of residence, Religion, Ethnicity, Frequency of reading newspaper or magazine, Frequency of listening radio, Frequency of watching Television, wealth Index, Knowledge and Attitude on harmful traditional practices have significant association with harmful traditional practices at 0.25 level of significance.

Table 4.9 Bivariate probit regression result

Covariates	B	S.E.	Z	sig.
Region			24.94	0.000
Tigray	-0.65482	414207	-15.81	0.000
Afar	1.797337	0.076945	23.36	0.000
Amhara	0.715715	0.045733	15.65	0.000
Oromia	1.263876	0.04877	25.92	0.000
Somali	1.849896	0.108889	16.99	0.000
Ben-Gumz	0.201822	0.057619	3.5	0.000
SNNP	0.274645	0.047752	5.75	0.000
Gambela	-1.5254	748907	-20.37	0.000
Harari	-0.10898	0.083296	-1.31	0.191
Dire Dawa	-1.11523	0.077465	-14.4	0.000
Addis Ababa(Ref.)				
Age	0.018702	0.001266	14.77	0.016
Residence				
Rural	0.502243	0.043516	11.54	0.000
Urban(Ref.)				
Education			-6.73	0.000
No education	0.693696	0.01539	45.07	0.000
Primary	-0.04464	0.035289	-1.26	0.206
Secondary	.0190176 .0538185 0.35 0.724	0.116	7.421	0.006
Higher(Ref.)				
Religion			3.62	0.000
Muslim	-0.01752	0.015783	-1.11	0.267
Others	1.378142	0.083491	16.51	0.000
Christian(Ref.)				
Ethnicity			14.23	0.000
Affar	-3.31168	0.12825	-25.82	0.000
Guragie	3.04	0.216	197.862	0.000
Oromo	-0.82985	0.120207	-6.9	0.000
Somalie	-1.4091	0.120775	-11.67	0.000
Tigraiy	-0.89659	0.158294	-5.66	0.000
Other Ethiopian Nationalities	-2.27275	0.1098	-20.7	0.000
Amhara(Ref.)				

Frq. of listening			-2.19	0.029
Not at all	0.056737	0.026571	2.14	0.033
Less than once a week	0.015133	0.065232	0.23	0.817
At least once a week	0.589647	0.016049	36.74	0.000
Almost Every day(Ref.				
Frq. of reading			-4.05	0.000
Not at all	-0.15047	0.039383	-3.82	0.000
Less than once a week	-0.12034	0.097484	-1.23	0.217
At least once a week	0.649295	0.013316	48.76	0.000
Almost Every day(Ref.)				
Frq. Of watching TV			-4.18	0.000
Not at all	0.076799	0.035802	2.15	0.032
Less than once a week	0.032401	0.077837	0.42	0.677
At least once a week	0.602246	0.013768	43.74	0.000
Almost Every day(Ref.)				
Wealth index			2.13	0.033
Porest	-0.02603	0.035957	-0.72	0.469
Poorer	0.042725	0.034541	1.24	0.216
Middle	0.122804	0.035601	3.45	0.001
Richer	0.509106	0.032928	15.46	0.000
Richest(Ref.)				
Knowledge			-16.01	0.000
Don't have knowledge	0.675211	0.013046	51.76	0.000
Have knowledge(Ref.)				
Attitude			-6.62	0.000
Continued	-0.65823	0.043889	-15	0.000
Discontinued	0.214682	0.160819	1.33	0.182
Indifferent(Ref)				

4.4.3 Analysis of Multiple Probit Regression

Further, multiple probit regression analysis is done using the significant variables in the bivariate analysis. The result of the multiple probit regression for the fitted model is given in Table 4.10.

