College of Natural Science

Department of Statistics

Teenagers Fertility Variation among Regional States of Ethiopia; Application of Multilevel Logistic Regression and Population Averaged Models.

By:

Adane Erango

A Thesis Submitted to Jimma University School of Graduate Studies College of Natural Sciences Department of Statistics In Partial Fulfillment of The Requirements For The Degree of Masters of Science In Biostatistics.

Jimma, Ethiopia

November 2018

TEENAGER FERTILITY VARIATION AMONG REGIONAL STATES OF ETHIOPIA; APPLICATION OF MULTILEVEL LOGISTIC REGRESSION AND POPULATION AVERAGED MODELS.

MSc. THESIS

BY:

ADANE ERANGO

ADVISOR: AGATAMUDI LAKISHMANA RAO (PhD)

CO-ADVISOR: YASIN NEGASH (MSc.)

JIMMA, ETHIOPIA

December, 2018

Declaration

I declare that the thesis is the result of my work, has not been submitted to any other university for achieving any academic degree or diploma awards and all source of materials used for the thesis have been duly acknowledged. I have submitted this thesis to Jimma University in the partial fulfillment for the requirements of Degree of Master of Science in Biostatistics. The thesis can be deposited in the university library to be made available to borrowers for reference.

Declared by:

Adane Erango

Date**:** ______________/______________/______________ Signature: ______________________

APPROVAL SHEET

COLLEGE OF NATURAL SCIENCE DEPARTMENT OF STATISTICS, SCHOOL OF GRADUATE STUDIES, JIMMA UNIVERSITY

As research advisors, we hereby certify that we have read the thesis prepared by Adane Erango under our guidance, entitled ***''Teenagers Fertility Variation among Regional States of Ethiopia; Application of Multilevel and Population averaged Models''***, in its final form and have found that its format, citations, and bibliographical style are consistent and acceptable.

AGATAMUDI LAKSHMANA RAO (PhD) ________________ ____________________

Advisor **Signature** **Date**

YASIN NEGASH (MSc.) ________________ ____________________

Co-Advisor **Signature** **Date**

As the members of MSc. thesis examiners board of open defense examination, we certify that we have read and evaluated the thesis and examined the candidate. Hence, we recommend that it can to fulfill the requirements for the degree of Master of Science in Biostatistics.

______________________________ _________________ _________________

Name of Chairman **Signature** **Date**

______________________________ _________________ _________________

Name of Internal Examiner **Signature** **Date**

______________________________ _________________ _________________

Name of External Examiner **Signature** **Date**

______________________________ _________________ _________________

Name of Advisor **Signature** **Date**

______________________________ _________________ _________________

Name of Co-Advisor **Signature** **Date**

ABSTRACT

Background: Teenage fertility is defined as a teenage girl, usually within the ages of 13-19, becoming fertile. Worldwide, adolescents suffer from a disproportionate share of early marriage, unwanted pregnancies, unsafe abortions, and sexually transmitted infections. Sub-Saharan African countries have high total and teenage fertility rates compared to the rest of the world. Ethiopia has a high incidence of unwanted pregnancies and incomplete and risky/septic abortions, particularly among adolescents.

Objectives: The general objective of this study is to assessing teenager fertility variation among regional states of Ethiopia.

Methods: This study has been used data from the 2016 Ethiopia Demographic and Health Survey (2016 EDHS). The researcher has been selected all teenagers aged 15-19 years of age at time of interview. The total sample of this study contains 3498 teenagers from nine regional states and two administrative cities in Ethiopia. The population of this study have been included all female teenagers aged 15-19 years of age at the time of interview. Multilevel Logistic and Population Average Models have been applied to assess the determinants and teenagers fertility variation in regional states of Ethiopia.

Results: The total of 3498 teenagers from nine regional states and two administrative cities in Ethiopia were eligible for this study. Among these eligible teenagers, 359 (10.263%) teenagers have been born child at the time of interview. From the significant regions the highest OR was seen in Harari(OR=4.44) and the lowest OR was seen in Affar(2.6) regions respectively.

Conclusion: Multilevel Logistic Regression with random slope and random intercept is found to be a good model to fit, compared with Random intercept models and Logistic regression model. From Marginal (GEE) models the model with independent correlation structure is good to fit the given data. The variables: Religion, Education Level, Wealth index, Contraceptive method, Knowledge of Ovulatory cycle and Exposure to any mass media were found significant.

Key words: Teenage fertility, Generalized Linear Models, Generalized Estimating equation, multilevel logistic regression, Cluster-Specific Variation and Repeated Measurements.

ACKNOWLEDGEMENT

I thank the **Lord God** for the protection rendered to me during the period of study, and for the knowledge and capability to write this thesis, and I thank **St**. Mary the mother of Christ. I thank my advisors Dr. Agatamudi Lakshmana Rao and Mr. Yasin Negash (MSc.) who worked hard towards my successful completion of the study. I therefore appreciate for the time given to me every time I approached them for advice and guidance.

I thank Jimma University for fully funded sponsorship of this study and all staffs of JU specially my staff members for their help and constructive advice. I thank department of statistics and all staffs there.

I also thank my parents Mr. Erango Boyamo, even I lost him and Mrs. Lawo Madebo for the funding to accomplish this study. I appreciate the support given to me and big thanks to all my family members, brothers, sisters who have been a source of encouragement and support, May the Lord God reward them with a gift of life and success in whatever they do.

Table of contents

Contents **Page**

List of Tables

Table **page**

List of Figures

Acronyms

AIC	Akaki-Information Criteria
AIDS	Acquired Immune Deficiency Syndrome
ASFR	Age-Specific fertility rate
BIC	Bayesian Information Criteria
CSA	Central Statistical Agency
EAs	Enumeration Area
EDHS	Ethiopia Demographic Health Survey
DBS	Dried Blood Spot
FGC	Female Genital Cutting
GEE	Generalized Estimating Equation
GLM	Generalized Linear Model
HIV	Human Immune Virus
MoH	Ministry of Health
PHC	Population and Housing Census
MQL	Marginal Quasi Likelihood
PL	Pseudo Likelihood
STIs	Sexually Transmitted Infections
TFR	Total Fertility Rate (Teenage Fertility Rate)
UNFPA	United Nations Population Fund
UNICEF	United Nations Children’s Fund
WHO	World Health Organization

CHAPTER ONE

Introduction

1.1. Background

Teenage fertility is defined as a teenage girl, usually within the ages of 13-19, becoming fertile. The term in everyday speech usually states that girls who have not reached legal adulthood, which varies across the world, who become fertile (1). Adolescence is the transitional period from childhood to adulthood characterized by significant physiological, psychological and social changes. The term "adolescent" is often used synonymously with "teenager" (2).This definition is not common in many societies, where the transition from child to adult was traditionally rapid, often marked by a special event with a symbolic or educational aspect (3). Teenage childbearing and the consequences associated with it remain a major concern worldwide (4). World Health Organization defines the age group 10-19 years of age as adolescents, 13-19 years of age as teenagers and 15-24 years of age as youth. Thus this implies that teenagers are included in adolescence. Those in the age group 10-24 years are called young people (5). Adolescence is characterized by undeveloped behavioral decision-making, exploration, experimentation, subjection to peer influences, and lack of knowledge about disease and protective measures against it (6, 7).

Worldwide, adolescents injured from inconsistent share of early marriage, unwanted pregnancies, and risky abortions, sexually transmitted infections (STIs) including HIV/AIDS, female genital cutting, underfeeding and anemia, sterility, sexual and gender based violence, and other serious reproductive health problems(8,9). Half of the world's populations are under 25. Some 1.8 billion are aged 10-25, history's largest generation of adolescents, and about 85% live in the developing world. Most people become sexually active before their 20th birthday. From the girls in least developed countries 49% of girls marry before they turn 18. From young unmarried girls 10% – 40% have had an accidental pregnancy according to community studies (1). According to UNFPA report, each year an estimated 14 million adolescents between the ages of 15 and 19 give birth globally, of which more than 90% occurs in developing countries (10, 11). In many countries, unmarried adolescent mothers are likely to experience social ostracism, which may result in rejection by their family and peers. Complications from pregnancy and childbirth are the leading cause of death for adolescent girls between the ages of 15 and 19 in poor countries. Girls in this age group are twice as likely to die from pregnancy and childbirth-related causes, compared with greater than 19 years women. Children born to teenage mother are 50% more likely to die before the age of one year than those born to women in their twenties. Additionally, among teenagers who become pregnant only few of them seek antenatal and delivery care from health professionals (12, 13).

Sub-Saharan African countries have high teenage and total fertility rates, compared to the rest of the world. The level of average births of minimum and maximum among 15-19 year olds were 156 births per 1000 teenagers in nine out of ten sub-Saharan African countries (13). Teens observe older peers' earnings (income) and base own earnings expectations in part on these observations (14). When wages are low and there is not much wage growth over the lifetime, there is lower opportunity cost to having a child today versus waiting (14). Teen births are higher in areas with greater poverty and with a lower lifetime earnings profile, Echoes work from other disciplines describing the feelings of despair or hopelessness experienced by individuals living in poverty (14). As a result, family planning programs were useful to women in the reproductive age group 20-49 and teenagers were deprived of contraception services. The regional average rate of births per 1000 females 15–19 years of age is 143, fluctuating from 45 in Mauritius to 229 in Guinea. This is very high compared to the world average of 65. From Africa, in some Sub-Saharan countries, one female out of five adolescents gives birth each year, so almost all females are likely to have had a child by age 20. In some African countries, (30–40)% of all adolescent females experience motherhood before the age of 18(15). From the mid–1970s to the early 1990s a trend towards lower adolescent fertility rates in a number of countries was observed – the largest changes being found in Kenya and Senegal (16).

Ethiopia has a high frequency of unwanted pregnancies and incomplete and dangerous abortions, particularly among adolescents (17). According to EDHS 2016, In Ethiopia, the median age at first birth among women age 25-49 is 19.2 years. This means that half of women age 25-49 give birth for the first time before age 20. The median age at first birth seems to have changed little between 2000 and 2016. Among women age 25-49, median age at first birth was 19.0 years in 2000 and 2005, after which it increased slightly to 19.2 years in 2016(18).

Amir (19) applied logistic regression technique to assess factors associated with teenage marital pregnancy among Bangladeshi women. Tewodros (20), analyzed the Determinants of adolescent fertility in Ethiopia by applying logistic regression. And also Admias (21) applied Logistic Regression on Determinants of teenager's fertility EDHS (2011) Data in Ethiopia also there are another studies discussed in literature, But the models proposed for those studies fail to handle sources of variation concerned with geographical area in Bangladesh and regional variation in Ethiopia. I.e. the model fails to handle random effects or regional variation in teenage fertility. Hence in this study the investigator used both models namely multilevel logistic and population average regression models to handle these problems of variation.

1.2. Statement of the Problem

According to adolescent health consultant, James E Rosen (22), adolescents aged less than 16 years face four times the risk of maternal death than women aged in their 20s, and the death rate of their neonates is about 50% higher and According to the study an estimated 16 million girls aged between 15 and 19 give birth every year, with 95% of these births occurring in developing countries. This makes up 11% of all births of the worldwide. However, global averages mask important regional differences. Births to adolescents as a percentage of all births range from about 2% in China to 18% in Latin America and the Caribbean. Seven countries account for half of all adolescent births in World, which are: Bangladesh, Brazil, the Democratic Republic of the Congo, Ethiopia, India, Nigeria and the United States of America (23).

According to the study done by Sibusico(24) in Sub-Saharan African countries, East Africa had highly decreased from 33.1% at 1992 to 16.3% at 2011, Southern Africa had a moderate decreasing and increasing pattern and also East Africa(including Ethiopia) had low decreasing pattern. Even the study was used multilevel logistic model the study was not specific to Ethiopia. Anteneh(25)applied logistic regression in Determinants of Adolescent Fertility among Rural Women of Ethiopia using 2011 EDHS data, from 2,510 adolescents 14.4% had given birth at the time of interview and found that Adolescent Fertility was associated with: Education, Religion, sex of house hold, Wealth Index, Work status and contraceptive method. The models applied to this study was binary logistic but this model fails to handle cluster-specific variation.

Tewodros and Jemal (20) finds that the teenage fertility rate was 13.6% and another 3.1% were pregnant for the first time at the time of interview. Admias (21) applied Logistic Regression on Determinants of teenagers' fertility EDHS (2011) Data in Ethiopia. And finds that the highest OR=7.025 for uneducated teenage and lowest OR=0.600 for Orthodox followers. According to the study done by Anteneh(25) the variables:- Religion, Contraceptive method, Education level and wealth Index are significant at 5% level of significance.

In Ethiopia and Nigeria, more than 25% of fistula patients had become pregnant before the age of 15 and more than 50% before the age of 18. Although the problem can be corrected with surgery, treatment is not widely available in most countries where fistula occurs and millions of women are left to suffer with a condition that leads to incontinence, bad odors and other side-effects including psychological problems and social isolation (23).

For teenage fertility and unwanted child bearing: low knowledge of ovulatory cycle, low contraceptive used and low knowledge of family planning were due to the shortage of education and communication with the health professionals, improve mass media message follow up about those factors, are the main reason in Ethiopia (25) and which is the same with the study in Bangladish done by Md. Hasinur(26). As listed above, many studies have been studied in the level of: world, Africa and Ethiopia, but some studies were out of this specific, area and age group; the models that the studies used were fail to handle random (cluster-specific) variation. To handle this problem the researcher used the two usual models namely multilevel logistic and population average logistic regression.

Research Question:

1. What factors of determinants are influential for teenagers' fertility in Ethiopia?
2. Is there a significant regional teenage fertility Variation in Ethiopia?
3. Are there variables that vary from region to region?

1.3. Objectives

1.3.1. General Objective:

- To assess teenagers' fertility variation among regional states of Ethiopia. Application of multilevel logistic and population average models.

1.3.2. Specific objectives

1. To identify the determinants of teenagers' fertility.
2. To find whether there is a significant regional teenage fertility Variation in Ethiopia.
3. To find out the variables those vary through region to region.

1.4. Significance of the Study

The outcome obtained from this study will be useful in many ways. Governmental and non-governmental organizations will take remedial measures and set appropriate strategy to reduce practice of teenage fertility and aware teenagers' about the problems of teenage fertility. The results of the study will also be helpful for policy making, monitoring and evaluating the activities for the government and different concerned agencies. And it helps individuals (women and men) to have sufficient awareness about the teenage fertility.

CHAPTER TWO

2. Literature Review

Over View of Teenagers Fertility

The 1.2 billion adolescents aged 10-19 around the world make up 16% of the world's population. The majority (86%) of adolescents live in developing countries. By the time they are 19 years old, half of adolescent girls in developing countries are sexually active, about 40% are married and close to 20% have children (27). There were 21 million pregnancies among adolescent girls aged 15–19 years in developing countries in 2016; nearly half (49%) were unintended (43% in Asia, 45% in Africa and 74% in Latin America and the Caribbean). An estimated 23 million adolescent girls have an unmet need for modern contraception and are at risk of unintended pregnancy. Additionally, 777,000 girls under the age of 15 gave birth in the same year (27). Likewise, Singh and Darroch (28) in their study on adolescent pregnancy and childbearing levels and trends in developed countries, they found that Japan and most Western European countries have very low pregnancy rates which are below 40 per 1000 adolescents. A lag in falling teenage birth rates was noticed in some South Asian countries which were characterized by highest traditional rates of adolescent marriage as a result, birth rates only started falling in the mid-1970s (28).

In a study on social interactions and contemporary fertility transition, (29) which reported that rapid fertility transitions have been observed in many developing countries. In the 1970s, teenagers constituted about 20 to 25 percent of the total population of developing countries (14). Globally, it has been indicated that more than 1.5 billion people are under the age of 25(28). Between the early 1960s and late (1985-90) total fertility rate of the developing world as a whole declined by approximately 36 percent that is from 6.0 to 3.8 births per woman(21).

The World Bank data set of 2017, reported the data for Adolescent fertility rate (births per 1000 women aged 15-19), for 57 years. According to this data set, from 1960 to 1961 the highest fertility was recorded in Cote D'Ivore with fertility rate of 229.12 per 1000 Adolescents in 1961, from 1962 to 1979 the highest fertility was recorded in Angola with fertility rate of 225.6 per 1000 adolescents in 1979, and from 1980 to 2016 the highest fertility rate was recorded in Niger with fertility rate of 194 per 1000 adolescents (30). All the listed countries are from developing

Countries indicating high fertility rate was recorded in those Countries and Ethiopia is one of developing countries. And likewise, from 1960 to 1970 the lowest fertility rate was recorded in Japan with fertility rate of 4.0944 per 1000 adolescents, from 1971 to 1973 the lowest fertility was recorded in Korea with fertility rate 4.2494 per 1000 adolescents, from1974 to 1984 the lowest fertility was recorded in Japan with fertility rate of 4.1844 per 1000 adolescents, And from 1985 to 2016 the lowest fertility was recorded in Korea, with fertility rate of 0.2864 per 1000 adolescents (30). Although teenage fertility rates in developing countries are declining, the rates remain extremely high relative to those in the developed countries (31). This has been reinforced by Alli and Maharaji(32) who observed that the provision of youth friendly services is a relatively recent practice in developing countries.

In many African countries fertility has been declining despite an increase in teenage premarital fertility which has been noted in several studies (33,34,35). Fertility decline has been documented in some parts of sub-Saharan Africa especially in countries like Kenya, Zimbabwe, Botswana, Tanzania and South Africa (36, 37). Cohen (1998) (38) contends that the decline in fertility noticed in these countries is attributed to greater use of modern contraception and rising age at marriage. Although total fertility has been declining, an increase in premarital fertility has been reported in many countries (33,34, 35). This has been supported by (39) who observed that teenage fertility in sub-Saharan Africa continues to be higher compared to other regions. In 16 out of 22 sub-Saharan African countries, teenage fertility rates were over 150 births per 1000 women in the mid-1970s to early 1980s and from 1980s to 1990s, adolescent age specific fertility rates declined at a moderate rate (39).Women becoming mothers before their 20th birthday are usually considered at greater risk of health and social problems by both health practitioners and researchers (40).

According to(32), almost constant fertility rate (115 per 1000 adolescents) was recorded in Ethiopia from 1960 to 1967, the highest and almost constant fertility rate (120 or 121 per 1000 adolescents) was recorded from 1975 to 1985, and the smallest fertility rate (65 per 1000 adolescents) was recorded in 2016 (32). According to this report adolescent fertility rate was high in developing Countries and low in Developed Countries (32). As mentioned, the majority of Ethiopian girls marry in their teens, and the percentage of teenagers who have begun childbearing has been extremely large. And in the Ethiopian context, factors affecting teenage pregnancy estimated in 2000 were teenagers' age at first sex, knowledge and use of family planning methods, residence type, education and employment status, and exposure to mass media (41).

Francoise (42) focuses on Demographic and Behavioral Factors Associated with Adolescent Pregnancy in Switzerland using logistic regression model. Results shows that use of less-effective methods of contraceptive or nonuse was related to pregnancy and child bearing: Ever-users of rhythm or withdrawal were more likely than those who had never used these methods to have experienced a pregnancy and child bearing, as were never-users of condoms and those who did not practice contraception at last intercourse. Median birth intervals increase with increasing education and wealth. For example, birth intervals among women with more than a secondary education are 13.7 months longer than intervals among women with no education (47.7 months versus 34.0 months). Likewise, birth intervals among women in the highest wealth quintile are 10.9 months longer than those among women in the lowest quintile (43.0 versus 32.1 months)(43).