Table 4.10 Final Model results of Multiple probit Regression

Coovariates	Coef.	Std. Err.	z	P>z
Age	0.0244243	0.0017781	13.74	0.000
Education			-5.39	0.000
No Education	0.0785408	0.0443292	1.77	0.076
Primary	0.0256998	0.0600067	0.43	0.668
Secondary	0.04500	-5.24	0.000	
Higher education(Ref.)				
Wealth Index			3.28	0.001
Poorest	0.0643187	0.0545302	1.18	0.238
Poorer	0.080302	0.0533542	1.51	0.132
Middle	0.0320264	0.0575046	0.56	0.578
Richer	-0.26437	0.05561	-4.75	0.000
Richest(Ref.)				
Region			9.39	0.000
Tigray	0.6451602	0.1507983	4.28	0.000
Affar	0.2479023	0.0897788	2.76	0.006
Amhara	0.4427886	0.0901067	4.91	0.000
Oromia	0.9474588	0.2448149	3.87	0.000
Somalie	0.3121778	0.0957927	3.26	0.001
Ben-Gumze	0.2838389	0.0879387	3.23	0.001
SNNP	-0.5971734	0.1043576	-5.72	0.000
Gambela	-0.0764603	0.0913429	-0.84	0.403
Harari	-0.6757201	0.0769773	-8.78	0.000
Dire Dawa	-0.27826	0.13158	-2.11	0.034
Addis Ababa (Ref.)				
Religion			-0.58	0.565
Muslim	0.2810171	0.0855354	3.29	0.001
Others	0.07930	0.15153	0.52	0.601
Christian(Ref.)				
Ethnicity			-6.7	0.000
Affar	0.5272654	0.1697637	3.11	0.002
Guragie	0.9917027	0.0834929	11.88	0.000
Oromo	0.8335876	0.0706325	11.8	0.000
Somalie	0.9481671	0.2355675	4.03	0.000
Tigray	0.4688858	0.1146206	4.09	0.000
Others	-0.84004	0.13359	-6.29	0.000
Amhara(Ref.)				
Attitude			-11.24	0.000
Discontinued	-0.1772606	0.0696227	-2.55	0.011
Indifferent	0.325034	0.0218384	14.88	0.000
Continued(Ref.)				
Rural	-0.1607122	0.063603	-2.53	0.012

Urban				
_cons	-0.9926184	0.160887	-6.17	0.000

4.4.4 Goodness of Fit for Probit Regression

4.4.4.1 The Likelihood Ratio Test

As shown in Table 4.11 the likelihood ratio chi-square value for probit regression is 2830.68, this indicates that there is a significant difference in -2LL between the final model and the model with only the constant at 0.05 level of significance. That is, the variables in the final model explain the data quite well. Pseudo R^2 value is also given in table 4.11.

Table 4.11: Model Fitting Information for Probit Regression

Likelihood Ratio Chi-Square	*degree of Freedom*	*Significance*	*pseudo R^2*
2830.68	***12***	***0.000***	***0.2391***

4.4.4.2 Classification Table for Probit Regression

As shown in Table 4.12 the fitted model has an over all predictive accuracy of 67% with 69.1% of the yes group, and 23.7 % of the no group being correctly classified and this indicates that the model classification power is reasonably enough to consider the model reliability.

Table 4.12: Classification Table

Observed	Predicted		
	Women faced harmful traditional practices		Percentage
	No	Yes	correct
No	8759	1610	23.7
Yes	342	809	69.1
Overall percentage			67

4.5 Probit and Logistic Regression

As shown in table 4.6 and 4.12 the logistic and probit model have an overall predictive accuracy of 83.1 % and 67 % respectively. From this result the study preferred the logistic regression model to identify the correlates of harmful traditional practices on Ethiopian women.

4.6 Interpretation of Final Model Results

The final model results are those obtained by the multiple logistic model presented in Table 4.4 As can be seen in table 4.4, uneducated women, primary and secondary educated women were 2.143 ,1.917,1.343 times more likely to face harmful traditional practices than higher educated women. Women who reside in Somalie, Dire Dewa, Affar were 6.099, 4.006, 2.139 times more likely to face harmful traditional practices than women in Addis Ababa. Muslim women were 1.614 times more likely to face harmful traditional practices than Christian in Ethiopia. Similarly, women who reside in rural area were 1.250 times more likely to face harmful traditional practices than women in urban area. And women who had willingness to discontinue the practices were 0.285 times more likely to face harmful traditional practices than women who hadn't willingness to continue the practice.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

This study is an attempt to examine the correlates of harmful traditional practices on women in Ethiopia. Several factors have been considered in the study. Multiple logistic and probit regression analysis are employed to identify the most important correlates to harmful traditional practices on Ethiopian women and estimating the variables used in the model. However; some variables such as frequency of reading, frequency of watching TV, frequency of listening radio and knowledge on harmful traditional practices are dropped from the inclusion in the model because of their insignificant relation to harmful traditional practices.

The result of analysis indicates that age, region, educational level, religion, ethnicity, place of residence, wealth index, and attitude on harmful traditional practices have significant effect on harmful traditional practices.

From the regression analysis, the result showed that uneducated women were 2.143 times more likely to face harmful traditional practices than higher educated women. Furthermore, the analysis indicated that Muslim women were 1.614 times more likely to face harmful traditional practices than Christian. More over the result showed that Women who reside to Somalie were 6.099 times more likely to face harmful traditional practices than women in Addis Ababa. Similarly, the result of the analysis showed that women who have awareness on harmful traditional practices were less likely to face the problem. Like wise, women who reside in rural areas were 1.250 times more likely to face harmful traditional practices than women in urban area.

Model adequacy diagnostic tests showed that there were no outliers and influential values. In order to decrease the practice, strategies should be designed in a way that it would focus on and address correlates as well as other factors that are useful to minimize the problem.

5.2 Recommendation

Harmful traditional practices have a serious impact on women health. To decrease these practices, it is recommended that:

1. Since women's education is important in reducing the practice of harmful traditional practices, special strategy has to be set to improve the educational status of women and gender issues.
2. To eradicate this practice, strategies and policies has to be set to increase the continuous awareness of women on harmful traditional practices.
3. To reduce the problem some better attentions by the Governmental and Non- Governmental Organizations has to be given to women who reside in Somalie and rural areas.
4. To alleviate the problem on harmful traditional practices, all the concerned bodies have to put all their efforts to bring changes on religious outlooks within this regard.
5. This study used the data from the Ethiopian Central Statistical Agency DHS survey conducted in 2005 on which some harmful traditional practices and important exogenous variables are not included in the survey. Hence, we further recommended similar studies with additional harmful traditional practices and

exogenous variables to provide complete and useful insight for formulating appropriate policies.

REFRENCES

Adriana_Kaplan(2009). Perceptions, Degree of Knowledge, Attitudes and Practices of the Primary Healthcare Professionals in Relation to Female Circumcision. Health Services Research 2009, 9: 11doi: 10.1186/1472-6963-9-11.

A.Gage and R.van Rossen(1999). Socio Economic Correlates of and Gender Differences in Attidunal Support for the Discontinuation of Female Circumcision. International Journal of Gynecology & Obstetrics, Volume 92, Issue 1, Pages 92-96.

Almoroth(2001). Male Complications and Attitudes with Regard to Female Circumcision. Social Science & Medicine 53 (2001) 1455–1460.

Dire, MA & Lindmark,G(1991). Female Circumcision in Somalia and Women's Motives. Acta Obstetricia et Gynecologica Scandinavica. 70(7-8):581-5, 1991.

El-Dareer AA(1982). Prevalence and Epidemiology of Female Circumcision in Sudan. World Health Organization, Regional Office for the Eastern Mediterranean, 1982 :312-34, WHO EMRO Technical Publication No. 2, Vol.2.

Esthero, Asekun-Olarinmoya and Oluwatoyin a.Amulan (2008). The Level of Practice of Female Circumcision and the Impact of a Health Education Intervention in Shao Community (Nigeria). The European Journal of Contraception & Reproductive Health Care, Volume 13, Issue 3 2008 , pages 289 – 297.

Getu Degu & Mlikias(2001). Knowledge, Attitudes and Practices on Traditional Harmful Health in Dembia District Northwest Ethiopia. Ethiop.J .Health.Dev.2002;16:199-207.

Getahun H. (1997). *Marriage Through Abduction (Telefa) in Rural North West Ethiopia. Ethiop Med J. 2001 Apr;39(2):105-12.*

Hosmer D. and Lemeshow S. (1989). *Applied Logistic Regression. John Wiley and Sons. Inc., New York.*

Islam and Uddin(2001). Women's Attitudes Towards Female Circumcision in Sudan. International Family Planning Perspectives, 2001, 27(2):71-76.

Maddala G.S (1992). Introduction to Econometrics. New York. Macmillan, Inc.

Mbacke (1998). The prevalence and Major Correlates of Female Genital Mutilation in Ghana's Kassena-Nankana District. African Journal of Reproductive Health, 1998 Oct;2(2):13-24.

Okemgbo (2002). Prevalence, Patterns and Correlates of Domestic Violence in Selected Igbo Communities of Imo State, Nigeria. African Journal of Jeproductive Health 2002;6(2):101-14.

Omaima El-etal(2002). The prevalence and social correlates of circumcision among girls aged 10–19, Population Council, Cairo, Egypt.

Snow et al (2002). The Prevalence and Social Determinants of Female Circumcision . Women's Health Action Research Centre, Benin City, Nigeria.

Suzuki C. and Meekers D. (2005). The Effect of Exposure to Communication Messages on Support for Female Circumcision in Egypt. Global Public Health, Volume 3, Issue 4 October 2008 , pages 383 – 39.

William H. Greene (2003). Econometrics Analysis. New Jersey.