Tewodros and Jemal(20) analyzed the Determinants of adolescent fertility in Ethiopia by logistic regression. The study finds that teenage fertility rate was 13.6% and another 3.1% were pregnant for the first time at the time of interview. Delayed marriage or non-marriage and postpartum sterility were the determinants of fertility both in urban and rural part of the country while use of contraceptive was determinant in urban area. The other deferential of fertility were age at first marriage, women education, place of residence and age.

Empirical Review

Region: According to (21) Tigray, Oromia, Benshangul-Gumuz, and Harari have OR greater than one. A study conducted by kosunen(45) on the trends and regional variation in teenage pregnancy, abortion and fertility rates found that there were remarkable regional differences in teenage fertility rates in Finland. Rapid teenage fertility decline was reported in Finland where teenage pregnancy and abortion rates declined by more than half from 1970s to early 1990s and this was the lowest in the Nordic countries in 1996(45). Highest teenage fertility rates were reported in remote areas of the country and from the 1970s till 1990s these regional differences remained the same although significant decreases in the incident of teenage pregnancies were recorded in the country. Highest rates in remote areas are attributed to less effective use of contraceptives among teenagers and reduction of sex education in schools (45). This is similar to what this research seeks to explore, that is, if there are any geographical differences in teenage fertility in Ethiopia.

Religion: The main religions in Ethiopia are Orthodox Christianity (43% of women and 45% of men) and Muslim (31% each of women and men). Twenty-three percent of women and 22% of men are Protestants (18). According to (21) the religious differences the odds of currently being fertile in Orthodox and Muslim religion is less than one that they have OR=OR=0.600, OR=0.768 respectively as compared to other religion.

Ethnicity: Tewodros and Jemal(8) studied on determinants of adolescent fertility in Ethiopia and reported that for the major Ethnic groups in Ethiopia :Oromo (32.7%), Amhara (32.1%), Tigray (7.3%), Guragae (4.9%) and 23% for the remaining ethnic groups.

Sex of House hold: According to the study done by Anteneh (25) the sex of house hold had association with teenagers' fertility in chi-square test of association but not significantly associated with teenagers' fertility in regression analysis.

Residence: Teenagers in rural areas are three times more likely to have begun childbearing than their urban peers: 15% of rural teenagers have had a live birth or are pregnant, as compared with 5% of urban teenagers (43). By region, teenage childbearing is highest in Affar (23%) and Somali (19%) and lowest in Addis Ababa (3%) and Amhara (8%).

Education: Teenage childbearing decreases with increasing education. The percentage of teenagers who have begun childbearing rises from 3% among those with more than a secondary education to 12% among those with a primary education and 28% among those with no education (43). In particular an adolescent with no more than primary schooling is 13.8% higher than an adolescent with at least secondary education. The OR for uneducated teenage was 7.025 and OR for teenage in primary education level was 1.678(21). The number of children per woman declines with increasing education. Women with no education have 3.8 more children than women with more than a secondary education (5.7children versus 1.9 children)(44).

Wealth Index: There are regional variations in wealth. The wealthiest households are concentrated in Addis Ababa (100%) and the poorest households in the Affar Region (74%)(18). Teenage childbearing is less common in the wealthiest households: 6% of women age 15-19 from the highest wealth quintile have begun childbearing, as compared with 24% of those from the lowest quintile (43). Adolescents from the second, middle and fourth wealth quintile were more likely to get pregnant compared to those from the highest wealth quintile (45). According to (43) women in the lowest wealth quintile have 3.8 more children than women in the highest wealth quintile (6.4 children versus 2.6 children). Across regions, the median birth interval ranges from 25.1 months in Somali to 47.6 months in Addis Ababa.

Knowledge of Ovulatory Cycle: to say the least, when planning a family. However, not everyone will have or make the time to learn about their reproductive hormones and natural ways to predict fertility. From the fertile teenagers 16.8% have no knowledge of ovulatory cycle and 10.4% knows at any time (21).

Contraception: Any deliberate practice undertaken to reduce the risk of conception by sexually active women (and their male partners) is considered as contraception. The tool used to prevent or reduce the frequency of conception is known as contraceptive. From the fertile teenagers 9.8% used no methods and 7.7% used modern methods (21).

Mass Media: Nearly three in four (74%) women and 62% of men have no access to radio, television, or newspapers on a weekly basis. There are wide variations by place of residence in median number of years of education completed. Urban women have completed a median of 7.7 years of education, while the median among rural women is 0.0(18). According to (21) from a fertile teenage 8.9% teenagers follow any mass media and 18.4% not follow any mass media.

Methodological Review

Amir and Sayem(19) applied the logistic regression technique to assess factors associated with teenage marital pregnancy among Bangladeshi women. This study revealed that participants aged 20-24 years had higher likelihood (OR 1.971, 95% CI 1.132 to 3.434), whereas participants aged 25-29 years had lower likelihood (OR 0.054, 95% CI 0.016 to 0.190) of experiencing teenage marital pregnancy compared to participants aged 15-19 years. In addition, participants desired for greater than 2 children had significant higher odds (OR 3.573, 95% CI 1.910 to

6.684) and participants born in urban area had significant lower odds (OR 0.458, 95% CI 0.228 to 0.919) for teenage marital pregnancy.

According to the study done by four Authors (46) in Assossa General Hospital Teenage pregnancy is estimated as 20.4%. The median age of subjects, at first sexual intercourse and at first marriage, was 16 and 17 years respectively. High proportion of (46.8%) teenagers had engaged in premarital sex. Among sexually active teenage females, 46.7% experienced their first sexual encounter by coercion. Being young [AOR= 0.21, 95% CI= 0.06-0.67], single [AOR= 0.06, 95%CI= 0.03-0.12], housemaid [AOR= 3.93, 95% CI=1.71-9.04] and use of family planning [AOR= 2.39, 95% CI= 1.20-4.75] have statistically significant association with teenage pregnancy.

Amsalu Arega (47), Modeling Delivery Care Service Utilization of Mothers in Ethiopia. And compares three alternative models; Population Average (GEE), Alternating Logistic Regression(ALR) and Generalized Linear Mixed (GLMM) Models and finds ALR fit the data well. Chiavegatto and Kawachi(48) analyzed Income inequality is associated with adolescent fertility in Brazil: a longitudinal multilevel analysis of 5,565 municipalities and found that positive association between income inequality and adolescent fertility.

Yobyo and, Kahsay (49) analyzed Why Do Women Deliver at Home? Multilevel Modeling of Ethiopian National Demographic and Health Survey Data and found that lower Educational levels, ANC visits, non-exposure to media, Higher parity and perceived distance problem to reach health facilities were positively associated with home delivery. Daniel (50), analyzed Modeling Determinants of Low Birth Weight for Under-Five Children in Ethiopia applied three models and found that Alternating logistic regression model was good fits the data for population-averaged effects of the given factors on birth weight than generalized estimating equation model and generalized linear mixed model with two random intercepts. Dechasa(51) applied multilevel logistic regression to analyze Regional Heterogeneity Of Under-Five Child Mortality In Ethiopia. Abebe(52) applied Multilevel logistic regression to analyze Obstetric Fistula In Ethiopia.

Due to the fact that Ethiopia, is one of developing countries and youthful population, the risk also well known in this country. Therefore the study considered only the teenage fertility in the age group 15-19 socio-cultural, economic and environmental determinants including contraceptive method used and knowledge of ovulatory cycle and most of the unintended births are occurred at the teenage stage. Therefore, special emphasis is given to the socio-cultural, economic and environmental determinants of teenage fertility.

As shown in background and literature: there were so many studies have been done in international and national level but most of the studies are epidemiological, specific to some area and based on the general populations of reproductive age, and some National studies applied Logistic Regression but this model fails to handle cluster-specific or subject-specific variation.

CHAPTER THREE

3.1.Data Sources and Methodology

3.1.1. Source of Data

This study has been used data from The 2016 Ethiopia Demographic and Health Survey (2016 EDHS). The 2016 Ethiopia Demographic and Health Survey (EDHS) is the fourth Demographic and Health Survey conducted in Ethiopia. It was implemented by the Central Statistical Agency (CSA) at the request of the Ministry of Health (MoH). Data collection took place from January 18, 2016, to June 27, 2016.

3.1.2. Sample Design

The sampling frame used for the 2016 EDHS is the Ethiopia Population and Housing Census (PHC), which was conducted in 2007 by the Ethiopia Central Statistical Agency (CSA). The census frame is a complete list of 84,915 enumeration areas (EAs) created for the 2007 PHC which covers on average of 181 households. With the exception of EAs in 6 zones of the Somali region, each EA has accompanying cartographic materials. These indicate geographic locations, boundaries, main access, and landmarks in or outside the EA that help to identify the EA. In Somali, cartographic frames were used in three zones, where a sketch map that indicates the EA geographic boundaries is available for each EA; in the remaining six zones, satellite image maps were used to provide a map for each EA.

The 2016 EDHS sample was stratified and selected in two stages. Before that each region was stratified into urban and rural areas and then Samples of EAs were selected independently. Implicit stratification and proportional allocation were applied at each of the lower administrative Levels, according to administrative units in different levels, and probability proportional to size selection at the first stage of sampling were used.

In the first stage, a total of 645 EAs (202 EAs in urban areas and 443 EAs in rural areas) were selected with probability proportional to the EA size and with independent selection in each sampling stratum. And then household listing operation was carried out in all the selected EAs for 1year (42). The resulting lists of households served as a sampling frame for the selection of households in the second stage. Some of the selected EAs were large. To minimize the task of household listing, each large EA selected for the 2016 EDHS was segmented. And then only one segment was selected for the survey, with probability proportional to the segment size. Household listing was conducted only in the selected segment, that is, a 2016 EDHS cluster is either an EA or a segment of an EA.

In the second stage of selection, a fixed number of 28 households per cluster were selected with an equal probability systematic selection from the newly created household listing. All women age 15-49 and all men age 15-59, who were either permanent residents of the selected households or visitors who stayed in the household the night before the survey.

From the selected EAs a total of 18,008 households were selected for the sample, of which 17,067 were occupied. Of the occupied households, 16,650 were successfully interviewed, yielding a response rate of 98 percent. In the interviewed households, 16,583 eligible women were identified for individual interviews; interviews were completed with 15,683 women, yielding a response rate of 95 percent. A total of 14,795 eligible men were identified in the sampled households and 12,688 were successfully interviewed, yielding a response rate of 86 percent. In general, response rates were higher in rural than in urban areas, especially for men.

3.2.Study Population

The population of this study included all female teenagers aged 15-19 years of age at the time of interview, whereas the corresponding sample have been taken from these populations of all teenagers (aged 15-19 years of age) at the time of interview who participated in the survey.

3.3.Variables of the Study:

3.3.1. Dependent Variable

This study has been used, being fertile (teenage fertility (have ever given birth)) by the time of interview and not being fertile as a dependent variable. The dependent variable is a dichotomous random variable currently have ever give birth by the time of interview (coded as 1) and not being fertile (coded as 0).

3.3.2. Independent Variables

Predictor variables are those variables which are supposed to affect or determine a dependent variable. Since based on the reviewed literatures, some of the common predictors that are expected to influence on determinants of teenage fertility in Ethiopia were recorded as given below for the purpose of the analysis.

Table 3.1. *Independent variables*

Variables	**Factor categories**	
1.Place of Residence	0=Urban	
	1=Rural	
2.Region	1=Tigray	2=Affar
	3=Amhara	4=Oromiya
	5=Somali	6=Benishangul
	7=SNNP	8=Harari
	9=Gambela	10=Addis Ababa
	11=DireDawa	
3.Religion group of a teenager	0=Orthodox	
	1=Muslim	
	2=Other	
4. Knowledge of ovulatory cycle	0=Before and during the period	
	1=Afer the period ended	
	2=Middle of the cycle	
	3=At any time	
	4=Do not know	
5.Exposure to any mass media	0=No	
	1=Yes	

6.Working status of teenager	0=Not Working
	1=Working
7.Teenager's education level	0=No education
	1=primary
	2=secondary and above
8.Economic status of teenager	0=poor
	2=middle
	3=rich
9.Contraceptive method used status currently	0=no method
	1=modern method

3.4. Statistical Methods

3.4.1. Overview of Generalized Linear Models

Generalized linear models (GLMs) are used to fit fixed effect models to certain types of data that are not normally distributed. The word Generalized represents that the data is not limited to normally distributed and linear implies that models use a linear combination of variables to 'predict' the response(53). All generalized linear models have three components: The **random component** identifies the response variable fertility status and assumes a probability distribution for it. (54) The **systematic component** specifies the explanatory variables for the model. These enter linearly as predictors on the right-hand side of the model equation (54). The **link function** specifies a function of the expected value (mean) of fertility status, which the GLM relates to the explanatory variables through a prediction equation having linear form (53).

3.4.2. Logistic Regression

Logistic regression is a statistical technique for predicting the probability of an event, given a set of predictor variables. The binary logistic regression procedure empowers one to select the predictive model for dichotomous dependent variables. It describes the relationship between a fertility status and a set of explanatory variables. The explanatory variables may be continuous or discrete. The logistic model, as a non-linear regression model, is a special case of generalized linear model where the assumptions of normality and constant variance of residuals are not satisfied (55). Logistic regression analysis is the most popular regression technique available for modeling dichotomous variable (fertility status). The special case in which the response variable has only two categories is, of particular interest and lends itself to an especially nice treatment. This is because, with only two categories, there is essentially only one way to define the odds (56). Binary logistic regression model is used to investigate the effect of predictors on the probability of having teenagers child bear at the time of survey is defined as follows:

1) Dependent Variable is given as:

$$Y_{ij} = \begin{cases} 1 \text{ If Being fertile or Teenage having child} \\ 0 \text{ other wise} \end{cases} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots .3.00$$

i=1,2,….,M & j=1,2,….,N, Where M is being fertile teenage in each region j and N is the number of region.

Let π(x) denotes the proportion of success (Being fertile teenage):

P ($y_{ij} = 1$) = π, P ($y_{ij} = 0$) = 1-π, And y_i~Bernoulli (π)

The logistic Analysis is defined as follows.

Let $\mathbf{X_{nx(k+1)}}$denote the binary logistic regression data matrix of k predicator variables of the Teenage fertility is given as:

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1K} \\ 1 & x_{21} & \ddots & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \sim n \times (k+1), \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} \sim (k+1)x1 \ldots\ldots\ldots\ldots\ldots\ldots\ldots .....3.01$$

It is mathematical modeling approach that can be used to describe the relationship of several predictor variables $X_1, X_2, \ldots . X_K$ to a dichotomous dependent variable teenage fertility where teenage fertility is typically coded as 1 or 0 for two possible categories. Consider a group of k predictor variables denoted by the vector X’=$X_1, X_2, \ldots X_k$. Then the conditional probability that

i^{th} teenager has experienced by child bearing given the vector of predictor variables is denoted by $p(y = 1/x) = \pi(x)$, Then, the logistic regression model for explaining data is given by;

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k)}{1 + \exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k)} = \frac{e^{(X_i \beta)}}{1 + e^{(X_i \beta)}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.02$$

Where: $\pi_i (i = 1,2, \ldots, n)$ is the probability of i^{th} teenager bearing child at the time of survey given the vector of predictors (x_i).

By algebraic manipulation, the logistic regression equation can be written in terms of an odds ratio for success:

Then, the logit(π) or $\log\left(\frac{\pi}{1-\pi}\right)$ of teenager bearing child(y=1) is modeled as a linear function of the explanatory variables as:

$$\text{Log}\left[\frac{\pi(x)}{1 - \pi(x)}\right] = \text{Log}[\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 \ldots + \beta_k X_k)]$$

$$= \quad \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \ldots\ldots\ldots\ldots\ldots\ldots 3.03$$

Where, $(0 \leq \pi(x) \leq 1)$**.** Bear in mind that natural logarithm is in base10 or base e; in our case it must to take natural logarithm with base **e** either of this it is better to use ln. Where β_0 is a Log-Odd of the equation, when other predictors kept as they have no effect on the model, and $\beta_1, \beta_2, \ldots, \beta_k$ are, the coefficients of the predictor variables, included in the model. The estimated logistic coefficients β_jth are interpreted as: - For every unit change in the predictor variables with respect to β's the value of Log-Odds increased by respective β by holding other predictors constant (57).

3.4.2.1.Assumptions of Logistic Regression

From the modern statistical approach the strength of inferences drawn is depends on the assumptions of the statistical model being satisfied. In order to have a valid analysis, our model has to satisfy the assumption of logistic regression listed below:-

1. The response must be categorical to have an output.
2. It assumes linear relationship between log-odds of response and predictor variables.

3. The predictor variables are not be regressed with each other OR there should not be severe co-linearity between independent variables.
4. The distribution of response assumed to be from the distribution of exponential family (like Poisson, Multinomial, Normal, Binomial and etc.).Binary logistic regression assumes binomial distribution of the response.
5. Groups for the predictors must to be mutually exclusive and exhaustive.
6. Compared to Linear regression larger sample is needed for logistic regression because maximum likelihood assumes large sample to have best estimate.

3.4.3. Coefficient Estimation in Logistic Regression

Based on the assumption stated above, the logistic regression needs to use maximum likelihood to estimate unknown parameters of logistic regression model.

3.4.3.1.Maximum Likelihood Estimation for Logistic Regression

Maximum likelihood is generally used to estimate the parameters for generalized linear models. The likelihood is simply the probability density computed from the observed data values with the parameters replaced by their estimates. Also it used to estimate the logistic regression model coefficients. The logic behind maximum likelihood estimation is to determine the values of β_j where j=0, 1… k which make the observed data most likely to have occurred. The method of maximum likelihood estimation is used very broadly in many statistical applications besides logistic regression. Maximum likelihood estimators often perform better than other types of estimation procedures in terms of being the most efficient use of data. Hence, maximum likelihood estimation is a very popular method of estimation in statistical practice (58). Consider the logistic regression model in equation (3.02) above: Then, we have the likelihood function with n observations and with y_i success and n-y_i failure.

$$\mathrm{Log}[\beta, Y = 1/X] = \prod_{i=1}^{n} \left[\frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)}\right]^{\sum_{i=1} y_i} \left[1 - \left(\frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)}\right)\right]^{1-\sum_{i=1} y_i} \quad \dots\dots\dots.\dots\dots\dots 3.04$$

And the log-likelihood function is

$$l = \sum_{i=1}^{n}\left[y_i \log\left(\frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)}\right)\right] + \left[\left(n - \sum_{i}^{n} y_i\right)\log\left(\frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)}\right)\right] \ldots\ldots\ldots.\ldots\ldots\ldots.\ldots.\ldots 3.05$$

Unfortunately, there do not exist formulas that give the estimates of β from a logistic regression in closed form as was the case in simple linear regression. Instead, iterative algorithms are needed to determine the maximum likelihood estimates of β. Many software packages have the ability to fit logistic regression models. The most popular algorithms for finding the maximum likelihood estimates are the Newton-Ralphson algorithm. The information in this case becomes (k+1)x(k+1) matrix of the partial second derivative with respect to the parameter β. The information matrix is inverted to give the covariance matrix for $\hat{\beta}$.

3.4.3.2. Odd and Odd Ratio in Logistic Regression.

For a fertile teenage probability π the odds are defined to be $\Omega = (\pi/1-\pi)$.The odds are non-negative, with Ω >1.0 when a success is more likely than failure. And the odds ratio is defined by the ratio of two odds which is $\theta = (\Omega 1/\Omega 2)$. Since an odds ratio is function of odds; as a result it is also non-negative as odds. For θ=1 corresponds to the predictors have no effect on teenagers fertility. When $1 < \theta < \infty$ subjects in a reference category are more likely to have a success than are subjects in interest category; that is, probability of success in reference (π1) > probability of success in interest category (π2) (46). In the logistic regression $\hat{\beta}$ s′ are related to the odds ratio by: - Log-odds=$\hat{\beta}$, Thus after some mathematical transformation the odd is estimated to be **exp ($\hat{\beta}$)** (58).

3.4.4. Variable Selection Method

The variables to be included in the model should be the minimum possible number that is parsimonious and bring optimum information. In our case the variable selection process begins with a univariable analysis of each variable. Tests to determine whether a systematic relation or association between each predictor variable with the response variable exists are made before the final model was selected. A univariable logistic regression and a likelihood ratio (LR) chi-square test would be employed to see the importance of each predictor variables to the outcome variable (59). Other approaches to variable selection are to use stepwise and forward selection procedure

In Forward selection procedure, we add terms sequentially until further additions do not improve the fit. The backward selection on the other hand begins with a complex model and sequentially removes terms. Stepwise selection procedure is the combination of forward selection and backward selection to identify the best model (59).

3.4.5. Goodness of fit in Logistic Regression

Once a model has been developed through various steps in estimating the parameters, there are several techniques involved in assessing the appropriateness, adequacy and usefulness of the model. At the beginning the importance of each of the explanatory variables would be assessed by carrying out statistical tests of significance of the parameters. Then the overall goodness of fit of the model would be assessed (54).

3.4.5.1.Likelihood Ratio Test

Likelihood ratio is defined as the ratio of two likelihoods. The logistic regression model is a special case of a generalized linear model. In GLM terminology, the likelihood-ratio statistic for this test is the deviance of the model is computed, which measures how close the predicted values from the fitted model match the actual values from the raw data(58).

The likelihood ratio test, also called log-likelihood test, it is based on the saturated model which is the model with the maximum number of parameters that can be estimated and the reduced model which is the model of interest with k parameters. Then the likelihood ratio:-

$\lambda = \left(\frac{L(\hat{\beta}_{max})}{L(\hat{\beta})}\right)$, then, the log-likelihood ratio becomes: - $\log(\lambda) = l(\hat{\beta}_{max}; y) - l(\hat{\beta}; y)$, Large value of $\log(\lambda)$ indicates a poor fit. And the deviance is defined as two time by the Log-likelihood ratio. Which is written as follows:-

$2 * (\text{Log} - \text{likelihood ratio}) = \text{Deviance} = 2\left(l(\hat{\beta}_{max}; y) - l(\hat{\beta}; y)\right)$, Large value of deviance indicates a poor fit. Also this log likelihood-ratio test uses the ratio of the maximized value of the likelihood function for the intercept only model L_0 over the maximized value of the likelihood function for the full model L_1. The likelihood test statistic is given by (60);

$$G^2 = -2\left[\frac{L_0}{L_1}\right] = -2[\log(L_0) - \log(L_1)] = 2(l_1 - l_0) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots .3.06$$

Where l_0 the log-likelihood of the model with the intercept term only and l_1 is the log-likelihood of the full model. The likelihood ratio statistic has a chi-square distribution and it tests the

H_O such that all logistic regression coefficients are zero except the constant.

H_1 such that at least one of the predictors is significantly related to the response variable.

The degrees of freedom are obtained by differencing the number of parameters in both models. It compared with chi-square value at the difference between degree of freedom of both model. And p-value indicates that the probability of the deviance based on chi-square is greater than the tabulated chi-square. If p-value is less than 5% level of significance leads to the rejection of **H_0** and concludes that, at least one of the coefficients of predictors is significantly different from zero.

3.4.5.2.The Hosmer and Lemeshow Test Procedure

The Hosmer–Lemeshow test uses a Pearson test statistic to compare the observed and fitted counts for this partition. The test statistic does not have exactly a limiting chi-squared distribution. However, Hosmer and Lemeshow noted that, when the number of distinct patterns of covariate values (for the original data) is close to the sample size, the null distribution is approximated by chi-squared with df = number of groups minus two(53). The mathematical formula for defining the test statistic is:-

$$\widehat{C} = \sum_{k=1}^{g} \frac{(O_k - E_k)}{V_k} = \sum_{k=1}^{g} \frac{(O_k - E_k)}{np_k(1-p_k)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots .3.07$$

Where O_k, E_k observed and expected number of events in the k^{th} group respectively and V_k is a variance correction factor for the k^{th} group . If the observed number of events differs from what is expected by the model, the statistic $\widehat{C}$ will be large and there will be evidence against the H_O. This statistic has an approximate Chi-Squared distribution with degrees of freedom equal to number of groups minus two.

3.4.5.3.Influence Diagnostics for Logistic Regression

As in ordinary regression, some observations may have too much influence in determining the parameter estimates. The fit could be quite different if they were deleted. Whenever a residual indicates that a model fits an observation poorly, it can be informative to delete the observation and re-fit the model to the remaining ones. However, a single observation can have a more exorbitant influence in ordinary regression than in logistic regression, since ordinary regression has no bound on the distance of y_i from its expected value. These diagnostics are algebraically related to an observation's leverage(61). Influence diagnostics for each observation include:

Cook's Distance: - Cook‗s distance is designed to measure the shift in $\hat{\beta}$ when a particular observation is omitted. It is a combined measure of the impact of that observation on all regression coefficients (63).Notice that D_i is large if the standardized residual is large and if the data point is far from the cancroids of the X-space. There are different opinions regarding what cut-off values to use for spotting outliers. A simple operational guideline of $D_i > 1$ has been suggested (63) lack of fit.

DFBETA :- The influential observations for the individual regression coefficients are identified by DFBETA $s_{j(i)}, j = 0,1,2,\dots,p$, where each DFBETA $s_{j(i)}$is the standardized change inβ_j, when the i^{th} observation is deleted from the analysis. Thus,

DFBETA $s_{j(i)}=\frac{(\hat{\beta}_j-\hat{\beta}_{j(i)})}{s_i\sqrt{C_{jj}}}$, Where C_{jj} is the $(j+1)^{th}$ diagonal element from$(x'x)^{-1}$. DFBETA $s_{j(i)}$ measures the change in $\hat{\beta}_j$ in multiples of its standard error. DFBETA less than one indicate there is no potential out layer.

3.4.6. Logistic Regression Models for the Analysis of Correlated Data

Two approaches are commonly used to model correlated binary data to estimate the determinants between Regional characters and individual-level fertility outcomes in multilevel studies: a random effects model and a population average model. The random effects model mimics the usual normal errors linear mixed effects model, where parameter estimates are conditional on the subject or cluster. Under the population average model estimates are, in a sense, averaged over the clusters. The random effects model is referred to in the literature as a "cluster-specific" or "conditional" model. Often the clusters are specific subjects, but we will use the cluster-specific terminology as this term is a bit more general than "subject specific. "It describes the case of multiple observations on a single subject and single observations on related subjects; in our case it indicates that, single observations on related subjects. The cluster-specific (multilevel or hierarchical) binary outcome model is formulated in the manner of the normal errors linear mixed effects model (64). An alternative to the cluster-specific model is the population average model or "marginal" model. In which responses are marginalized over all other responses and Parameters characterize the marginal expectation (65). Both the cluster-specific and population average model may be fit to data containing subject-specific and cluster-level covariates (64).

3.4.7. Over View of Multilevel Model

Multilevel modeling is applied to logistic regression and other generalized linear models in the same way as in linear regression: the coefficients are grouped into batches and a probability distribution is assigned to each batch. In multilevel study, the structure of data in the population is hierarchical, and a sample from such a population can be viewed as a multistage sample. Because of cost, time and efficiency considerations, stratified multistage samples are the norm for sociological and demographic surveys. For such samples the clustering of the data is, in the phase of data analysis and data reporting, a nuisance which should be taken into consideration. However, these samples, while efficient for estimation of the descriptive population quantities, pose many challenges for model-based statistical inference (66).

Cluster sampling scheme often introduces multilevel correlation among the observations that can have implications for model parameter estimates. For multistage clustered samples, the dependence among observations often comes from several levels of the hierarchy. The problem of dependencies between individual observations also occurs in survey research, where the sample is not taken randomly but cluster sampling from geographical areas is used instead. In this case, the use of single-level statistical models is not reasonable. Hence, in order to draw appropriate inferences and conclusions from multistage stratified clustered survey data, we may require complicated modeling techniques like multilevel modeling. Multilevel models contain variables measured at different levels of hierarchy.

The 2016 EDHS data set have been used for this study was based on a multistage stratified cluster sampling (44). The appropriate approach to analyzing teenage fertility data from this survey is therefore based on nested sources of variability. Here the units at lower level are individuals (teenager aged 15−19) who are nested within units at higher level (regions). Due to this nested structure, the odds of teenage experiencing the outcome of interest are not independent, because women from the same cluster (region) may share common exposure to community characteristics.

3.4.7.1.Multilevel Logistic Regression Model

First consider a two-level model for binary outcomes with a single explanatory Variable. The basic data structure of the two-level regression is a collection of N groups (‘units at two levels’) with in group *j*, (j = 1, 2, … , *N*) random sample of n_j level-one units (individual or number of teenagers living in the region j). And the total sample size becomes M $=\sum_1^N n_j$. The extension to three or higher levels is straightforward. Let y_{ij} is the binary outcome variable, coded ‘1’ or ‘0’ associated teenagers nested within region *j* . Also let π_{ij} be the probability that the response variable equals 1, $\pi_{ij} = pr(y_{ij} = 1)$. Like the ordinary logistic regression, π_{ij} is modeled using the link function, logit.

The two-level logistic regression model can be given as:

$$\text{logit}(\pi_{ij}) = \log\left[\frac{\pi_{ij}}{1-\pi_{ij}}\right] = \beta_0 + \beta_1 x_{ij} + u_{0j} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. \ldots\ldots 3.08$$

Where,u_{0j}, is the random effect at level 2

We can equivalently, split model (3.08) into two models: one for level 1 and the other for level 2.

$$\text{logit}(\pi_{ij}) = \log\left[\frac{\pi_{ij}}{1-\pi_{ij}}\right] = \beta_{0j} + \beta_1 x_{ij} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [\text{Model: level 1}]\ \ \text{And}$$

$$\beta_{0j} = \beta_0 + u_{0j} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. [\text{Model level 2}]$$

The intercept β_{0j} consists of two terms: a fixed component β_0 and a group-specific component, random effect, u_{0j}.We assumes the u_{0j} follow a normal distribution with mean zero and variance σ^2_{u0} (67).

3.4.7.2.Test of Hetrogenity

For the proper application of multilevel analysis the first logical step is to test heterogeneity of proportions between groups. Here we present two commonly used test statistics that are used to check for heterogeneity (67). To test whether there are indeed systematic differences between the groups, the well-known Chi-Square test for contingency table can be used. In this case the Chi-Square test statistic is:-

$$x^2 = \sum_{j=1}^{N} n_j \left[\frac{\hat{\pi}_j - \hat{\pi}}{\hat{\pi}(1-\hat{\pi})}\right] \sim x^2_{(N-1)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.09$$

It can be tested a chi-square distribution with $N-1$ degrees of freedom. This chi-squared distribution is an approximation valid if the expected number of success $(n_j\,\pi_j)$ and of failures $n_j\,(1-\pi_j)$ in each group all are at least one while 80 percent of them are at least five (54)

3.4.7.3.Estimating Between and Within Groups Variance

Consider a population having two-levels. The basic data structure of two-level logistic regression is a collection of N groups (units at level-two (regions)) and within region j (j=1, 2… N) a random sample of n_j level-one (teenager) units. Then the true variance between the group dependent probabilities (67), i.e. the population values of $\hat{\tau}_j$ is given by:

$$\hat{\tau}_j = S^2_{between} - \frac{S^2_{within}}{\tilde{n}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.10$$

Where: $\tilde{n}$ is defined as: $\tilde{n} = \frac{1}{N-1}\left[M - \frac{\sum_{j=1}^{N} n_j^2}{M}\right]$, $S^2_{between} = \frac{\hat{\pi}(1-\hat{\pi})}{\tilde{n}(N-1)} x^2$ And

$S^2_{within} = \frac{1}{M-N}\sum_{j=1}^{N} n_j\left(1-\hat{\pi}_j\right)$, when x^2 is given in equation 3.17

3.4.7.4.The Empty Multilevel Logistic Regression Model

The empty two-level model for a dichotomous outcome variable refers to a population of groups (level-two units (regions)) and specifies the probability distribution for group-dependent probabilities π_j in $y_{ij} = \pi_j + \varepsilon_{ij}$ without taking further dependent variables in to account. We focus on the model that specifies the transformed probabilities $f(\pi_j)$ to have a normal distribution. The link function $f(\pi_j)$ is:

$$f(\pi_j) = \beta_0 + u_{0j} \ldots\ldots 3.11$$

Where, β_0 is the population average of the transformed probabilities and u_{0j} the random deviation from this average for group j. Thus, for the logit link function, the log-odds have a normal distribution in the population of groups, and it is given by:

$$\text{logit}(\pi_j) = \beta_0 + u_{0j} \ldots\ldots 3.12$$

This model does not include a separate parameter for the level-one variance. This is because the level-one residual variance of the dichotomous outcome variable follows directly from the success probability, as indicated by equation, $\text{var}(\varepsilon_{ij}) = \pi_j(1 - \pi_j)$. The probability corresponding to the average value β_0 denoted by π_0 is defined by $f(\pi_0) = \beta_0$. For the logit function the so-called logistic transformation of β_0, is defined by:

$$\boldsymbol{\pi_0} = \textbf{logit}(\boldsymbol{\beta_0}) = \frac{\exp(\boldsymbol{\beta_0})}{1+\exp(\boldsymbol{\beta_0})} \ldots\ldots 3.13$$

3.4.7.5.The Random Intercept Model

The random intercept model is the model in which the intercept is allowed to vary across regions after controlling for covariates of teenagers' Fertility. It means that the overall intercept ($\boldsymbol{\beta_0}$) is shared by all regions based on the region specific (random effects (u_{oj})) and is assumed to be normally distributed with mean zero and variance= $\boldsymbol{\delta^2_{u0}}$. In the random intercept model, the intercept is the only random effect meaning that the groups differ with respect to the average value of the response variable, but the relation between response and explanatory variables cannot differ between groups (regions). The random intercept model expresses the legit of success probability as a sum of a linear function of the explanatory variables. Thus it is given as:

$$\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_{0j} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \cdots + \beta_k x_{kij}$$

$$\beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + u_{0j} \dots \dots \dots \dots 3.14$$

Where $= \beta_{0j}$is assumed to vary randomly and is given by the sum of an average intercept and group-dependent deviations. That is $\beta_{0j} = \beta_0 + u_{0j}$. The first part incorporating the regression coefficients $\beta_0 + \sum_{h=1}^{k} \beta_h x_{hij}$ is the fixed part of the model, because the coefficients are fixed. The remaining part u_{0j} is called the random part of the model. It is assumed to be the residual, u_{0j} are mutually independent and normally distributed with mean zero and variance σ_0^2 (67). Thus by solving for π_{ij} we have

$$\pi_{ij} = \frac{\exp\left(\beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + u_{0j}\right)}{1 + \exp\left(\beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + u_{0j}\right)} \dots \dots \dots \dots 3.15$$

Thus, a unit difference between the x_h values of two individuals in the same group is associated with a difference of β_h in their log-odds, or equivalently, a ratio of $\exp(\beta_h)$ in their odds (67).

3.4.7.6.The Random Slope Model

The multilevel modeling strategy accommodates the hierarchical nature of the data and corrects the estimated standard errors to allow for clustering of observations within units (67). A significant random effect may represent factors influencing the outcome variable that cannot be quantified in a large-scale social survey. A random effects model thus provides a mechanism for estimating the degree of correlation in the outcome that exists at the region level, while also controlling a range of all indicators may potentially influence the outcome.

The intercepts $= \beta_{0j}$ as well as the regression coefficients, or slopes, β_{1j} are group (region) dependent. These group dependent coefficients can be split into an average coefficient and the group dependent deviation:

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$\beta_{1j} = \beta_1 + u_{1j}$$ Then is given by

$$\text{logit}\left(\pi_{ij}\right) = \left[\beta_0 + u_{0j}\right] + \left[\beta_1 + u_{1j}\right]x_{1ij} = \beta_0 + \beta_1 x_{1ij} + u_{0j} + u_{1j}x_{1ij} \dots \dots \dots \dots 3.16$$

Now we have two parts which is known as fixed and random parts and we have two random effects at group level, the random intercept u_{0j} and the random slope u_{1j}.It assumed that both random effects have mean zero and the variances are denoted by:, δ_0^2, δ_1^2 and their covariance is δ_{01}Where β_0 is the average intercept of the response variable, β_1 is fixed regression coefficient given explanatory variable X_1, u_0 is the random coefficient in the model , $u_0 + u_1X_{1ij}$ the random part of the model can be considered as interaction by group and predictors.

The two random effects that characterized group (region) u_{0j} and u_{1j} are correlated. Further, it is assumed that, for different groups, the pairs of random (u_{0j}, u_{1j}) effects are independent and identically distributed. Thus, the variances and covariance of the level-two random effects are (u_{0j}, u_{1j}) denoted by:

$$\mathrm{var}(u_{0j}) = \delta_{00} = \delta_0^2$$
$$\mathrm{var}(u_{1j}) = \delta_{11} = \delta_1^2$$
$$\mathrm{cov}(u_{0j}, u_{1j}) = \delta_{01}$$

The model for a single explanatory variable shown above can be extended by adding more variables that have random effects on response variable. Suppose that k level one explanatory variables are given as $X_{1,}X_2, \ldots\ldots, X_k$ and consider the model where all predictor variables have varying slopes and random intercept is given by:

$$\mathrm{logit}[\pi_{ij}] = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \cdots + \beta_{kj}X_{kij} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots .3.17$$

$\beta_{0j} = \beta_0 + u_{0j}$ and $\beta_{hj} = \beta_h + u_{hj\,,}$where , $h = 1,2, \ldots.., k$,

Since $\beta_{0j} = \beta_0 + u_{0j}$ and $\beta_{hj} = \beta_h + u_{hj\,,}$where , $h = 1,2, \ldots.., k$, We have

$$\mathrm{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \sum_{h=1}^{k} \beta_h X_{ij} + u_{0j} + \sum_{h=1}^{k} \beta_{hj} X_{ij} \ldots\ldots\ldots\ldots\ldots\ldots .\ldots .\ldots\ldots\ldots 3.18$$

The first part of equation (3.18) is called the fixed part of the model and the second part is called the random part of the model. The random variables or effects $u_{0j}, u_{1j}, \ldots\ldots u_{kj}$ are assumed to be correlated within groups and independent between groups. So the components of the vector

$(u_{0j}, u_{1j}, \ldots \ldots u_{kj})$ are independently distributed as a multivariate normal distribution with zero mean vector and variance and co-variance Ω matrix given by:

$$\Omega = \begin{bmatrix} \delta_0^2 & \delta_{01} & \cdots & \cdots & \delta_{0k} \\ \delta_{10} & \delta_1^2 & \vdots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \delta_{k-10} & \vdots & \vdots & \delta_{k-1}^2 & \vdots \\ \delta_{ko} & \delta_{0k} & \cdots & \cdots & \delta_k^2 \end{bmatrix},$$

The diagonal elements are variances and non-diagonal elements are co-variances.