APPENDIX
LOGISTIC REGRESSION OUT PUT

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	Ethnicity			1398.888	6	.000			
	Ethnicity(1)	2.310	.230	100.587	1	.000	10.078	6.416	15.829
	Ethnicity(2)	1.301	.133	95.712	1	.000	3.672	2.830	4.765
	Ethnicity(3)	1.331	.072	337.060	1	.000	3.784	3.283	4.362
	Ethnicity(4)	3.356	.358	88.081	1	.000	28.670	14.225	57.781
	Ethnicity(5)	-1.534	.070	476.050	1	.000	.216	.188	.248
	Ethnicity(6)	.085	.066	1.673	1	.196	1.089	.957	1.238
	Constant	1.057	.037	833.205	1	.000	2.878		
Step 2(b)	Ethnicity			1313.217	6	.000			
	Ethnicity(1)	1.590	.235	45.906	1	.000	4.902	3.095	7.765
	Ethnicity(2)	1.431	.134	114.418	1	.000	4.183	3.218	5.436
	Ethnicity(3)	1.297	.073	312.470	1	.000	3.658	3.168	4.224
	Ethnicity(4)	2.811	.359	61.197	1	.000	16.635	8.224	33.645
	Ethnicity(5)	-1.617	.074	482.726	1	.000	.198	.172	.229
	Ethnicity(6)	.015	.067	.048	1	.827	1.015	.890	1.158
	Attitude			370.795	2	.000			
	Attitude(1)	-1.439	.075	368.310	1	.000	.237	.205	.275
	Attitude(2)	-1.043	.127	67.856	1	.000	.352	.275	.452
	Constant	2.244	.076	860.789	1	.000	9.428		
Step 3(c)	AGE	.046	.003	254.467	1	.000	1.047	1.041	1.053
	Ethnicity			1363.126	6	.000			
	Ethnicity(1)	1.588	.236	45.444	1	.000	4.894	3.084	7.765
	Ethnicity(2)	1.505	.135	124.519	1	.000	4.503	3.457	5.865
	Ethnicity(3)	1.351	.074	331.493	1	.000	3.863	3.340	4.468
	Ethnicity(4)	2.827	.360	61.677	1	.000	16.899	8.345	34.221
	Ethnicity(5)	-1.710	.076	509.549	1	.000	.181	.156	.210
	Ethnicity(6)	.037	.068	.297	1	.586	1.038	.908	1.186
	Attitude			331.085	2	.000			
	Attitude(1)	-1.382	.076	330.510	1	.000	.251	.216	.291
	Attitude(2)	-1.080	.127	71.839	1	.000	.340	.265	.436
	Constant	.964	.108	79.201	1	.000	2.623		
Step 4(d)	AGE	.045	.003	243.678	1	.000	1.046	1.041	1.052
	Region			258.086	10	.000			
	Region(1)	-.314	.149	4.420	1	.036	.731	.545	.979
	Region(2)	1.134	.233	23.660	1	.000	3.108	1.968	4.907
	Region(3)	.473	.087	29.788	1	.000	1.605	1.354	1.902
	Region(4)	.873	.105	68.797	1	.000	2.393	1.947	2.941

	Region(5)	2.011	.523	14.790	1	.000	7.471	2.681	20.822
	Region(6)	.989	.131	56.877	1	.000	2.689	2.079	3.477
	Region(7)	1.012	.120	71.409	1	.000	2.751	2.175	3.478
	Region(8)	.259	.145	3.191	1	.074	1.296	.975	1.722
	Region(9)	1.048	.124	71.739	1	.000	2.851	2.238	3.634
	Region(10)	1.543	.147	110.949	1	.000	4.681	3.512	6.238
	Ethnicity			389.698	6	.000			
	Ethnicity(1)	.969	.323	9.017	1	.003	2.634	1.400	4.956
	Ethnicity(2)	1.582	.142	124.047	1	.000	4.866	3.683	6.428
	Ethnicity(3)	1.078	.091	139.493	1	.000	2.938	2.457	3.514
	Ethnicity(4)	1.504	.490	9.408	1	.002	4.501	1.721	11.771
	Ethnicity(5)	-1.017	.140	52.640	1	.000	.362	.275	.476
	Ethnicity(6)	-.317	.102	9.672	1	.002	.728	.596	.889
	Attitude			305.014	2	.000			
	Attitude(1)	-1.378	.079	304.801	1	.000	.252	.216	.294
	Attitude(2)	-1.053	.130	65.753	1	.000	.349	.270	.450
	Constant	.466	.125	13.954	1	.000	1.594		
Step 5(e)	AGE	.043	.003	188.063	1	.000	1.043	1.037	1.050
	EDUCATION			71.706	3	.000			
	EDUCATION(1)	.833	.135	38.049	1	.000	2.301	1.766	2.998
	EDUCATION(2)	.714	.136	27.489	1	.000	2.042	1.564	2.667
	EDUCATION(3)	.290	.131	4.885	1	.027	1.336	1.033	1.728
	Region			246.212	10	.000			
	Region(1)	-.622	.155	16.014	1	.000	.537	.396	.728
	Region(2)	.961	.235	16.711	1	.000	2.613	1.649	4.142
	Region(3)	.188	.094	3.984	1	.046	1.207	1.003	1.452
	Region(4)	.669	.108	38.266	1	.000	1.953	1.580	2.415
	Region(5)	1.878	.518	13.130	1	.000	6.541	2.368	18.063
	Region(6)	.757	.135	31.620	1	.000	2.131	1.637	2.774
	Region(7)	.784	.122	41.107	1	.000	2.190	1.723	2.783
	Region(8)	.031	.148	.043	1	.835	1.031	.772	1.377
	Region(9)	1.048	.125	70.479	1	.000	2.851	2.233	3.641
	Region(10)	1.487	.148	101.504	1	.000	4.424	3.313	5.908
	Ethnicity			347.060	6	.000			
	Ethnicity(1)	.789	.325	5.898	1	.015	2.201	1.164	4.160
	Ethnicity(2)	1.515	.143	112.928	1	.000	4.550	3.441	6.016
	Ethnicity(3)	.975	.092	111.745	1	.000	2.652	2.213	3.177
	Ethnicity(4)	1.327	.487	7.444	1	.006	3.772	1.453	9.787
	Ethnicity(5)	-.985	.141	48.525	1	.000	.373	.283	.493
	Ethnicity(6)	-.385	.102	14.173	1	.000	.681	.557	.832
	Attitude			245.051	2	.000			
	Attitude(1)	-1.264	.081	243.918	1	.000	.282	.241	.331
	Attitude(2)	-1.076	.130	68.366	1	.000	.341	.264	.440
	Constant	.026	.174	.022	1	.883	1.026		
Step 6(f)	AGE	.043	.003	191.517	1	.000	1.044	1.038	1.050
	EDUCATION			62.606	3	.000			