3.4.7.7.Estimation Methods for Multilevel Logistic Regression Model

Like the methods for ordinary logistic regression, Parameter estimation for multilevel logistic model is not straight forward. Two classes of estimation method are most commonly used in standard software packages. The first is to avoid the difficulty of evaluating the integral which is a non-linear function of the parameters, by "linearizing" the model using a Taylor series approximation. These "linearized" models are then estimated using methods from linear mixed models. As the likelihood function is not actually used in the estimation, such procedures are referred to as "quasi-likelihood" or "pseudo-likelihood" estimation. The second approach is to evaluate the integral in the likelihood function using numerical integration techniques namely quadrature (Gauss–Hermite quadrature). The basic idea of quadrature is to replace the integral over the random effects normal distribution with a sum. Software packages offer numerous options for the approximation and parameter estimation algorithms (63). For this study, the researcher has used the second approach.

3.4.7.8.Model Selection and Information Criterion

In selecting a model, you should not think that you have found the "correct" one. Any model is a simplification of reality. However, a simple model that fits adequately has the advantages of model parsimony. If a model has relatively little bias, describing reality well, it provides good estimates of outcome probabilities and of odds ratios that describe effects of the predictors (54).

Akaike Information Criterion (AIC); the AIC penalizes a model for having many parameters. Even though a simple model is farther than a more complex model from the true relationship, for a sample the simple model may provide better estimates of the true expected values. The model with minimal AIC tries to find an optimal compromise between models fit and model complexity

$$AIC = \underbrace{-2\log[likelihood]}_{measure-of-goodness-of-fit} + \underbrace{2 * p(parameternumember)}_{Measures-of-Model-complexity} \qquad 3.19$$

(56). **Bayesian Information Criterion (BIC):** A Bayesian argument motivates the Bayesian information criterion (58).

$$\mathrm{BIC} = [\mathrm{G}^2 - \ln(\mathrm{n})(\mathrm{df})] = -2\ln(\mathrm{likelihood}) + \mathrm{k} * \ln(\mathrm{n}) \qquad 3.20$$

an alternative to AIC. It takes sample size into account. Compared to AIC, BIC gravitates less quickly toward more complex models as n increases. The model with minimal BIC tries to find an optimal compromise between models fit and model complexity. (60)

3.4.7.9.Population Average Model

Population average model is an alternative to cluster-specific or multilevel model in which the parameters characterize the marginal expectation and responses are marginalized over all other responses. Under this model we average probabilities of the outcome, in a sense, over the statistical distribution of the random effect and assume that this process yields the logit

$$\mathrm{Logit}(\pi) = \mathrm{Log}\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \mathrm{X}'_{ij}\beta_{PA} \qquad 3.21$$

Probabilities based on the logit in equation (3.21) represent the proportion of subjects in the population with outcome present among subjects with covariates, X_{ij}. Note that we have not specified the statistical distribution of the random effects, only that the population proportions have logit function given by equation (3.21). In population average model the coefficient describes the effect of the covariate in broad groups of subjects rather than in individual subjects (64). The coefficient estimates returned by the generalized estimating equations (GEE) typically used to estimate population average models (sometimes called marginal models) describe changes in the population mean given changes in covariates, while accounting for within-regions non-independence of observations when deriving the variability estimates of these coefficients(68).

3.4.6.1.1. Analysis Method for Population Average Model (GEE)

In the population average model, estimation is based on generalized estimating equations (GEE). The GEE approach uses a set of equations that look like weighted versions of the likelihood equations (64). Generalized Estimating Equation itself is not model but it is an analysis method which fit Marginal (population Average) Model (70). A key feature of this method is the option to estimate the correlation structure from the data without having to assume it follows a pre-specified structure (70). GEE is an extension of GLMs to account for correlated responses. It doesn't require distributional assumption for the observations, only regression model for the mean response. GEE Provide consistent regression coefficient estimates even if the correlation structure is miss-specified allows estimation of regression and association parameters without specifying the entire likelihood (65).

In most applications we are only interested in estimating the regression coefficients and need to account for correlation in responses to obtain correct estimates of the standard errors of the estimated coefficients. For this reason the choice of correlation structure to use in the GEE as the "working correlation." The idea is that one chooses a correlation structure for estimation that seems plausible for the setting and then this structure is used in adjusting the estimator of the variance. One of the advantages of the GEE approach is the "robustness" of the estimates to choice of correlation structure. Thus we use the GEE method for population average models with exchangeable correlation as the working correlation unless the nature of the data clearly suggests another choice.

We need some additional notation to fully describe the application of GEE to the population average model. We denote the logistic probability obtained from the logit in equation (3.08) as

$$\pi_{PA(X_{ij})} = \left[\frac{\exp\left(\beta_0 + X'_{ij}\beta_{PA}\right)}{1 - \exp\left(\beta_0 + X'_{ij}\beta_{PA}\right)}\right] \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \; 3.22$$

Where $\boldsymbol{\beta_0 + X'_{ij}\beta_{PA}}$ is the usual linear expression of the logit consisting of the set of predictors and the corresponding parameters estimated under the population average model. We use two matrices to describe the within-cluster covariance of the correlated observations of the outcome variable. The first is an $n_i \times n_i$ (recall from the previous section that this is the number of observations in the cluster) diagonal matrix containing the variances under the model in equation (3.22) denoted

$$A_i = \text{diag}\left[\pi_{PA(X_{ij})} \times \left(1 - \pi_{PA(X_{ij})}\right)\right] \ldots\ldots 3.23$$

And the second is the $n_i \times n_i$ exchangeable correlation matrix denoted

$$R_i(\rho) = \begin{bmatrix} 1 & \rho & \dots & \rho & \rho \\ \rho & 1 & \cdots & \vdots & \vdots \\ \rho & \rho & \ddots & \rho & \rho \\ \vdots & \vdots & \dots & 1 & \rho \\ \rho & \rho & \cdots & \rho & 1 \end{bmatrix} \ldots\ldots 3.24$$

Using the fact that the correlation is defined as the covariance divided by the product of the standard deviations it follows that the covariance matrix in the i^{th} cluster is

$$V_i = \sqrt{A_i} R_i(\rho) \sqrt{A_i} \ldots\ldots 3.25$$

Where, $\sqrt{A_i}$ is the diagonal matrix whose elements are the square roots of the elements in the matrix in equation (3.25). The contribution to the estimating equations for the i^{th} cluster is $\mathbf{D_i' V_i^{-1} S_i}$ where, $D_i' = X_i' A_i$, X_i is the $n_i \times (p+1)$ Matrix of covariate values and S_i the vector with j^{th} element the residual, $S_{ij} = y_{ij} - \boldsymbol{\pi}_{\mathbf{PA(X_{ij})}}$. The full set of estimating equation is

$$\sum_{i=1}^{m} D_i' V_i^{-1} S_i = 0 \ldots\ldots 3.26$$

And its solution is denoted as $\hat{\beta}_{PA}$ Implicit in the solution of these equations is an estimator of the correlation parameter, ρ. typically this is based on the average correlation among within-cluster empirical residuals and as such it is also adjusted with each iterative change in the solution for, $\hat{\beta}_{PA}$. They derive, as an estimator of the covariance matrix, the estimator that is often referred to as the information sandwich estimator. The "bread" of the sandwich is based on the observed information matrix under the assumption of exchangeable correlation. The "bread" for the i^{th} cluster is

$$\boldsymbol{\beta_i} = \mathbf{D_i' V_i^{-1} D_i}$$
$$= \mathbf{X_i' A_i} \left(\sqrt{\mathbf{A_i}} \mathbf{R_i}(\boldsymbol{\rho}) \sqrt{\mathbf{A_i}}\right)^{-1} \mathbf{A_i X_i'}$$

The "meat" of the sandwich is an information matrix that uses empirical residuals to estimate the within-cluster covariance matrix. The "meat" for the i^{th} cluster is

$$\mathbf{M_i = D_i' V_i^{-1} C_i V_i^{-1} D_i}$$

$$\mathbf{X_i' A_i \left(\sqrt{A_i} R_i(\rho) \sqrt{A_i}\right)^{-1} C_i \left(\sqrt{A_i} R_i(\rho) \sqrt{A_i}\right)^{-1} A_i X_i'} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.27$$

Where, $\mathbf{C_i}$**,** is the outer product of the empirical residuals. Specifically, the jk^{th} element of this $n_i \times n_i$ matrix is

$$\mathbf{C_{ij} = \left[y_{ij} - \pi_{PA(X_{ij})}\right] \times \left[y_{ij} - \pi_{PA(X_{ij})}\right].}$$

The equation for the estimator is obtained by evaluating all expressions at the estimator $\hat{\beta}_{PA}$ and the respective values of the covariates, namely

$$\mathbf{\widehat{Cov}\left(\hat{\beta}_{PA}\right) = \left(\sum_{i=1}^{m} \hat{\beta}_i\right)^{-1} \times \left(\sum_{i=1}^{m} \hat{M}_i\right) \times \left(\sum_{i=1}^{m} \hat{\beta}_i\right)^{-1}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.28$$

We note that some packages may offer the user the choice of using the information sandwich estimator, also called the robust estimator, in equation (3.28) or one based only on the observed information matrix for the specified correlation structure, the “bread” $\boldsymbol{\beta_i}$. One can use the estimated coefficients and estimated standard errors to estimate odds ratios and to perform tests for individual coefficients. Joint hypotheses must be tested using multivariable Wald tests because the GEE approach is not based on likelihood theory (64, 70).

CHAPTER FOUR

4. Results and Discussions

4.1. Background Information

The total of 3498 teenagers from nine regional states and two administrative cities in Ethiopia were eligible for this study. Among these eligible teenagers', 359 (10.263%) teenagers have been born child at the time of interview.

The rate is calculated as, fertile teenage divided by total teenagers in each category and the proportion is the ratio of fertile teenage to non-fertile teenage in each category. The rate of teenagers' fertility for each indicator is shown in Table (4.1)**.** From the total number of teenage fertility status in the region groups, fertility rates are: 8.27%, 20.68%, 7.04%, 13.73%, 15.36%, 10.55%, 6.91%, 15.57%, 14.21% ,1.85% and 7.17% for teenagers living in; Tigray, Affar, Amhara, Oromia, Somali, Benshangul, SNNPR, Gambella, Harari, Addis Ababa and Dire-Dawa respectively. As shown in the table; from the total teenage fertility status in the residence group fertility rates 4.25% of urban teenager and 13.59% of rural teenager were fertile at the time of interview. And fertility rates for religion group 5.25% of teenagers following Orthodox, 14.84% of teenagers following Muslim, and 11.01% of teenagers following other religions were fertile at the time of interview. And also fertility rates in Teenagers education level; 22.46%, 9.47%, 4.64% and 1.94% for Teenagers with; no education, Primary education, secondary education and higher education respectively.

Also according to table(4.1) Fertility rates in teenagers' wealth Index; 17.61%, 12.13%, and 4.93% of teenagers with wealth index; Poor, Middle, and Rich respectively were fertile, Fertility rates in a teenagers Knowledge of ovulatory cycle; 10.20%, 18.14%, 8.64%, 10.23%, 9.94%, and 6.41% of teenagers having knowledge of ovulatory cycles ; during her period, After period ended, Middle of the cycle, Before period begins, At any time, and Don't know, were fertile respectively, Fertility rates in teenagers using status of contraceptive methods; 8.28%, 9.94%, and 6.41% of teenagers using; no method, Traditional Method and modern method were fertile respectively, Fertility rate in teenagers' Working status ;10.71% of teenagers' with no work were fertile and 8.91% of teenagers with work were fertile, Fertility rates in teenagers' Exposure to any mass media; 14.88% of teenagers having no exposure to any mass media were fertile and 6.58% of teenagers exposed to any mass media were fertile. Similarly one can describe Ethnicity in the same way.

Table 4.1 *The rate of teenagers' fertility for each indicator*

Variables		**FertilityStatus**				
		Not Fertile	**Fertile**	**Total**	**Rate**	**Proportion**
Region	Tigray	388	35	423	8.27%	9.02%
	Afar	211	55	266	20.68%	26.07%
	Amhara	330	25	355	7.04%	7.58%
	Oromia	358	57	415	13.73%	15.92%
	Somali	270	49	319	15.36%	18.15%
	Benishangul	212	25	237	10.55%	11.79%
	SNNPR	364	27	391	6.91%	7.42%
	Gambela	179	33	212	15.57%	18.44%
	Harari	157	26	183	14.21%	16.56%
	Addis Adaba	424	8	432	1.85%	1.89%
	Dire Dawa	246	19	265	7.17%	7.72%
Residence	Urban	1193	53	1246	4.25%	4.44%
	Rural	1946	306	2252	13.59%	15.72%
Religion	Orthodox	1335	74	1409	5.25%	5.54%
	Muslim	1222	213	1435	14.84%	17.43%
	Others	582	72	654	11.01%	12.37%
Edu.Level	No education	442	128	570	22.46%	28.96%
	Primary	1835	192	2027	9.47%	10.46%
	Secondary	761	37	798	4.64%	4.86%
	Higher	101	2	103	1.94%	1.98%
WealthEndex	poor	992	212	1204	17.61%	21.37%
	Middle	413	57	470	12.13%	13.80%
	Rich	1734	90	1824	4.93%	5.19%

KnowlOvuCycle	During her period	132	15	147	10.20%	11.36%
	After period ended	537	119	656	18.14%	22.16%
	Middle of the cycle	592	56	648	8.64%	9.46%
	Before period begins	237	27	264	10.23%	11.39%
	At any time	707	78	785	9.94%	11.03%
	Don't know	934	64	998	6.41%	6.85%
Cont.Method	No method	3015	272	3287	8.28%	9.02%
	Traditional method	2	2	4	50.00%	100.00%
	Modern method	122	85	207	41.06%	69.67%
WorkingStatus	No	2342	281	2623	10.71%	12.00%
	Yes	797	78	875	8.91%	9.79%
Sex of household head	M	2123	250	2373	10.54%	11.78%
	F	1016	109	1125	9.69%	10.73%
ExpossuretoanymassMedia	No	1321	231	1552	14.88%	17.49%
	yes	1818	128	1946	6.58%	7.04%
Ethnicity	Amhara	714	41	755	5.43%	5.74%
	Oromo	731	98	829	11.82%	13.41%
	Tigrie	435	36	471	7.64%	8.28%
	Affar	183	54	237	22.78%	29.51%
	Somalie	299	49	348	14.08%	16.39%
	Guragie	134	3	137	2.19%	2.24%
	Sidama	68	7	75	9.33%	10.29%
	Nuwer	43	9	52	17.31%	20.93%
	WDG	120	10	130	7.69%	8.33%
	Berta	51	11	62	17.74%	21.57%
	Kefficho	42	1	43	2.33%	2.38%
	Gumuz	25	4	29	13.79%	16.00%
	Hadiya	44	3	47	6.38%	6.82%
	Silte	34	1	35	2.86%	2.94%
	Anyiwak	43	15	58	25.86%	34.88%
	Others	173	17	190	8.95%	9.83%

4.2. Test of Association between Teenagers' Fertility and its Indicators

From Chi-square Test of association between teenagers' fertility and categorical explanatory variables such as: Regions, Residence, Religion, teenagers' educational level, household wealth index, Knowledge of Ovulatory Cycle, Contraceptive Method used status, Ethnicity, and Exposure to any mass media are found significant at 5% level of significance indicating that, they have significant association with Teenagers' fertility. While teenagers' working status and sex of house hold are found insignificant at 5% significance level, suggesting that no significant associations with teenagers' fertility (**see Table 4.2** in **Appendix A**).

4.3. Logistic Regression Analysis

4.3.1. Results of Logistic Regression Analysis

When, we come to multiple logistic regression analysis the variable residence is not significant but it disturbs the variable region. When we analyze multiple logistic regression including **residence** only two region i.e (Gambella and Harari) found to be significant and the residence itself is not significant, and when we analyze Multiple logistic regression without residence five regions i.e (Affar,Oromia Somali,Gambella and Harari) are significant at 5% level of significance. So it is better to drop the variable Residence and the results of logistic regression analysis represented in Table (4.3).

This table provides the regression coefficients (**β**) and all-important Odds Ratio (**Exp (β)**) for each variable category. The result shows that Region ;Affar, Oromia, Somali, Gambella and Harari are found significant and all categories of Religion, Education level, Wealth Index, Knowledge of Ovulatory Cycle, Contraceptive Method and Exposure to any mass media are significant at 5% level of significance. And also the remaining categories of region (Tigray, Amhara, Benshangul,SNNPR and Dire-Dawa) are insignificant implying they have no significant difference with reference. The coefficients for all significant categories of region are significant and positive implying that has a positive relationship with log-odds of teenagers' fertility and each significant region is associated with increased odds of teenagers' fertility.

Logistic regression analysis revealed that the odds of Teenagers' living in Afar (OR=2.602, CI=[1.075,6.30]),Oromia(OR=2.74,CI=[1.185,6.34]),Somali(OR=2.75,CI=[1.135,6.645]),Gamb ella (OR=3.767,CI=[1.494,8.998]) and Harare(OR=4.44,CI=[1.802,10.93]) times more likely than those of teenagers' living in Addis Ababa to be fertile before nineteen years of age respectively. While the odds of teenagers' living in the remaining Five regions being fertile before nineteen years of age are not significant, indicating that teenagers fertility in these regions are almost similar to Addis Ababa.

The effect of Religion is significant and positive indicating that has a positive relationship with log-odds of teenagers' fertility and teenagers' following the Religion category others and Muslim are associated with increased odds of teenagers' fertility. The result implies that the Odds of teenagers' in the religion group others (Catholic, Protestant, Traditional and Others) is (OR=2.61, CI [1.55, 4.37]) times more likely than that of Orthodox and Odds of teenagers' in the religion Muslim are (OR=2.565, CI= [1.603, 4.104]) times more likely than that of teenagers' in Orthodox.

The effect of Education Level is highly significant and negative, implying that has negative relationship with log-odds of teenagers' fertility and teenagers' having primary, secondary and Higher education are associated with decreased odds of teenagers' fertility. The result shows that Odds of teenagers with primary education is (OR=0.501, CI= [0.3696, 0.6783]) times less likely than that of teenagers' with no education, Odds of teenagers' with secondary education is (OR=0.360, CI= [0.224, 0.579]) times less likely than that of teenagers with no education and Odds of teenagers with higher education is (OR=0.196, CI=[0.0446, 0.861]) times less likely than that of Teenagers with no education.

The outcome of wealth index is highly significant and positive, indicating that has a positive relationship with log-odds of teenagers' fertility and teenagers' of having wealth Index poor and Middle are associated with Increased Odds of Teenagers' fertility. The outcome shows that the Odds of teenagers with wealth index poor is (OR=2.759, CI [1.932, 3.940]) times more likely than that of teenagers with wealth Index Rich and Odds of teenagers with wealth Index Middle is (OR=2.289, CI=[1.521, 3.4456]) times more likely than that of teenagers with wealth index Rich.

The outcome of knowledge of Ovulatory cycle is positive and highly significant suggesting that, it has positive relationship with log-odds of teenagers' fertility and teenagers' having knowledge of Ovulatory cycle; during her period, After period ended, Middle of the cycle, Before period begins and At any time are associated with Increased Odds of teenagers' fertility. The result reveals that the Odds of teenagers with knowledge of Ovulatory cycle at, During her period is (OR=2.633, CI=[1.352, 5.125]) times more likely than that of don't know, Odds of teenagers with knowledge of Ovulatory cycle After period ended is (OR=4.544, CI=[3.165, 6.523]) times more likely than that of Don't know, Odds of teenagers' with knowledge of Ovulatory cycle at Middle of the cycle is (OR=2.674, CI=[1.752, 4.081]) times more likely than that of reference, Odds of teenagers' with knowledge of Ovulatory cycle before period begins is(OR=3.044, CI=[1.769, 5.237]) times more likely than that of the reference and Odds of teenagers with knowledge of Ovulatory cycle at any time is(OR=2.022, CI=[1,389, 2.943]) times more likely than that of the reference.

The Modern contraceptive Method used teenagers' have Odds (OR=14.644, CI= [9.976, 21.]) times more likely than that of teenagers' used no Method and Teenagers' used traditional method. And also the table shows that Odds of teenagers' exposed to any mass media is (OR=0.743, CI=[0.555, 0.9955]) times less likely than that of Teenagers' Not exposed to any Mass media. The constant term is highly significant which indicates that the odds of teenagers' fertility is (OR=0.009875, CI=[0.003978, 0.0245]) by controlling coefficients for other covariates zero.

Table4.3 *Results of logistic regression analysis*

Log likelihood = -903.14368

FertilityStatus |	**β.**	**Exp(β)**	**Std. Err.**	**z**	**P>|z|**	**[95% Conf.**	**Interval]**
Region (Refer. Addis Ababa)							
Tigray |	.8333965	2.301121	1.070872	1.79	0.073	[.9243067	5.72879]
Affar |	.9563657	2.602222	1.173923	2.12	0.034	[1.074849	6.300009]
Amhara |	-.0424294	.9584581	.4584621	-0.09	0.929	[.3753319	2.447546]
Oromia |	1.008851	2.742448	1.173524	2.36	0.018	[1.185496	6.344201]
Somali |	1.010381	2.746648	1.238324	2.24	0.025	[1.135114	6.646094]

Benishangul	.6719023	1.957958	.9048023	1.45	0.146	[.7914988	4.84347]
SNNPR	.3411415	1.406552	.6559913	0.73	0.464	[.5638588	3.508661]
Gambela	1.299103	3.666006	1.679499	2.84	0.005	[1.4936	8.998126]
Harari	1.489932	4.436795	2.040202	3.24	0.001	[1.801587	10.92656]
Dire Dawa	.7336432	2.082654	.9719429	1.57	0.116	[.8344004	5.198283]
Religion (Refer. Orthodox)							
Others	.957929	2.606293	.6885797	3.63	0.000	[1.552876	4.374312]
Muslim	.9420649	2.565273	.6149734	3.93	0.000	[1.603522	4.103858]
Edu_Level (Refer. No Education)							
Primary	-.6917661	.500691	.0775568	-4.47	0.000	[.3695891	.678298]
Secondary	-1.021477	.3600627	.0873063	-4.21	0.000	[.2238633	.5791265]
Higher	-1.629962	.1959371	.1480344	-2.16	0.031	[.0445673	.8614233]
WealthEndex (Refer. Rich)							
poor	1.014791	2.758787	.5017176	5.58	0.000	[1.9316	3.940207]
Middle	.8281833	2.289156	.4775991	3.97	0.000	[1.520844	3.445611]
KnowlOvuCycle (Refer. Don't Know)							
During her period	.968096	2.632927	.8948654	2.85	0.004	[1.352499	5.125553]
After period ended	1.513777	4.543859	.8382266	8.21	0.000	[3.16519	6.523038]
Middle of the cycle	.9835009	2.673801	.5768623	4.56	0.000	[1.751804	4.081055]
Before period begins	1.113163	3.043971	.8427379	4.02	0.000	[1.769217	5.237207]
At any time	.7040337	2.021892	.3873429	3.67	0.000	[1.388959	2.943245]

Cont_Method (Refer. No Meth& Trad.Meth)							
Modern Method |	2.684035	14.64406	2.867378	13.71	0.000	[9.976818	21.49467]
Media |(Refer. No)							
yes |	-.2966954	.7432704	.1108166	-1.99	0.047	[.5549303	.9955319]
_cons |	-4.617777	.0098747	.0045806	-9.95	0.000	[.0039781	.0245118]

Akaike's information criterion and Bayesian information criterion

Model |	Obs	ll(null)	ll(model)	df	AIC	BIC
. |	3,498	.	-903.1437	25	1856.287	2010.286

4.3.2. Assessment of Goodness Fit of Logistic Regression Analysis

Table(4.4) represents the assessment of Goodness of fit. The deviance statistic is the difference between two times likelihood of baseline model and full model. The deviance for empty model is **2314.1** and the deviance of full model is **1799.4** which implies that the full model fits the data well. As shown in the Table the likelihood of full model is greater than that of the likelihood of empty model which is significant with p-value=0.000* implying an evidence against null hypothesis that, there is no indicators that affect teenagers' fertility. And also the table shows that the AIC for full model is less than that of the AIC of empty model this also implies that the full model is good to fit the given data.

Table 4.4 *Represents the assessment of Goodness of fit*

Model	Log-likelihood	P-Value	Deviance	AIC
Empty	-1157.2213		2314.4	2316.4
Full	-903.14658	0.000	1799.4	1853.4

Omnibus Tests of Model Coefficients

As shown in the Table(4.5) blow the Chi-square =152.131 is significant at 5% level of significance with p-value less than 0.05. It indicates that, considering all indicators of Teenagers' Fertility together, the overall logistic regression model is; significantly predict the teenagers' fertility.

Table(4.5) *Omnibus Tests of Model Coefficients*

	Chi-square	**df**	**Sig.**
Step	4.836	1	0.000
Block	524.408	30	0.000
Model	524.408	30	0.000

Model Summary of Logistic Regression Analysis.

The "Pseudo" R^2 estimates in the Table(4.6) indicate that, approximately 13.8 % or 28.5% of the variance in teenagers' in Ethiopia whether fertile or not fertile can be predicted from the linear combination of all variables supposed to be an indicator of teenagers' fertility. The cox and Snell R^2(13.8%) usually an underestimate. The Nagelkerkerke R^2 is estimated at 28.5% implying that all determinants are useful in predicting teenage fertility in Ethiopia. Include those variables having significant association with teenage fertility by using logistic regression analysis for EDHS 2016 data is appropriate.

Table (4.6) Model Summary of logistic Regression

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
	1790.035	0.138	0.285

Hosmer and Lemeshow Test of Goodness of fit

Table (4.7) blow shows that the Hosmer and Lemeshow test statistic; it indicates the fit of predicted and actual value of teenagers fertility based on the data. Since the Hosmer and Lemeshow goodness-of-fit test statistic is not significant at significance level of 5% which implies that fail to reject the null hypothesis which states that there is no difference between observed data and Model predicted values, indicating that the fitted model is good to fit the data in given level of significance.

Table(4.7) *Hosmer and Lemeshow Test*

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	7.825	8	0.451

Outliers and Influential Diagnostics

Cook’s distance is a measure used to identify influential cases. It is a measure of how much the residual of all cases would change if a particular case is excluded from the computation of the regression coefficients. Guide line for Cook's influence statistics is; greater than or less than one implying potential outlier or no potential outlier respectively.

The maximum value of Cook’s Distance, and leverage for each indicators of teenagers’ fertility is less than one. This result implies that among the indicators of teenagers’ fertility, there is no potential influential observation (see Table (4.8) in Appendix A). And also the table indicates that DFbeta for variable ethnicity i.e DFbeta for Gumuz and Ayniwak is greater than one; implying the two variables accounted as influential observations, then as a result the variable ethnicity disturbs over all variables in the analysis. And so, it is better to drop the variable ethnicity from the analysis. And also according to univariable logistic regression analysis the variables; Teenagers’ working status and sex of house hold are found not significant at 5% level of significance which is the same with part in the test of association. So it is better to drop the variables Teenagers’ working Status and Sex of house hold. Therefore the multiple, logistic regression is left with the remaining variables.

4.4. Results of Multilevel Logistic Regression Analysis

In multilevel logistic regression analysis two-levels are used: - Teenagers as the first level unit and regions as the second-level unit. For Multilevel Logistic Regression analysis there are three Quadrature's of Integration Methods, these are: - Mean and Variance adaptive guess Hermit Quadrature, Mode and Curvature adaptive guess Hermit Quadrature and Non-adaptive gauss hermit quadrature. The results for all this three Integration Methods are almost the same; so, for this study, the researcher has been used Mean and Variance adaptive gauss Hermit quadrature (mvaghermite) Method of Integration.

4.4.1. Test of Heterogeneity between Regions

Before attempting to multilevel analysis, one has to test the heterogeneity of teenagers' Fertility among regional states of Ethiopia. As shown in Table (4.9) the Pearson Chi-square = 101.880 which is highly significant with P-value = 0.000, indicating strong indication of heterogeneity for teenagers' fertility across regional states of Ethiopia. Therefore, attempting multilevel logistic analysis is better to fit the given data set.

Table (4.9) *Tests of Heterogeneity*

Chi-Square Tests			
Statistics	Value	df	p-value
Pearson Chi-Square	101.880[a]	10	0.000
N of Valid Cases	3498		
(*significant at 5% level)			

4.4.2. Multilevel Empty Logistic Regression analysis

The single level and multilevel empty models are differs only with respect to the variance component and it is tested by the deviance of likelihood-ratio based on Chi-square as shown in Table (4.10) The probability of Chi-square = 72.02 with p-value =0.000** is highly significant in 5% level of significance. Therefore, multilevel empty model is found to be significant, suggesting that there is a significant region effects on teenagers' fertility variation in Ethiopia in turn it suggests that multilevel analysis is better than single level analysis.

In empty multilevel model the overall log-odds of teenagers' fertility is estimated by -2.224 and between region variance of teenagers' fertility is estimated by $\delta_{u0}^2 = .372$ which is highly significant at 5% level of significance representing the variations of teenagers' fertility among regional states of Ethiopia. The intra-regional correlation coefficient of regional states of Ethiopia for empty multilevel model is estimated at 0 .1016 implying that 10.16% of the variance in Teenagers' Fertility could be attributed to differences across regions.

Test of significance of random intercept is Ho: $\delta_0^2 = 0$ versus $\delta_0^2 \neq 0$, in table (4.10), the 95% confidence interval for the estimate of random intercept is not negative and different from zero. This indicates that there is a significant Heterogeneity (variance difference) between Regional states of Ethiopia. And also it tells as random intercept is significantly different from zero.)

Table (4.10) Multilevel empty logistic regression analysis.

Log likelihood = -1121.2121 Prob > chi2 = .

FertilityS~s	Coef(β).	Std. Err.	z	P>|z|	[95% Conf. Interval]	
_cons	-2.224467	.1940935	-11.46	0.000	[-2.604883	-1.84405]

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
Region: Identity				
var(_cons)= $\boldsymbol{\delta_{u0}^2}$	.3722353	.1858914	[.1398731	.9906058]
ICC (rho)	.1016461	.0456019	[.0407825	.2314273]

Likelihood-ratio test of rho=0: chibar2(01) = 72.02 Prob >= chibar2 = 0.000

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	3498	.	-1121.212	2	2246.424	2258.744

4.4.3. Results of Random Intercept Logistic Regression Analysis

The random intercept model has larger log-likelihood and smaller IC suggesting that the random intercept model with all indicators of teenagers' Fertility is better to fit the data than Empty Multilevel model. The probability of Chi-square =8.35 is significant with p-value = 0.0019 implying that after controlling all indicators of Teenagers' Fertility, the intercept varied across the region with variance of 0.111, telling that there is significant variation of teenagers' Fertility among Regional states of Ethiopia. And also the variance is decreased from $\boldsymbol{\delta_{u0}^2}$ 0=0.372 in Empty Multilevel model to $\boldsymbol{\delta_{u0}^2} = 0.111$ in to random intercept model indicating that teenager fertility indicators are accounted for decreasing variations through regional states of Ethiopia.

Table (4.11) *Results of random intercept model*

Log likelihood = -916.3357 Prob > chi2 = 0.0000

Fertility Status	Coef(β).	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]	
Religion								
Others	1.079449	2.943057	.2418177	4.46	0.000	[.6054948	1.553403]
Muslim	1.10813	3.028689	.2163134	5.12	0.000	[.6841635	1.532096]
Edu_Level								
Primary	-.7141343	.4896158	.1532446	-4.66	0.000	[-1.014488	-.4137803]
Secondary	-1.02962	.3571425	.2404933	-4.28	0.000	[-1.500979	-.5582621]
Higher	-1.679848	.1864024	.7552079	-2.22	0.026	[-3.160028	-.1996672]
WealthEndex								
poor	1.036376	2.818981	.1764254	5.87	0.000	[.6905882	1.382163]
Middle	.8100004	2.247909	.204769	3.96	0.000	[.4086606	1.21134]
KnowlOvuCycle								

During her period	.9549164	2.598453	.3388741	2.82	0.005	[.2907355	1.619097]
After period ended	1.488304	4.429575	.1834475	8.11	0.000	[1.128753	1.847854]
Middle of the cycle	.9867811	2.682586	.2143875	4.60	0.000	[.5665894	1.406973]
Before period begins	1.10642	3.023516	.2749776	4.02	0.000	[.5674743	1.645367]
At any time	.6982363	2.010204	.1901679	3.67	0.000	[.3255139	1.070959]
Cont_Method							
Modern Method	2.624092	13.79205	.1929167	13.60	0.000	[2.245983	3.002202]
Media							
yes	-.3016984	.7395611	.1482861	-2.03	0.042	[-.5923339	-.0110629]
_cons	-3.930277	.0196382	.2998857	-13.11	0.000	[-4.518042	-3.342512]

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
Region: Identity				
var(_cons)	.1108515	.0775842	[.0281185	.43701]

LR test vs. logistic regression: chibar2(01) = 8.13 Prob>=chibar2 = 0.0022

Akaike's information criterion and Bayesian information criterion

Model	**Obs**	**ll(null)**	**ll(model)**	**df**	**AIC**	**BIC**
.	3498	.	-916.3357	16	1864.671	1963.231

Residual intra-class correlation

Level	**ICC**	**Std. Err.**	**[95% Conf. Interval]**	
Region	.0325965	.0220704	[.0084746	.117259]

LR test vs. logistic regression: chibar2(01) = 8.35 Prob>=chibar2 = 0.0019

4.4.4. Random Slope Multilevel Logistic Regression Analysis

Table (4.12) Shows that the comparisons of Different Models with Different random slopes. As seen in the table below p-value for all variables is less than 0.05 suggesting that the variations of all variables from region to region is significant at 5% level of significance. And also the table indicates that the variables; Religion, Wealth Index, Knowledge of Ovulatory Cycle and Media Exposure have very small regional variations and the variables **Education Level** and **Contraceptive Methods** have high regional variations compared to the others, indicating that they are two most varying variables through region and are permitted to analyze random slope Model.

Table(4.12) *Comparison of Different models with different Random slopes*

Models	Source of Random Slope	**Loglik**	AIC	BIC	Chi2()	P-Value	Regional Variation
Model1	Religion	**-916.3357**	1865.748	1976.627	11.05	0.00114	**0.0374**
Model2	Edu.Level	**-914.48**	1864.96	1975.84	11.84	0.0079	**0.0844**
Model3	Wealth Index	**-916.33**	1868.659	1979.538	8.14	0.0432	**0.0002**
Model4	KnoOvuCycle	**-915.2054**	1866.411	1977.29	10.39	0.0155	**0.009**
Model5	Cont.Method	**-906.9147**	1849.829	1960.709	26.97	0.000	**1.467**
Model6	Media Exposs.	**-915.878**	1867.755	1978.634	9.05	0.0287	**0.00224**
Model7	Logistic	**-903.1437**	1856	2010.286	-	-	**-**
Model8	Cont.Method and Edu.Level	**-901.3247**	1844.649	1974.008	38.15	0.000	**-**
Model9	Cont.Method and Religion	Hessian is not negative semi-definite (Itr=6, loglik= -903.69056)					
Mode10	Religion&Education level	**-913.2903**	1868.58	1997.94	14.22	0.0273	**-**

Then as shown in Table (4.12) the resulting better random slope multilevel Logistic Regression model is Model8 which has smallest IC and largest log-likelihood as seen in the above table.

Variance Covariance Matrix of random effects for Region Level

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\delta}_0^2 & \boldsymbol{\delta}_{20} & \boldsymbol{\delta}_{50} \\ \boldsymbol{\delta}_{02} & \boldsymbol{\delta}_2^2 & \boldsymbol{\delta}_{52} \\ \boldsymbol{\delta}_{05} & \boldsymbol{\delta}_{25} & \boldsymbol{\delta}_5^2 \end{bmatrix} = \begin{bmatrix} \mathbf{0.0214262} & \mathbf{-0.0449959} & \mathbf{0.1995792} \\ \mathbf{-0.0449959} & \mathbf{0.194058} & \mathbf{-0.5556981} \\ \mathbf{0.1995792} & \mathbf{-0.5556981} & \mathbf{2.046371} \end{bmatrix}$$

Test of Hypothesis for part of random effect

H_0 : $\boldsymbol{\delta_{00} = 0}$ versus $\boldsymbol{\delta_{00}} \neq 0$ as shown in Table (4.12) the 95% confidence interval for the estimate of random intercept is not negative and different from zero. This indicates that there is a significant Heterogeneity (variance difference) between Regional states of Ethiopia. Implying Teenagers' fertility varies from region to region.

H_0 : $\boldsymbol{\delta_{22} = 0}$ versus $\boldsymbol{\delta_{22}} \neq 0$ and $\boldsymbol{\delta_{55} = 0}$ versus $\boldsymbol{\delta_{55}} \neq 0$ As seen in the table (4.12) the 95% Confidence interval for random slopes for both test is non-negative and different from zero implying that random slope for Teenagers' Education Level and Contraceptive Method used in the region is significant. This Indicates that Teenagers Fertility varies from region to region concerning with Teenagers education level and Contraceptive Method used that means there is a significant variation of teenagers Education level and Contraceptive Method used from region to region. The var(u_{0j}), var(u_{2j}), and var(u_{5j}), are the estimated variance of intercept, slope of teenagers education level and slope of teenagers Contraceptive Method used respectively.

The parameter interpretation for multilevel random slope plus Intercept Logistic Regression analysis:-

The Analysis revealed that within a region and holding other variables constant, the odds of teenagers who follows other religion group were (OR=2.276) times more likely than that of Orthodox and odds of teenagers' who follows Muslim were (OR=2.83) times more likely than that of teenagers in Orthodox.

The effect of Education Level is highly significant and negative implying that within a region, have negative relationship with log-odds of teenagers fertility and teenagers' having primary, secondary and Higher education are associated with decreased odds of teenagers' fertility. The result shows that within a region and holding other variables constant Odds of teenagers with primary education is (OR=0.426) times less likely than that of teenagers with no education, Odds of teenagers' with secondary education is (OR=0.2555) times less likely than that of teenagers with no education and Odds of teenagers with higher education is (OR=.103) times less likely than that of Teenagers with no education.