	EDUCATION(1)	.791	.135	34.304	1	.000	2.206	1.693	2.875
	EDUCATION(2)	.687	.136	25.526	1	.000	1.988	1.523	2.594
	EDUCATION(3)	.288	.131	4.858	1	.028	1.334	1.032	1.724
	Region			212.224	10	.000			
	Region(1)	-.623	.155	16.064	1	.000	.536	.396	.727
	Region(2)	.785	.237	10.954	1	.001	2.192	1.377	3.488
	Region(3)	.141	.095	2.230	1	.135	1.152	.957	1.386
	Region(4)	.621	.108	32.838	1	.000	1.860	1.504	2.300
	Region(5)	1.772	.522	11.521	1	.001	5.881	2.114	16.360
	Region(6)	.627	.137	21.047	1	.000	1.872	1.432	2.447
	Region(7)	.807	.123	42.876	1	.000	2.241	1.760	2.853
	Region(8)	.013	.148	.008	1	.928	1.014	.758	1.356
	Region(9)	.904	.127	50.484	1	.000	2.469	1.924	3.168
	Region(10)	1.394	.149	87.890	1	.000	4.030	3.011	5.393
	Religion			36.897	2	.000			
	Religion(1)	.469	.081	33.915	1	.000	1.598	1.365	1.871
	Religion(2)	-.328	.251	1.698	1	.193	.721	.440	1.180
	Ethnicity			285.278	6	.000			
	Ethnicity(1)	.535	.328	2.658	1	.103	1.708	.897	3.252
	Ethnicity(2)	1.367	.145	89.227	1	.000	3.923	2.954	5.210
	Ethnicity(3)	.878	.093	88.300	1	.000	2.405	2.003	2.888
	Ethnicity(4)	.990	.493	4.040	1	.044	2.691	1.025	7.068
	Ethnicity(5)	-.975	.141	47.568	1	.000	.377	.286	.498
	Ethnicity(6)	-.414	.104	15.906	1	.000	.661	.540	.810
	Attitude			236.765	2	.000			
	Attitude(1)	-1.247	.081	235.448	1	.000	.287	.245	.337
	Attitude(2)	-1.071	.131	67.330	1	.000	.343	.265	.442
	Constant	.012	.174	.005	1	.944	1.012		
Step 7(g)	AGE	.043	.003	192.203	1	.000	1.044	1.038	1.050
	EDUCATION			53.842	3	.000			
	EDUCATION(1)	.790	.139	32.529	1	.000	2.204	1.680	2.892
	EDUCATION(2)	.672	.137	24.117	1	.000	1.959	1.498	2.561
	EDUCATION(3)	.288	.131	4.845	1	.028	1.334	1.032	1.723
	WEALTH INDEX			13.395	4	.009			
	WEALTH INDEX(1)	-.210	.106	3.899	1	.048	.811	.658	.998
	WEALTH INDEX(2)	-.022	.101	.045	1	.832	.979	.802	1.194
	WEALTH INDEX(3)	.140	.099	2.019	1	.155	1.151	.948	1.396
	WEALTH INDEX(4)	.124	.098	1.599	1	.206	1.132	.934	1.371
	Region			201.546	10	.000			
	Region(1)	-.584	.163	12.841	1	.000	.558	.405	.768
	Region(2)	.778	.238	10.698	1	.001	2.177	1.366	3.470
	Region(3)	.114	.103	1.228	1	.268	1.121	.916	1.371
	Region(4)	.601	.113	28.440	1	.000	1.824	1.462	2.275
	Region(5)	1.843	.528	12.181	1	.000	6.317	2.244	17.784

	Region(6)	.581	.140	17.091	1	.000	1.787	1.357	2.354
	Region(7)	.753	.128	34.846	1	.000	2.124	1.654	2.727
	Region(8)	.003	.151	.000	1	.985	1.003	.746	1.348
	Region(9)	.891	.127	48.911	1	.000	2.437	1.899	3.128
	Region(10)	1.392	.149	87.606	1	.000	4.022	3.005	5.382
	Religion			37.817	2	.000			
	Religion(1)	.483	.081	35.401	1	.000	1.621	1.382	1.900
	Religion(2)	-.269	.253	1.128	1	.288	.764	.466	1.255
	Ethnicity			284.321	6	.000			
	Ethnicity(1)	.687	.335	4.191	1	.041	1.987	1.030	3.835
	Ethnicity(2)	1.364	.145	88.850	1	.000	3.913	2.947	5.197
	Ethnicity(3)	.874	.094	87.113	1	.000	2.397	1.995	2.879
	Ethnicity(4)	1.008	.497	4.113	1	.043	2.740	1.034	7.259
	Ethnicity(5)	-.988	.141	48.873	1	.000	.372	.282	.491
	Ethnicity(6)	-.397	.104	14.484	1	.000	.672	.548	.825
	Attitude			238.048	2	.000			
	Attitude(1)	-1.260	.082	236.906	1	.000	.284	.242	.333
	Attitude(2)	-1.060	.131	65.369	1	.000	.346	.268	.448
	Constant	.023	.174	.018	1	.893	1.024		
Step 8(h)	AGE	.043	.003	193.703	1	.000	1.044	1.038	1.051
	EDUCATION			46.372	3	.000			
	EDUCATION(1)	.762	.139	30.007	1	.000	2.143	1.631	2.815
	EDUCATION(2)	.651	.137	22.510	1	.000	1.917	1.465	2.508
	EDUCATION(3)	.295	.131	5.094	1	.024	1.343	1.040	1.736
	WEALTH INDEX			14.454	4	.006			
	WEALTH INDEX(1)	-.338	.124	7.464	1	.006	.713	.559	.909
	WEALTH INDEX(2)	-.148	.119	1.546	1	.214	.862	.682	1.089
	WEALTH INDEX(3)	.016	.116	.020	1	.888	1.016	.809	1.277
	WEALTH INDEX(4)	.006	.114	.003	1	.958	1.006	.805	1.258
	Region			202.850	10	.000			
	Region(1)	-.656	.167	15.508	1	.000	.519	.375	.719
	Region(2)	.760	.238	10.226	1	.001	2.139	1.342	3.409
	Region(3)	.041	.109	.141	1	.707	1.042	.842	1.289
	Region(4)	.545	.116	22.149	1	.000	1.725	1.374	2.164
	Region(5)	1.808	.525	11.847	1	.001	6.099	2.178	17.078
	Region(6)	.520	.143	13.155	1	.000	1.683	1.270	2.229
	Region(7)	.684	.132	27.043	1	.000	1.982	1.532	2.566
	Region(8)	-.060	.154	.152	1	.697	.942	.697	1.273
	Region(9)	.880	.127	47.629	1	.000	2.410	1.877	3.095
	Region(10)	1.388	.149	87.158	1	.000	4.006	2.993	5.360
	Religion			37.134	2	.000			
	Religion(1)	.479	.081	34.753	1	.000	1.614	1.376	1.892
	Religion(2)	-.267	.253	1.114	1	.291	.766	.466	1.257
	Ethnicity			280.508	6	.000			

Ethnicity(1)	.641	.336	3.639	1	.056	1.898	.983	3.667
Ethnicity(2)	1.364	.145	88.840	1	.000	3.911	2.945	5.193
Ethnicity(3)	.855	.094	82.584	1	.000	2.351	1.955	2.827
Ethnicity(4)	1.010	.494	4.174	1	.041	2.745	1.042	7.230
Ethnicity(5)	-.982	.141	48.278	1	.000	.375	.284	.494
Ethnicity(6)	-.412	.105	15.541	1	.000	.662	.539	.813
Attitude			235.974	2	.000			
Attitude(1)	-1.255	.082	234.710	1	.000	.285	.243	.335
Attitude(2)	-1.062	.131	65.631	1	.000	.346	.267	.447
RESIDENCE(1)	.223	.108	4.271	1	.039	1.250	1.012	1.545
Constant	.019	.174	.012	1	.914	1.019		

a Variable(s) entered on step 1: Ethnicity.
b Variable(s) entered on step 2: Attitude.
c Variable(s) entered on step 3: AGE.
d Variable(s) entered on step 4: Region.
e Variable(s) entered on step 5: EDUCATION.
f Variable(s) entered on step 6: Religion.
g Variable(s) entered on step 7: WEALTH INDEX.
h Variable(s) entered on step 8: RESIDENCE.