And also for a random slope part with in a region and holding other variables constant the Odds specific to teenagers education level is multiplied by exp[$\mathbf{u_{2j}EduLevel_{ij}}$]. Where j=1, 2,….., 11

That means mathematically:-

$\frac{\pi_{ij}/u_{0j}}{1-\pi_{ij}/u_{0j}} = \exp(\beta_0 + \cdots + \beta_{02}\text{Edu Level}_{ij} + \cdots \ldots\ldots + u_{2j}\text{EduLevel}_{ij})$. So is a multiple of exp[$\mathbf{u_{2j}EduLevel_{ij}}$].

The outcome of wealth index is highly significant and positive indicating that within a region teenagers', of having wealth Index poor and Middle are associated with Increased Odds of teenagers' fertility. The outcome shows that within region and holding other variables constant the Odds of teenagers, with wealth index poor is (OR=2.883) times more likely than that of teenagers with wealth Index Rich and Odds of teenagers with wealth Index Middle is (OR=2.335) times more likely than that of teenagers with wealth index Rich.

The outcome of knowledge of Ovulatory cycle is positive and highly significant suggesting that with in a region and holding other variables constant, teenagers' having knowledge of Ovulatory cycle; during her period, After period ended, Middle of the cycle, Before period begins and At any time are associated with Increased Odds of teenagers' fertility. The result reveals that within a region and holding other variables constant ; the Odds of teenagers with knowledge of Ovulatory cycle at During her period is (OR=2.61) times more likely than that of teenagers' who don't know, Odds of teenagers with knowledge of Ovulatory cycle After period ended is (OR=4.564) times more likely than that of teenagers' Don't know, Odds of teenagers' with knowledge of Ovulatory cycle at Middle of the cycle is (OR=2.9) times more likely than that of reference, Odds of teenagers' with knowledge of Ovulatory cycle before period begins is (OR=2.986) times more likely than that of the reference and Odds of teenagers with knowledge of Ovulatory cycle at any time is(OR=2.071) times more likely than that of the reference.

And also for a random slope part within a region and holding other variables constant the Odds specific to teenagers Contraceptive Method used is multiplied by exp[$\mathbf{u_{5j}Cont\ Method_{ij}}$]. That means mathematically it can be represented as:-

$\left[\frac{\pi_{ij}/u_{0j}}{1-\pi_{ij}/u_{0j}}\right]$= exp $(\beta_0 + \cdots \ldots .. + \beta_{05}\text{Cont Method}_{ij} + \cdots \ldots \ldots + u_{5j}\text{Cont Method}_{ij})$. So is a multiple of exp[$\mathbf{u_{2j}Cont\ Method_{ij}}$].

Within region and holding other variables constant; the odds of Modern contraceptive Method used teenagers' is (OR=18.46) times more likely than that of reference category.

Table(4.12) Results of Random slope Multilevel Logistic Regression Analysis

Log likelihood = -901.32466 Prob > chi2 = 0.0000

FertilityStatus	**Coef(β).**	Odds ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]	
Religion								
Others	.8224863	2.276152	.2466771	3.33	0.001	[.3390081	1.305965]
Muslim	1.040384	2.830303	.1984133	5.24	0.000	[.6515008	1.429267]
Edu_Level							
Primary	-.8529002	.4261771	.2067004	-4.13	0.000	[-1.258026	-.4477748]
Secondary	-1.364583	.2554872	.3763602	-3.63	0.000	[-2.102235	-.6269306]
Higher	-2.27489	.1028082	.8928239	-2.55	0.011	[-4.024793	-.5249874]
WealthEndex							
poor	1.058784	2.882862	.1793554	5.90	0.000	[.7072535	1.410314]
Middle	.8481119	2.335234	.2103075	4.03	0.000	[.4359168	1.260307]
KnowlOvuCycle							
During her period	.9589383	2.608925	.3467006	2.77	0.006	[.2794176	1.638459]
After period ended	1.518106	4.563572	.1846378	8.22	0.000	[1.156222	1.879989]
Middle of the cycle	1.064731	2.900057	.2190734	4.86	0.000	[.6353546	1.494106]

Before period begins	1.093862	2.985783	.2831784	3.86	0.000	[.5388424	1.648881]
At any time	.7278221	2.070566	.1928218	3.77	0.000	[.3498983	1.105746]
Cont_Method							
Modern Method	2.915632	18.46047	.5033395	5.79	0.000	[1.929105	3.902159]
Media							
yes	-.2769216	.7581139	.1506539	-1.84	0.049	[.5721979	.6183547]
_cons	-3.782807	.0227587	.2855193	-13.25	0.000	[-4.342415	-3.2232]

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
Region: Unstructured				
var(Cont_M~d)	2.046371	1.270492	[.6060591	6.909615]
var(Edu_Le~l)	.194058	.1339266	[.0501748	.750546]
var(_cons)	.0214262	.0375757	[.0006889	.6663766]
cov(Cont_M~d,Edu_Le~l)	-.5556981	.3540378	[-1.249599	.1382032]
cov(Cont_M~d,_cons)	.1995797	.1977573	[-.1880174	.5871769]
cov(Edu_Le~l,_cons)	-.0449959	.0710352	[-.1842223	.0942304]

LR test vs. logistic regression: chi2(6) = 38.15 Prob > chi2 = 0.0000

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	3498	.	-901.3247	21	1844.649	1974.008

Model Diagnostics for Random slope plus Intercept Multilevel Model.

The diagnostic plots below indicates that the normal distributuion of random effects in random Intercept plus slope model. Which means random effects are normaly distributed with mean zero and variance **Ω,**

$$\mathbf{u} \sim (0, \Omega) = \left(0, \begin{bmatrix} 0.0214262 & -0.0449959 & 0.1995792 \\ -0.0449959 & 0.194058 & -0.5556981 \\ 0.1995792 & -0.5556981 & 2.046371 \end{bmatrix} \right)$$

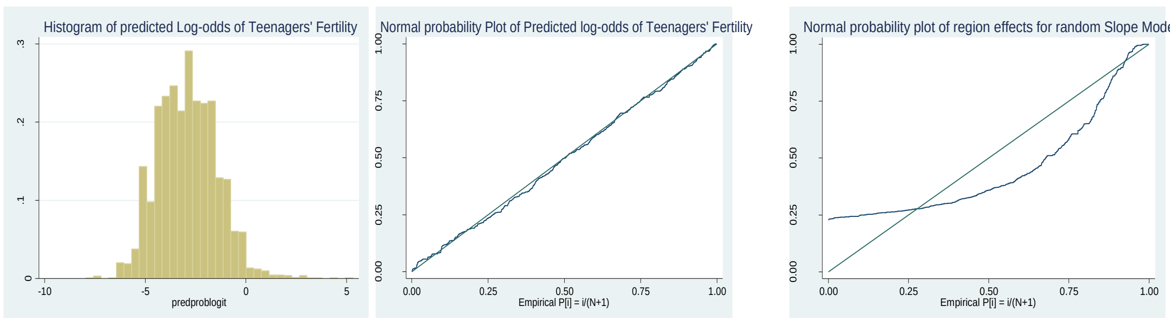


Fig3 *Graphs of Model diagnostics*

4.5. Population Averaged Model (GEE)

Comparison of Empirical and Model based standard errors for two proposed Models

According to Table (4.13), when we compare standard errors of exchangeable correlation structure, model based is a good fit compared to Empirical, and also when we compare standard errors of Independent correlation structure also model based is a good fit.

The Table (4.13) blow also shows that, when we compare model based for exchangeable and model based for Independent, almost the two models are the same, but a little, more a model with Independent correlation structure modifies the analysis.

Table4.13 Empirical and Model Based Standard Errors for two proposed Models

	Exchangeable			Independent		
		Model based	Empirical		Model based	Empirical
Coeff.	Estimates	(S.E)	(S.E)	Estimates	(S.E)	(S.E)
B0	-3.8140	**.3043**	.3071	-3.997729	**.2711035**	.3328751
B1	1.0829	**.2241**	.2584	1.180237	**.1990544**	.2711719
B2	1.1447	**.1952**	.2091	1.259729	**.1695048**	.2101965
B3	-.7176	**.1497**	.0969	-.73334	**.149481**	.0971103
B4	-1.0165	**.2345**	.3849	-1.001062	**.2367383**	.4035261
B5	-1.6547	**.7012**	.6219	-1.719955	**.7546418**	.6837189
B6	.9800	**.1713**	.1855	1.079612	**.1670749**	.1934779
B7	.7270	**.1980**	.1909	.7716518	**.1989092**	.2163864
B8	.9233	**.3335**	.4128	.9526163	**.3374469**	.4145705
B9	1.4500	**.1824**	.1836	1.452186	**.1801321**	.1777699
B10	.9873	**.2109**	.1985	.999043	**.2109318**	.1835781
β11	1.1068	**.2678**	.2824	1.109708	**.2707682**	.2732709
β12	.6944	**.1874**	.2099	.6886615	**.1872505**	.2181018
β13	2.4827	**.1901**	.4301	2.53907	**.1860938**	.4464966
β14	-.2826	**.1447**	.0930	-.3059393	**.1464631**	.1040002

Graphical Comparisons of the two models

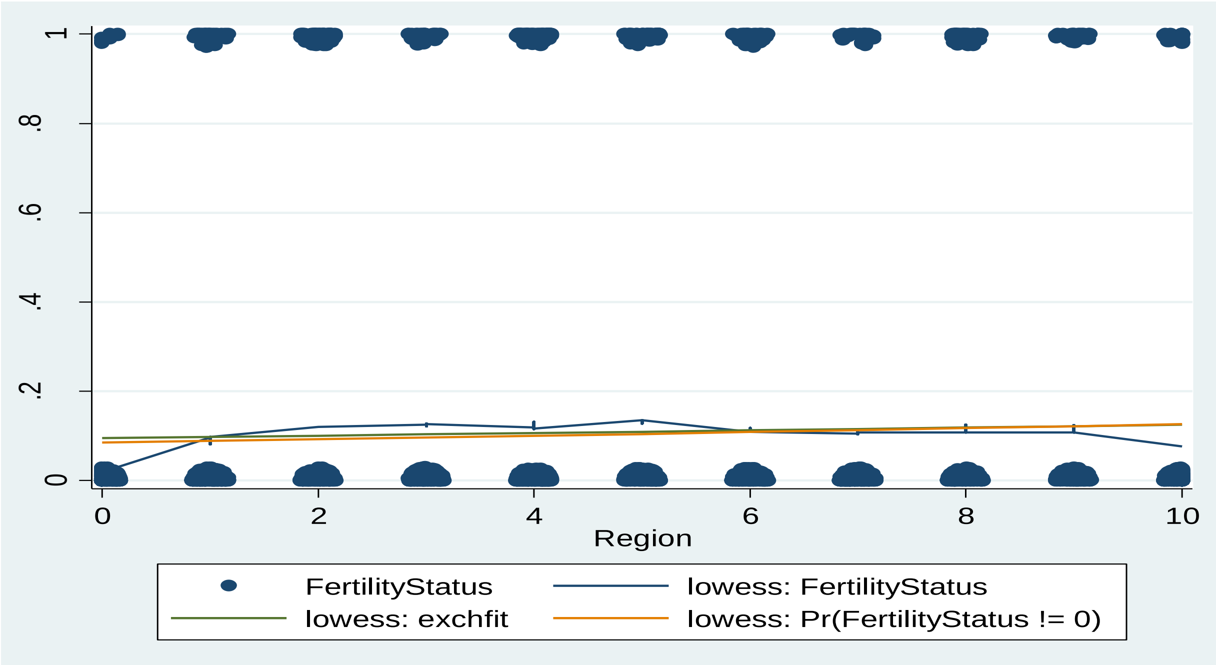


Fig4 *Graph of Model comparison*

As we seen in the Graph above the yellow line(indfit) highly touch the fertility status line as compared to the green(exchfit), it implies that, this model is a good fit for the given data. So the marginal model (GEE) with Independent correlation structure is the final model thus interpretation is based on this model.

Parameter Interpretation in Population Averaged Model:-

The only difference between multilevel and population averaged model is that, unlike multilevel, parameters, in population averaged model is marginalized (averaged). As indicated in Table(4.14), On average, for the Ethiopian teenagers' in this study, the odds of teenagers' Fertility in religion group others increases by 225% compared with average teenagers Fertility in Orthodox, and on average for the Ethiopian Teenagers' the odds of teenagers' Fertility in religion Muslim is about 252% higher than that of average teenagers' fertility in Orthodox.

On average, for the Ethiopian teenagers' in this study, the odds of teenagers' Fertility with primary education decreased by 52% compared to fertility of teenagers' with no education, the odds of teenagers' Fertility with secondary education decreased by about 63% compared to teenagers' with no education and the odds of teenagers' with Higher education is decreased by 82% compared to teenagers' fertility with no education. On average for Ethiopian teenagers' in this study, the odds of teenagers' fertility with wealth Index poor is increased by 194% compared to teenagers fertility with wealth Index rich and odds of teenagers' with wealth Index Middle is 116% higher than that of teenagers 'fertility with wealth Index rich.

On average, for the Ethiopian teenagers' in this study, Odds of teenagers with knowledge of Ovulatory cycle at During her period is increased by 159% compared to teenagers with no knowledge of ovulatory cycle, odds of teenagers with knowledge of Ovulatory cycle at After period ended is increased by 327% compared to teenagers' with no knowledge of ovulatory cycle, Odds of teenagers with knowledge of Ovulatory cycle at Middle of the cycle is increased by 171% compared to teenagers with no knowledge of ovulatory cycle, Odds of teenagers with knowledge of Ovulatory cycle at Before period begins is increased by 203% compared to teenagers with no knowledge of ovulatory cycle and Odds of teenagers with knowledge of Ovulatory cycle at any time is increased by 99% compared to teenagers with no knowledge of ovulatory cycle.

On average, for the Ethiopian teenagers' in this study, odds of teenagers' fertility using modern method is 11.67 times higher than that of using no methods and Traditional methods. The odds of teenagers' fertility exposed to any mass media is decreased by 26% compared to that teenagers' not exposed to any mass media.

Table(4.14) *Analysis of population Averaged Model*

Fertility Status	Coef.	OR	Std. Err.	z	P>|z|	[95% Conf. Interval]
Religion						
Others	1.180237	3.255144	.1990544	5.93	0.000	[.7900971 1.570376]
Muslim	1.259729	3.524467	.1695048	7.43	0.000	[.9275058 1.591953]
Edu_Level						
Primary	-.73334	.4803021	.149481	-4.91	0.000	[-1.026317 -.4403627]

Secondary	-1.001062	.3674891	.2367383	-4.23	0.000	[-1.46506	-.537063]
Higher	-1.719955	.1790742	.7546418	-2.28	0.023	[-3.199026	-.2408842]
WealthEndex							
poor	1.079612	2.943537	.1670749	6.46	0.000	[.7521512	1.407073]
Middle	.7716518	2.163337	.1989092	3.88	0.000	[.381797	1.161507]
KnowlOvuCycle							
During her period	.9526163	2.592483	.3374469	2.82	0.005	[.2912325	1.614]
After period ended	1.452186	4.272445	.1801321	8.06	0.000	[1.099134	1.805239]
Middle of the cycle	.999043	2.715682	.2109318	4.74	0.000	[.5856242	1.412462]
Before period begins	1.109708	3.033473	.2707682	4.10	0.000	[.5790124	1.640404]
At any time	.6886615	1.991049	.1872505	3.68	0.000	[.3216572	1.055666]
Cont_Method							
Modern Method	2.53907	12.66789	.1860938	13.64	0.000	[2.174333	2.903807]
Media							
yes	-.3059393	.7364313	.1464631	-2.09	0.037	[-.5930017	-.0188768]
_cons	-3.997729	.0183573	.2711035	-14.75	0.000	[-4.529082	-3.466376]

4.6. Discussion

This study suggested that teenagers' Fertility, in Ethiopia, is about 10.26%. That means, there are about 103 fertile teenagers' for every 1000 teenagers in Ethiopia. According to the study, done by Singh and Darroch(28), this result shows high rate of fertility compared to western countries, the result shows low fertility rate compared to report from EDHS 2011 and the previous study done by Admias (21) in 2013, and the original article written by Dereje and Jemal(20) in 2014 which estimates teenage pregnancy rate as 20.4%.

The variable Ethnicity is significant which is inconsistent with (46) but it disturbs the overall model because of influential observations in Gumuz and Agnuk(see Table 4.3.6 in Appendix A) as a result the variable Ethnicity is excluded from this study. And it is also not significant in the previous study done by same Authors (46) in Assossa General Hospital in 2015. The variable residence is significant in univariable analysis and associated with teenagers' fertility, in chi-square test of association but it is not significant in multiple logistic regression analysis this is inconsistent with the previous study done by Admias(21) in Gondor University.

Analysis from logistic regression indicates that the factors: - Region, Religion, Education level, Wealth Index, Knowledge of Ovulatory Cycle, Contraceptive Method and Media exposure have significant effects on teenagers' fertility. These results are consistent with; the previous study done by Admias(21) in Gondor University except variables, Residence and Working status which are not significant in this study, consistent with previous study in Brazil, and is consistent with the previous study done by Tewodros and Jemal(20) by using 2005EDHS data.

As seen in the Logistic regression analysis, teenagers' fertility has significant association with geographical regions with positive coefficients which was similar with the study done by Admias(21) in Gondor University. It indicates that the Odds of teenagers living in Affar, Oromia, Somali, Gambella and Harare are more likely to become fertile before nineteen years of age than teenagers living in Addis Ababa which is inconsistent with Admias(21).This is logically, due to teenagers' fertility difference among regional states Ethiopia, and while the Odds of the remaining five regions are not significantly different from Addis Ababa.

Since test of heterogeneity is highly significant indicating significant variation between regions which is consistent with Dechasa(69) and Abebe(70); thus the multilevel modeling is good to fit EDHS 2016 data. According to IC and log-likelihood the good model to fit this data is multilevel logistic regression with one random intercept and two random slopes. Unlike the studies done by Dechasa (51) and Abebe (52) in Jimma University at 2014 and 2015 respectively, it selected the slope model as the good model, i.e region is a source of random intercept and the two variables

(Contraceptive Method and Education level) which are mostly varying through region to region are the sources of random slopes.

The empty multilevel random intercept model indicates that the variance of constant is .372 and highly significant implying that there is a significant fertility variation between regional states of Ethiopia which is consistent with the study done by Kosunen(33) and when compared to single level empty model; it has large log-likelihood and small IC indicating multilevel modeling is better to fit teenagers fertility data in Ethiopia. Also the result indicates that the ICC for empty model is 0.1016461 suggesting that 10.16% of the variance in teenagers' fertility could be attributed to difference across region.

The full random intercept model is highly significant with p-value = 0.0019 and variance for constant is 0.111 which decreased from 0.372 in empty multilevel model to full multilevel model indicating that determinants of teenagers fertility accounted for decreasing variations through regional states of Ethiopia.

The final multilevel model analysis revealed that, the effects of religion, that teenagers follow, teenagers' education level, teenagers wealth index, teenagers knowledge of Ovulatory Cycle ,Contraceptive method used status and teenagers exposure to any mass media are found to be significant indicating significant effects on teenagers fertility with in region.