Iteration History(a,b,c)

Iteration		-2 Log likelihood	Coefficients
			Constant
Step 0	1	12603.728	1.169
	2	12543.841	1.331
	3	12543.724	1.339
	4	12543.724	1.339

a Constant is included in the model.
b Initial -2 Log Likelihood: 12543.724
c Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	10686.173(a)	.140	.219
2	10225.136(a)	.172	.269
3	9950.387(a)	.190	.298
4	9678.101(a)	.208	.325
5	9606.659(a)	.213	.332
6	9568.419(a)	.215	.336
7	9555.097(a)	.216	.338
8	9550.805(a)	.216	.338

a Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Classification Table[a]

Observed			Predicted Harmful No	Predicted Harmful Yes	Percentage Correct
Step 1	Harmful	No	725	1825	28.4
		Yes	450	9275	95.4
	Overall Percentage				81.5
Step 2	Harmful	No	665	1885	26.1
		Yes	283	9442	97.1
	Overall Percentage				82.3
Step 3	Harmful	No	621	1929	24.4
		Yes	273	9452	97.2
	Overall Percentage				82.1
Step 4	Harmful	No	851	1699	33.4
		Yes	407	9318	95.8
	Overall Percentage				82.8
Step 5	Harmful	No	854	1696	33.5
		Yes	385	9340	96.0
	Overall Percentage				83.0
Step 6	Harmful	No	863	1687	33.8
		Yes	377	9348	96.1
	Overall Percentage				83.2
Step 7	Harmful	No	867	1683	34.0
		Yes	383	9342	96.1
	Overall Percentage				83.2
Step 8	Harmful	No	858	1692	33.6
		Yes	380	9345	96.1
	Overall Percentage				83.1

a. The cut value is .500

Hosmer and Lemeshow Test

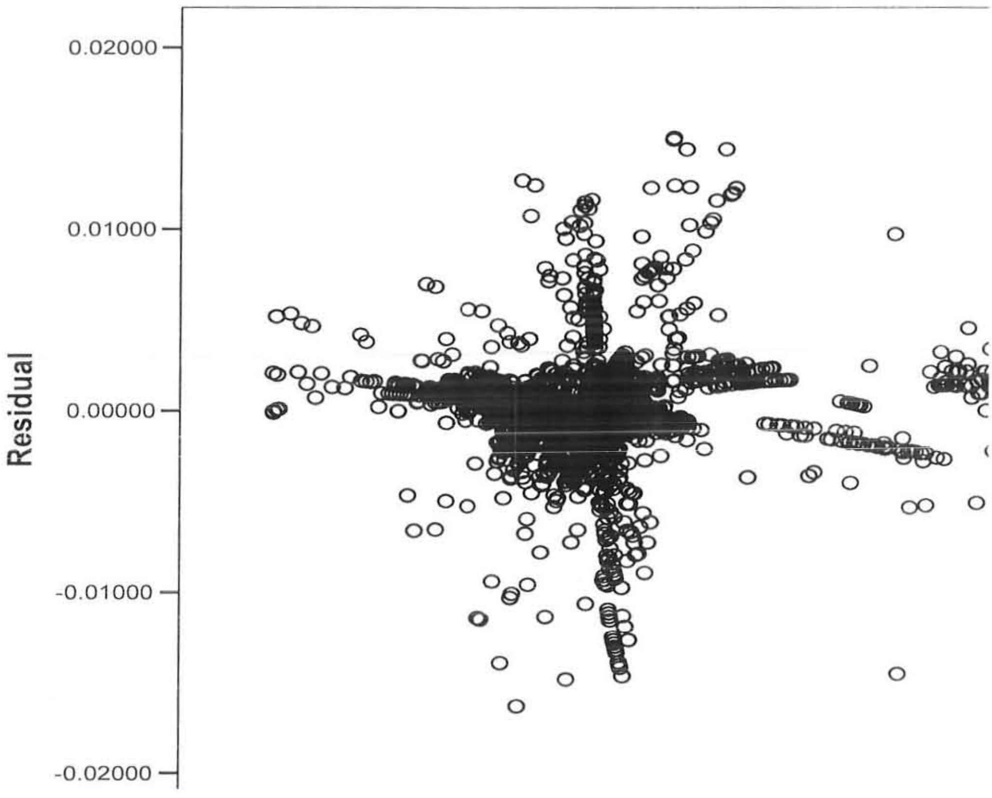
Step	Chi-square	df	Sig.
1	.000	2	1.000
2	8.491	5	.131
3	30.503	8	.000
4	37.015	8	.000
5	16.958	8	.031
6	13.889	8	.085
7	7.626	8	.471
8	14.576	8	.068

Contingency Table for Hosmer and Lemeshow Test

		Harmful = No		Harmful = Yes		
		Observed	Expected	Observed	Expected	Total
Step 1	1	1840	1840.000	4343	4343.000	6183
	2	399	399.000	1395	1395.000	1794
	3	51	51.000	218	218.000	269
	4	290	290.000	4082	4082.000	4372
Step 2	1	97	92.332	120	124.668	217
	2	1566	1564.138	3007	3008.862	4573
	3	354	366.507	1063	1050.493	1417
	4	228	240.096	1454	1441.904	1682
	5	225	212.788	1928	1940.212	2153
	6	58	43.948	519	533.052	577
	7	52	60.192	1947	1938.808	1999
Step 3	1	585	577.986	735	742.014	1320
	2	432	474.506	809	766.494	1241
	3	480	421.878	814	872.122	1294
	4	310	332.418	926	903.582	1236
	5	242	262.423	1017	996.577	1259
	6	204	190.536	1079	1092.464	1283
	7	164	139.487	1083	1107.513	1247
	8	83	94.814	1140	1128.186	1223
	9	52	58.081	1263	1256.919	1315
	10	28	27.871	1172	1172.129	1200
Step 4	1	526	594.645	742	673.355	1268
	2	524	474.747	742	791.253	1266
	3	429	406.476	835	857.524	1264
	4	347	335.553	917	928.447	1264
	5	225	259.623	1035	1000.377	1260
	6	215	191.315	1052	1075.685	1267
	7	150	139.164	1110	1120.836	1260
	8	90	96.150	1173	1166.850	1263
	9	48	54.146	1212	1205.854	1260
	10	26	28.180	1220	1217.820	1246
Step 5	1	592	626.845	667	632.155	1259
	2	488	483.805	780	784.195	1268
	3	417	392.083	831	855.917	1248
	4	330	326.995	933	936.005	1263
	5	240	255.551	1030	1014.449	1270
	6	209	183.594	1039	1064.406	1248
	7	152	136.002	1115	1130.998	1267
	8	79	91.787	1185	1172.213	1264
	9	48	53.901	1209	1203.099	1257
	10	25	29.438	1249	1244.562	1274
Step 6	1	622	638.053	641	624.947	1263

	2	482	484.676	779	776.324	1261
	3	428	396.696	832	863.304	1260
	4	302	326.234	960	935.766	1262
	5	261	251.306	1001	1010.694	1262
	6	192	183.220	1068	1076.780	1260
	7	145	130.473	1121	1135.527	1266
	8	69	86.724	1193	1175.276	1262
	9	54	54.475	1209	1208.525	1263
	10	25	28.144	1234	1230.856	1259
Step 7	1	634	642.663	629	620.337	1263
	2	493	486.066	769	775.934	1262
	3	403	397.711	862	867.289	1265
	4	312	322.577	950	939.423	1262
	5	256	250.022	1006	1011.978	1262
	6	185	182.906	1077	1079.094	1262
	7	148	129.593	1114	1132.407	1262
	8	70	86.002	1192	1175.998	1262
	9	54	54.400	1208	1207.600	1262
	10	25	28.059	1231	1227.941	1256
Step 8	1	631	643.266	631	618.734	1262
	2	487	487.483	774	773.517	1261
	3	417	398.907	845	863.093	1262
	4	315	321.819	948	941.181	1263
	5	243	250.275	1023	1015.725	1266
	6	181	182.533	1082	1080.467	1263
	7	161	129.292	1101	1132.708	1262
	8	69	85.023	1193	1176.977	1262
	9	52	53.913	1210	1208.087	1262
	10	24	27.489	1231	1227.511	1255

Scatter Plot for Age



Predicted probabilities

Results of diagnostic tests for outliers and influential values

	Minimum	*Maximum*
Normalized Residuals	*-2.135*	*1.49841*
Cook's Influence Statistics	*0.00001*	*0.0112*
DFBETA for Constant	*-0.00692*	*0.00799*
DFBETA for Age	*-0.00016*	*0.00008*
DFBETA for Region	*-0.00042*	*0.00045*
DFBETA for Educational level	*-0.00045*	*0.00253*
DFBETA for Religion	*-0.00171*	*0.00058*
DFBETA for Ethnicity	*-0.00004*	*0.00004*
DFBETA Place of residence	*-0.00427*	*0.00296*
DFBETA for Wealth Index	*-0.00128*	*0.00118*
DFBETA for attitude on harm.trad.practices	*-0.00161*	*0.0014*

DECLARATION

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

Name: **Zemene Yohannes Ayalew**

Signature: ....

Place: **Faculty of Science, Addis Ababa University**

Date: **June, 2009**

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

.... 03/07/2009

Fentaw Abegaz, Ph.D.