In a case of population averaged model (GEE), four models (two **model** based and two **Empirical**) are compared by their standard errors and graphical methods, and from the four models, two models with **model** based standard errors (Model based exchangeable and Model based independent) are selected, and then the two models are compared and the model with Independent correlation structure is selected as the final model. And also the two selected models are compared by graphical methods of comparison, and the same model is selected as the final model.

The final population averaged (marginal (GEE)) model indicates that, the religion effect, the effect of education level, the effect of wealth index, the effect of knowledge of Ovulatory Cycle, the effect of contraceptive method used and the effect of exposure to any mass media are found to significant implying significant average effects on teenagers' fertility in Ethiopia.

The result, that the probability of teenagers' fertility was increasing, for teenagers who have knowledge about ovulatory cycle, is not expected. This may be, caused by, teenagers' who had knowledge of ovulatory cycle tried to protect fertility by using periods concerning with ovulatory cycle or It may be, caused by teenagers' who have no knowledge of ovulatory cycle abstain themselves from sexual intercourse or they could use effective contraceptive method.

Also the result, that the probability of teenagers' fertility was increasing, for teenagers' who have been used modern contraceptive method, which is unexpected and it may be, caused by luck of knowledge about how to use modern methods of contraception.

CHAPTER FIVE

5. Conclusion and Recommendations

5.1. Conclusions

The main aim of this study was to asses socio-economic, demographic, other determinants associated with teenagers' fertility outcome in Ethiopia. The rate of teenage who fertile before nineteenth was about one tenth of the total teenage in this study and their proportion was less than half of the quarter of all teenagers in this study. The study identified that; Region, Religion, Education level, Wealth Index, Knowledge of Ovulatory cycle, Contraceptive Method and Media were significant predictors for Teenagers fertility in Ethiopia and Contraceptive Method and Education level were the source of slope model in multilevel modeling for this study. Among the cluster-specific models the random slope multilevel model provided the best fit for the data under consideration for the analysis of teenagers' fertility variation among regional states of Ethiopia.

Teenagers are nested with in regions, as a result the study concluded that: using standard error multilevel logistic regression model was better fitted for the analysis of teenagers' fertility outcomes when the researcher interested in cluster-specific variation and Marginal models when the researcher interested in averaged outcome.

In the final model (The random slope Multilevel model), for non-marginal models, found that; Religion, Education level, Wealth Index, Knowledge of Ovulatory cycle, Contraceptive Method and Media are associated with increased risk of teenagers' fertility outcome in Ethiopia and the same factors are associated in increased risk of teenagers' fertility in selected marginal model (Model based model with independent correlation structure).

5.2. Recommendation

Based on the findings of this study, the researcher recommends the responsible bodies in the following important issues. Since there is a significant teenagers' fertility variation among regional states of Ethiopia; then the Ethiopian government has to take remedial action as general and Regional states of Ethiopia as specific, on maternal health policy and design strategies to protect teenagers' fertility problem considering the major factors that affect teenagers' fertility and contributing to variations among regional states to minimize teenagers fertility based on the following recommendations:

- Supporting teenagers to complete at least secondary education and above.
- Encouraging and giving special support for teenagers who have wealth index poor and middle.
- Special attention is paid, for teenagers' who have knowledge of ovulatory cycle, because they are more likely fertile than teenagers' don't have knowledge of ovulatory cycle which is unusual it may be caused by gap of knowledge.
- Also a special attention must be taken for teenagers who used Modern contraceptive method, because those teenagers are more likely to become fertile before their nineteen, this may be caused by luck of knowledge about how to use modern contraceptive methods.
- Trying to expand any mass media in rural and urban parts of the country.
- Each religious institution must aware teenagers about the problem that follow being fertile before nineteenth years of age.
- Every researcher would do similar analysis within each region, concerning their variations and including another determinants those are not included in this study.
- Every community (especially mothers) must to aware teenagers about the period of ovulatory cycle, to avoid pregnancy before nineteen years of age.
- For correlated data it is better to use subject-specific or cluster-specific Models if interested in subject (cluster) specific random effects and Marginal models if interested in average effects.
- From marginal models try to use Alternating logistic regression (ALR) which was not included in this paper but it may be preferable than GEE.

Reference

1. Unicef. Young people and family planning: Teenage pregnancy. UNICEF, New York. 2008.
2. WHO . Adolescent Pregnancy :Issues in Adolescent Health and Development. Geneva: WHO. 2004.
3. UNFPA A. a.convention in Ethiopia in January 1997 by the funds Africa division: adolescent.1997
4. Chawhanda C. Proximate Determinants of Teenage Fertility in Zimbabwe (Doctoral dissertation, University of KwaZulu-Natal, Durban).
5. World Health Organization. Programming for adolescent health and development: report of a WHO/UNFPA/UNICEF Study Group on Programming for Adolescent Health.
6. Seifu A. Reproductive health needs of urban and rural out of school adolescents in Eastern Gojjam. Addis Ababa: Addis Ababa University. 2001:42-5.
7. United Nations Fund for Population Activities. Giving adolescent girls the chance to reach their full potential. 2008 [cited 2015].
8. Alemayehu T, Haider J, Habte D. Determinants of adolescent fertility in Ethiopia. Ethiopian Journal of Health Development. 2010; 24(1).
9. Neema S, Musisi N, Kibombo R. Adolescent sexual and reproductive health in Uganda: a synthesis of research evidence. New York: Alan Guttmacher Institute; 2004 Dec.
10. United Nations Fund for Population Activities. Giving adolescents today and tomorrow: Breaking the cycle of adolescent pregnancy.
11. Reynolds HW, Wong EL, Tucker H. Adolescents' use of maternal and child health services in developing countries. International family planning perspectives. 2006 Mar 1:6-16.
12. Bearinger LH, Sieving RE, Ferguson J, Sharma V. Global perspectives on the sexual and reproductive health of adolescents: patterns, prevention, and potential. The lancet. 2007 Apr 7;369(9568):1220-31.
13. Save the Children. Mothers & children in war & conflict: state of the world's mothers 2002. Save the Children; 2002.

14. Melanie Guldi Derterminants of Teenage Fertility. 2017 University of Central Flirida.
15. Senanayake P, Ladjali M. Adolescent health: changing needs. International Journal of Gynecology & Obstetrics. 1994 Aug; 46(2):137-43.
16. Singh S. Adolescent childbearing in developing countries: a global review. Studies in family planning. 1998 Jun 1:117-36.
17. Rossiter CE, Chong H, Lister UG, BANO Q, Briggs ND, Ekwempu CC, MEMBERR M. 4. The influence of maternal age and parity on child-bearing with special reference to primigravidae aged 15 years and under. BJOG: An International Journal of Obstetrics & Gynaecology. 1985 Dec 1; 92:23-31.
18. EDHS., CSA AND ICF. (2016 , 2012).
19. Sayem AM, Nury AT. Factors associated with teenage marital pregnancy among Bangladeshi women. Reproductive health. 2011 Dec; 8(1):16.
20. Tewodros A, Jemal H, Dereje H. Utilization of antenatal care services among teenagers in Ethiopia: a cross sectional study. Ethiopian Journal of Health Development. 2010;24(3):221-5
21. Adimias Wendimagegn Applying Logistic Regression on Determinants of Teenagers Fertility EDHS (2011) Data in Ethiopia.2013.
22. James E Rosen. adolescence pregnancy is an issue that calls for more education and support to encourage girls to delay mother hood until they are ready:. WHO bulletinvolumes 2011. 87/6/2011.
23. WHO. Adolescent pregnancy: a culturally complex issue. Bull World Health Organ. 2009 June, 2009; 87(6): 410–411.
24. Sibusiso Mkwananzi* and Clifford Odimegwu* Teen Pregnancy in Sub-Saharan Africa :the Application of social Disorganization Theory. 2015 Annual Meeting –April 30-May 2,San,Diago,CA.
25. Anteneh Mulgeta Determinants of Adolescent Fertility among Rural Women of Ethiopia. 2016
26. Md. Hasinur Rahaman Khan: Multilevel Logistic Regression Analysis of Contraceptive Prevalence in Bangladesh. 2008.
27. World Health Organization: Reducing Early and Unintended pregnancies among adolescents. Family Planning. 2017

28. Singh S, Darroch JE. Adolescent pregnancy and childbearing: levels and trends in developed countries. Family planning perspectives. 2000 Jan 1:14-23.
29. Bongaarts J, Watkins SC. Social interactions and contemporary fertility transitions. Population and development review. 1996 Dec 1:639-82.
30. Adolescent Fertility rate (Births per 1,000 women ages 15-19) sited in https://data.worldbank.org/indicator/SP.ADO.TFRT. 2017
31. World Conference on Women, United Nations. Report of the Fourth World Conference on Women: Beijing, 4-15 September 1995. na; 1996.
32. Alli F, Maharaj P, Vawda MY. Interpersonal relations between health care workers and young clients: barriers to accessing sexual and reproductive health care. Journal of community health. 2013 Feb 1; 38(1):150-5.
33. Lesthaeghe R, Jolly C. The start of the sub-Saharan fertility transition: Some answers and many questions. Journal of International Development. 1995 Jan; 7(1):25-45.
34. Gage AJ. Sexual activity and contraceptive use: the components of the decisionmaking process. Studies in family planning. 1998 Jun 1:154-66.
35. Cohen B. Fertility levels, differentials, and trends. Demographic change in sub-Saharan Africa. 1993 Jan 15:8-67.
36. Robinson WC. Kenya enters the fertility transition. Population studies. 1992 Nov 1; 46(3):445-57.
37. Rutenberg N, Diamond I. Fertility in Botswana: The recent decline and future prospects. Demography. 1993 May 1; 30(2):143-57.
38. Cohen B. The emerging fertility transition in sub-Saharan Africa. World Development. 1998 Aug 1; 26(8):1431-61.
39. McDevitt TM, Adlakha A, Fowler TB, Harris-Bourne V. Trends in Adolescent Fertility and Contraceptive Use in the Developing World, US Bureau of the Census. Washington DC. 1996.
40. MELHUISH E, PHOENIX A. Motherhood under twenty: Prevailing ideologies and research. Children & Society. 1987 Dec; 1(4):288-98.
41. Govindasamy P, Kidanu A, Bantayerga H. Youth reproductive health in Ethiopia: OCR macro; 2002. Google Scholar.

42. Narring F, Michaud PA, Sharma V. Demographic and behavioral factors associated with adolescent pregnancy in Switzerland. Family Planning Perspectives. 1996 Sep 1:232-6.
43. CSA. Ethiopian Demographic and Health Survey . 2016 Addis Abeba: CSA.
44. UNFPA.National Adolescent Fertility Study.2016
45. Kosunen EA, Vikat A, Gissler M, Rimpelä MK. Teenage pregnancies and abortions in Finland in the 1990s. Scandinavian journal of public health. 2002 Oct; 30(4):300-5.
46. Beyene A, Muhiye A, Getachew Y, Hiruye A, Mammo D, Hailemariam D, Derbew M, Enquselassi F. Assessment of the magnitude of teenage pregnancy and its associated factors among teenage females visiting Assosa General Hospital. Ethiopian medical journal. 2015 Jul 1;53.
47. Amsalu Arega Modeling Delivery Care Service Utilization Of Mothers In Ethiopia.2014
48. Chiavegatto Filho AD, Kawachi I. Income inequality is associated with adolescent fertility in Brazil: a longitudinal multilevel analysis of 5,565 municipalities. BMC public health. 2015 Dec;15(1):103.
49. Yebyo H, Alemayehu M, Kahsay A. Why do women deliver at home? Multilevel modeling of Ethiopian National Demographic and Health Survey data. PLoS One. 2015 Apr 15;10(4):e0124718.
50. Daniel Biftu Modeling Determinants of Low Birth Weight for Under Five- Children in Ethiopia.2015
51. Dechasa Bedada Statistical Analysis Of Regional Heterogeneity Of Under-Five Child Mortality In Ethiopia 2014.
52. Abebe Debu Determinants Of Obstetric Fistula In Ethiopia: An Application Of Binary And Multilevel Logistic Regression Models 2015.
53. Agresti. An introduction to Categorical Data Analysis. 1996 and 2008. New York: John Wiley and Sons,.
54. McCullagh P, Nelder JA. Generalized linear models. CRC press; 1989 Aug 1.
55. Christensen R. Log-linear models and logistic regression. Springer Science & Business Media; 2006 Apr 6.
56. Dayton CM. Logistic regression analysis. Stat. 1992 Sep: 474-574.
57. Generalized Linear Models. 2012

58. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression. John Wiley & Sons; 2013.
59. Geremew Muleta. Applied Generalized Linear Models (GLM) using R:An introduction to Statistical modeling using R.
60. Cook RD, Weisberg S. Residuals and influence in regression. New York: Chapman and Hall; 1982.
61. Agresti A. An introduction to categorical data analysis. Wiley; 2018 Nov 20.
62. Dickey DA, Rawlings JO, Pantula SG. Applied regression analysis.
63. Logistic Regression Models for the Analysis of Correlated Data.
64. Molenberghs G, Verbeke G. Missing data concepts. Models for Discrete Longitudinal Data. 2005:481-8.
65. Khan H, Shaw E. Multilevel logistic regression analysis applied to binary contraceptive prevalence data.
66. Tom AB, Bosker TA, Bosker RJ. Multilevel analysis: an introduction to basic and advanced multilevel modeling. Sage; 1999 Sep 29.
67. Harvey G. Multilevel statistical models.
68. Hubbard AE, Ahern J, Fleischer NL, Van der Laan M, Satariano SA, Jewell N, Bruckner T, Satariano WA. To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. Epidemiology. 2010 Jul 1:467-74.
69. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. Regression methods in biostatistics: linear, logistic, survival, and repeated measures models. Springer Science & Business Media; 2011.
70. Narring F, Michaud PA, Sharma V. Demographic and behavioral factors associated with adolescent pregnancy in Switzerland. Family Planning Perspectives. 1996 Sep 1:232-6.

Appendixes

Appendix A

Estimation of between and within region variance								
Region	n_j	$n_j * n_j$	$\widehat{\pi_j}$	$(1-\widehat{\pi_j})$	$(n_j * \widehat{\pi_j}) * (1-\widehat{\pi_j})$	$(\widehat{\pi_j}-\widehat{\pi})$	$\frac{(\widehat{\pi_j}-\widehat{\pi})^2}{\widehat{\pi}(1-\widehat{\pi})}$	$n_j * \frac{(\widehat{\pi_j}-\widehat{\pi})^2}{\widehat{\pi}(1-\widehat{\pi})}$
Tigray	423	178929	0.083	0.917	32.10402	-0.0199	0.016983	7.18385
Affar	266	70756	0.207	0.793	43.62782	0.1041	0.465648	123.8624
Amhara	355	126025	0.070	0.930	23.23944	-0.0322	0.044541	15.81211
Oromia	415	172225	0.137	0.863	49.17108	0.0347	0.05176	21.48033
Somali	319	101761	0.154	0.846	41.47335	0.0510	0.111574	35.59213
Benishangul	237	56169	0.106	0.894	22.36287	0.00286	0.00035	0.082962
SNNPR	391	152881	0.069	0.931	25.13555	-0.0336	0.048408	18.92739
Gambela	212	44944	0.156	0.844	27.86321	0.0530	0.120753	25.59964
Harari	183	33489	0.142	0.858	22.30601	0.0394	0.066814	12.22691
Addis Adaba	432	186624	0.019	0.981	7.851852	-0.0841	0.30378	131.2329
Dire Dawa	265	70225	0.072	0.928	17.63774	-0.0309	0.041083	10.88697
	3498	**1194028**			**312.7729**			**402.8876**

Where, $\pi = \frac{\text{total numember of fertile teenage in this study}}{\text{Total numember of teenage in this study}} = \frac{359}{3498} = \mathbf{0.10263}$

$$x^2=\sum_{j=1}^{N} n_j \frac{(\hat{\pi}_j-\hat{\pi})}{\hat{\pi}(1-\hat{\pi})} = \mathbf{402.8876}$$

$\tilde{n}$ Is defined as: $\tilde{n} = \frac{1}{N-1}\left[M - \frac{\sum_{j=1}^{N} n_j^2}{M}\right] = \frac{1}{11-1}\left[3498 - \frac{1194028}{3498}\right] = \mathbf{351.665}$

$$s_{between}^2 = \frac{\hat{\pi}(1-\hat{\pi})}{\tilde{n}\,(N-1)} x^2 = \frac{0.10263(1-0.10263)}{351.665(10)} * 402.8876 = \mathbf{0.010551172265393}$$

$$s_{within}^2 = \frac{1}{M-N}\sum_{j=1}^{N} n_j \pi_j(1-\pi_j) = \frac{312.7729}{3498-11} = \mathbf{0.0896968454258675}$$

Tables in this study

Table 4.2 Pearson Chi-square test of association

Chi-Square Tests				
Variables	Statistics	Value	df	P-value
Region	Pearson Chi-Square	101.88	10	0.000**
Residence	Pearson Chi-Square	75.89	1	0.000**
Religion	Pearson Chi-Square	71.499	2	0.000**
Education Level	Pearson Chi-Square	128.566	3	0.000**
Wealth Index	Pearson Chi-Square	128.541	2	0.000**
Knowledge of Ovulatory Cycle	Pearson Chi-Square	62.203	5	0.000**
Contraceptive Method used	Pearson Chi-Square	234.179	2	0.000**
Working Status	Pearson Chi-Square	2.305	1	0.129
Ethnicity	Pearson Chi-Square	109.831	15	0.000**
Sex of House Hold	Pearson Chi-Square	0.594	1	0.441
Media Exposure	Pearson Chi-Square	64.684	1	0.000**

Table4.3 Results of Univariable Analysis

FertilityStatus |	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]
Region |					

Tigray |	4.780933	1.903432	3.93	0.000	[2.190907	10.43281]
Affar |	13.81518	5.355583	6.77	0.000	[6.462198	29.53471]
Amhara |	4.015155	1.657384	3.37	0.001	[1.787894	9.017019]
Oromia |	8.438556	3.243039	5.55	0.000	[3.973206	17.92236]
Somali |	9.618528	3.743455	5.82	0.000	[4.485674	20.62479]
Benishangul |	6.250006	2.592627	4.42	0.000	[2.771948	14.0921]
SNNPR |	3.931323	1.607247	3.35	0.001	[1.764172	8.760651]
Gambela |	9.770959	3.947857	5.64	0.000	[4.426065	21.57032]
Harari |	8.777079	3.642113	5.23	0.000	[3.891685	19.79531]
Dire Dawa |	4.0935	1.756179	3.29	0.001	[1.765693	9.490178]
_cons |	0188679	.0067335	-11.13	0.000	[.0093746	.0379747]
Residence |						
Rural |	3.53951	.542458	8.25	0.000	[2.621137	4.779655]
_cons |	.0444258	.0062364	-22.18	0.000	[.03374	.058496]
Religion |						
Others |	2.231819	.3857219	4.65	0.000	[1.59055	3.131633]
Muslim |	3.144546	.4422057	8.15	0.000	[2.387023	4.142469]
_cons |	.0554307	.0066199	-24.22	0.000	[.0438626	.0700497]
Edu_Level |						
Primary |	.3613079	.0454563	-8.09	0.000	[.2823498	.4623463]
Secondary |	.1678918	.0329068	-9.10	0.000	[.1143388	.2465273]
Higher |	.0683787	.0493075	-3.72	0.000	[.0166389	.2810068]
_cons |	.2895928	.0290676	-12.35	0.000	[.2378752	.3525545]
WealthEndex |						
poor |	4.117473	.5433326	10.72	0.000	[3.179129	5.332777]
Middle |	2.65908	.4730845	5.50	0.000	[1.876254	3.768522]
_cons |	.0519031	.0056113	-27.36	0.000	[.0419923	.064153]
KnowlOvuCycle |						
During her period |	1.658381	.5001003	1.68	0.093	[.9183249	2.994829]

After period ended	3.233997	.5310182	7.15	0.000	[2.344085 4.461756]
Middle of the cycle	1.38049	.2628084	1.69	0.090	[.9505767 2.004838]
Before period begins	1.662579	.4002367	2.11	0.035	[1.037219 2.664981]
At any time	1.61006	.2831629	2.71	0.007	[1.140619 2.272709]
_cons	.0685225	.0088539	-20.75	0.000	[.0531922 .0882711]
Cont_Method					
Modern Method	7.671563	1.187049	13.17	0.000	[5.664666 10.38947]
_cons	.0908187	.0057303	-38.02	0.000	[.0802542 .1027738]
WorkingStatus					
No	1.225979	.1647605	1.52	0.130	[.9420835 1.595427]
_cons	.097867	.0116108	-19.59	0.000	[.0775623 .1234872]
SexofHH					
F	.9110512	.1101943	-0.77	0.441	[.718766 1.154777]
_cons	.1177579	.007874	-31.99	0.000	[.1032937 .1342475]
Media					
yes	.4026307	.0466923	-7.84	0.000	[3207708 .5053811]
_cons	.1748675	.0124709	-24.45	0.000	[.1520564 .2011007]

Table4.7 Influential cases

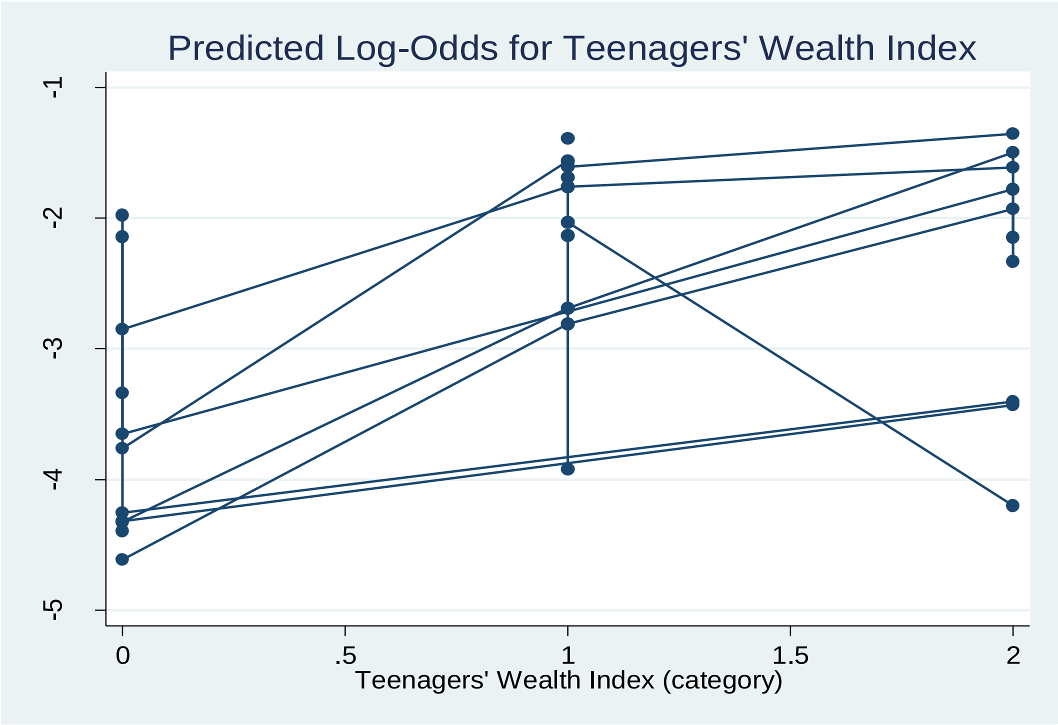
Variables	Cases	Minimum	Maximum
Cook's Influence statistics	3498	0.00003	0.12297
Leverage Value	3498	0.00029	0.00546
DFBeta			
Constant	3498	-0.00073	0.0473
Region			
Tigray	3498	-0.00408	0.05283
Afar	3498	-0.05283	0.02864
Amhara	3498	-0.05283	0.01825
Oromia	3498	-0.05283	0.04011
Somali	3498	-0.05283	0.01759
Benishangul-Gumuz	3498	-0.05283	0.02047
SNNPR	3498	-0.05283	0.04017
Gambela	3498	-0.05283	0.03713

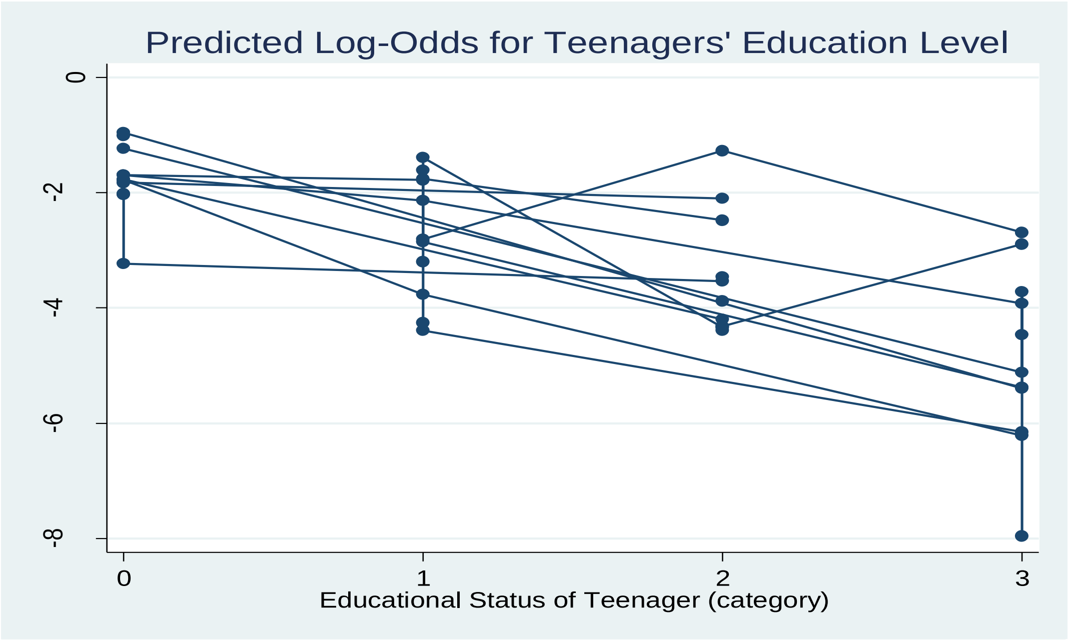
Harari	3498	-0.05283	0.03045
Addis Ababa	3498	-0.05283	0.03867
Dire Dawa	3498	-0.05283	0.12529
Residence			
Urban	3498	-0.00051	0.00327
Rural	3498	-0.00327	0.0188
Religion			
Orthodox	3498	-0.00172	0.01391
Muslim	3498	-0.01391	0.01352
Others	3498	-0.01391	0.0047
Level of Education			
No education	3498	-0.01	-0.5049
Primary	3498	-0.5049	0.01
Secondary	3498	-0.5049	0.01
Higher	3498	-0.5049	0.02706
Wealth Endex			
poor	3498	-0.0058	0.01112
Middle	3498	-0.01112	0.00472
Rich	3498	-0.01112	0.01758
Knowledge of Ovulatory Cycle			
During her period	3498	-0.01564	0.06712
After period ended	3498	-0.01564	0.06712
Middle of the cycle	3498	-0.01564	0.00842
Before period begins	3498	-0.01564	0.01788
At any time	3498	-0.01564	0.03718
Don't know	3498	-0.01564	0.01284
Contraceptive Method used			
No method	3498	-0.00824	0.01182
Traditional method	3498	-0.01182	0.00824
Modern method	3498	-0.66667	0.66667
Working Status			
No	3498	-0.00032	0.00279
Yes	3498	-0.63027	0.062997
Sex of Household			
M	3498	-0.00032	0.00279
F	3498	-0.00279	0.00342
Exposure to any Mass media			
No	3498	-0.00055	0.00782
yes	3498	-0.00782	0.00433

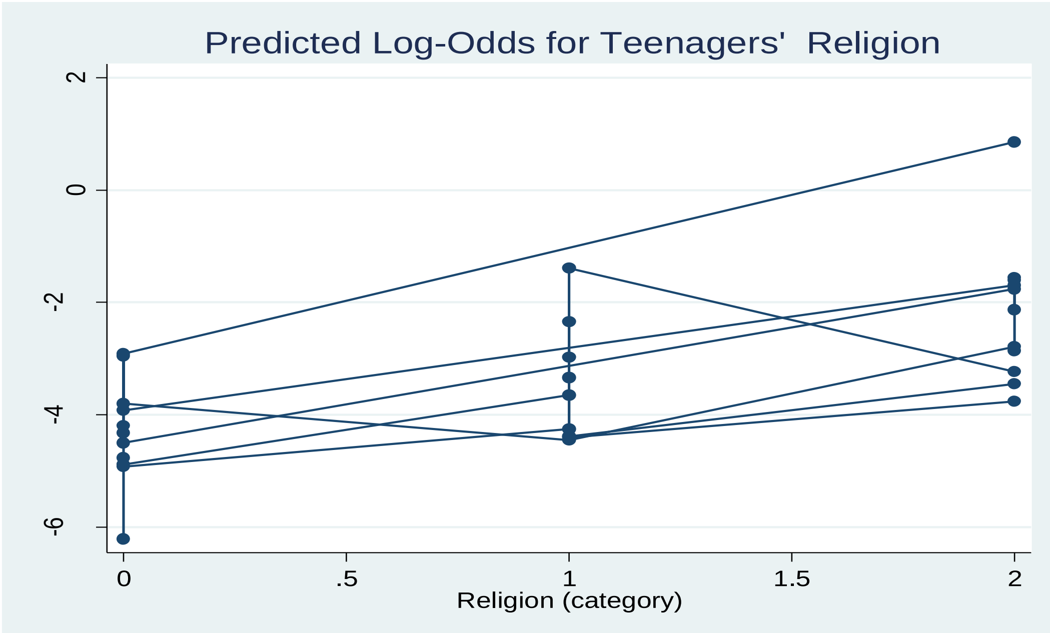
Ethnicity			
Amhara	3498	-0.00581	0.05913
Oromo	3498	-0.05913	0.02441
Tigrie	3498	-0.05913	0.01022
Affar	3498	-0.05913	0.02784
Somalie	3498	-0.05913	0.0186
Guragie	3498	-0.05913	0.02047
Sidama	3498	-0.05913	0.33578
Nuwer	3498	-0.05913	0.14478
WDG	3498	-0.05913	0.11329
Berta	3498	-0.05913	0.10078
Kefficho	3498	-0.05913	0.0924
Gumuz	3498	-0.05913	1.02381
Hadiya	3498	-0.05913	0.25893
Silte	3498	-0.05913	0.34058
Anyiwak	3498	-0.05913	1.02941
Others	3498	-0.05913	0.06784

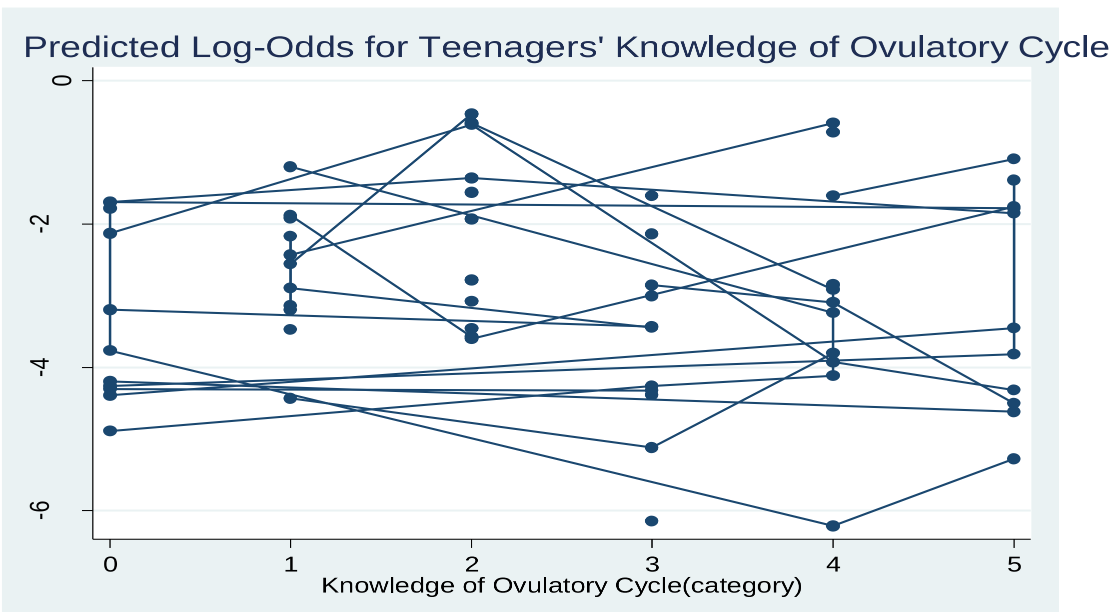
Appendix B

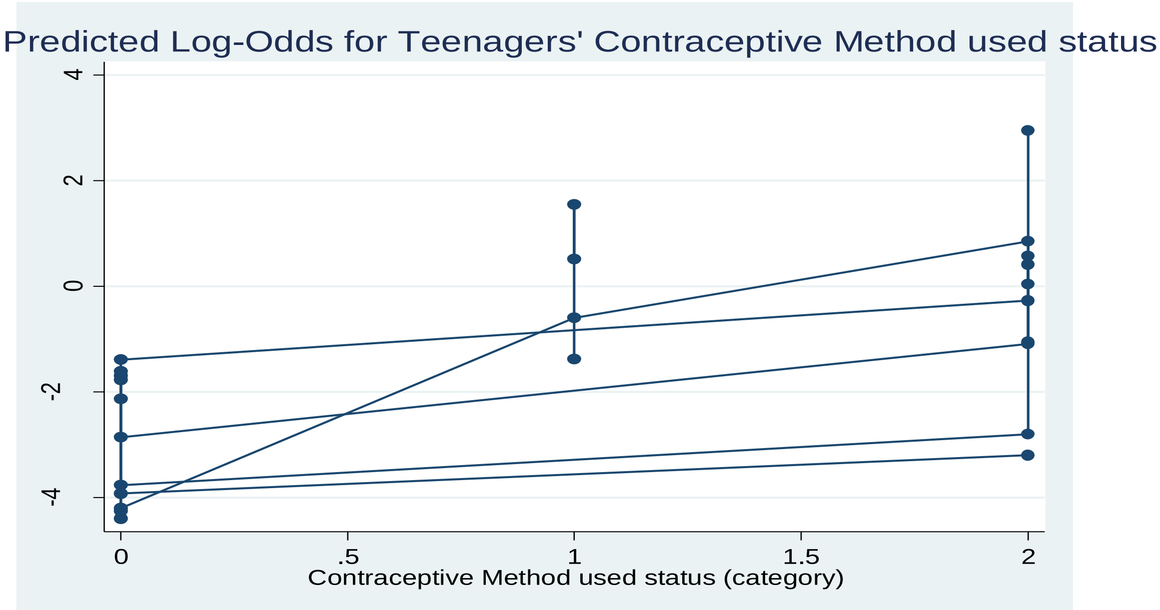
Tagging Indicators with Region





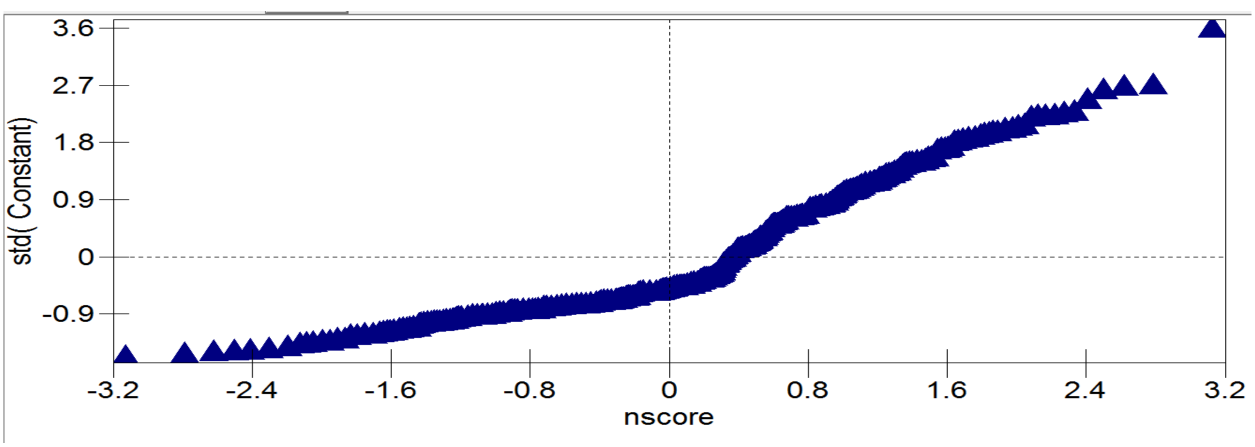

Predicted Log-Odds for Teenagers' Religion
2
0
-2
-4
-6
0
.5
1
1.5
2
Religion (category)


Predicted Log-Odds for Teenagers' Knowledge of Ovulatory Cycle
0
-2
-4
-6
0
1
2
3
4
5
Knowledge of Ovulatory Cycle(category)

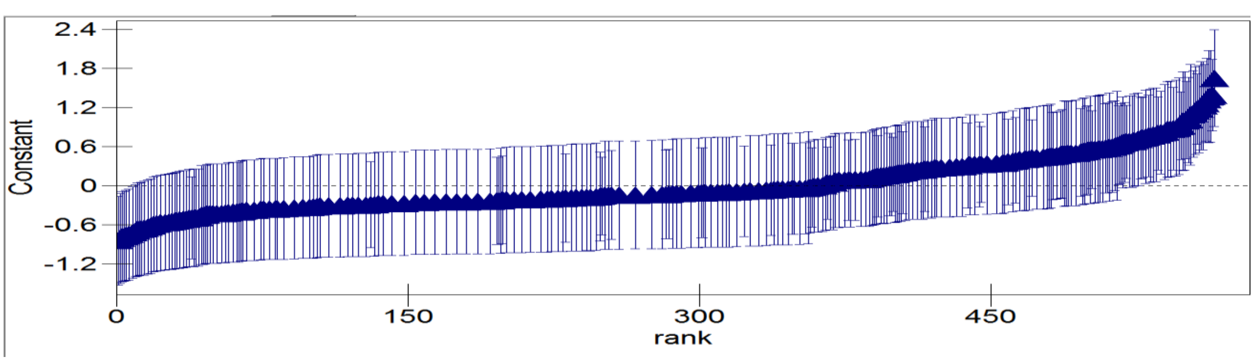


Graphs for model Diagnostics(From MLwiN)

Standard Residuals cross Normal scores for Constant



The Plot with the Region Effects with 95% Confidence Intervals



Model Diagnostics