

**ADDIS ABABA UNIVERSITY
GRADUATE STUDIES PROGRAM
COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE
DEPARTMENT OF STATISTICS**



**Utilization of Reproductive Health Services among Undergraduate
Graduating Class Students in Selected Public Universities of Ethiopia:
Classical and Bayesian approaches**

Yebelay Berelie

Advisor: Dejen Tesfaw (PhD)

**A Thesis Submitted to the Graduate Programs of Addis Ababa University in
Partial Fulfillment of the Requirements for the Degree of Master of Science in
Statistics (Bio-Statistics)**

**June, 2017
Addis Ababa University
Addis Ababa, Ethiopia**

**ADDIS ABABA UNIVERSITY
GRADUATE STUDIES PROGRAM
COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE
DEPARTMENT OF STATISTICS**

**Utilization of Reproductive Health Services among Undergraduate
Graduating Class Students in Selected Public Universities of Ethiopia:
Classical and Bayesian approaches**

Yebelay Berelie

Advisor: Dejen Tesfaw (PhD)

**A Thesis Submitted to the Graduate Programs of Addis Ababa
University in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Statistics (Bio-Statistics)**

Addis Ababa University
Addis Ababa, Ethiopia

June, 2017

DECLARATION

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

Name: Yebelay Berelie

Signature: _____

Date: _____

Place of submission: Department of Statistics, College of Natural Sciences, Addis Ababa University.

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

Advisor's Name: Dejen Tesfaw (PhD)

Signature: _____

Date: _____

Addis Ababa University
School of Graduate Studies

This is to certify that the thesis prepared by Yebelay Berelie, entitled: Utilization of Reproductive Health Services among Undergraduate Graduating Class Students in Selected Public Universities of Ethiopia: Classical and Bayesian approaches and submitted in partial fulfillment of the requirements for the Degree of Master of Science in Statistics (Biostatistics) complies with requirements of the university and meets the accepted standards with respect to originality and quality.

Signed by the examining committee:

Name	Signature	Date
Advisor: Dejen Tesfaw (PhD)	_____	_____
Examiner: Prof. Eshetu Wencheke	_____	_____
Examiner: Birhanu Teshome (PhD)	_____	_____

Contents

ACKNOWLEDGEMENTS	VII
ABSTRACT	VIII
LIST OF TABLES	IX
LIST OF FIGURES	IX
ACRONYMS AND ABBREVIATIONS	X
1. INTRODUCTION	1
1.1 BACKGROUND OF THE STUDY	1
1.2 STATEMENT OF THE PROBLEM	2
1.3 OBJECTIVE OF THE STUDY	3
1.4 SIGNIFICANCE OF THE STUDY	3
2. LITERATURE REVIEW	4
3. DATA AND METHODOLOGY	7
3.1 SOURCE OF DATA	7
3.2 VARIABLE DESCRIPTION	8
3.3 METHODS OF ANALYSIS	9
3.3.1 <i>Introduction to Logistic Regression model</i>	9
3.3.2 <i>The Multiple Logistic Regression Model</i>	10
3.3.3 <i>Maximum Likelihood (ML) Estimation of the Parameters</i>	11
3.3.4 <i>Goodness-of-fit of the Model</i>	12
3.3.5 <i>Statistical tests of individual parameters</i>	13
3.4 MULTILEVEL LOGISTIC REGRESSION MODEL	14
3.4.1 <i>A Two Level Logistic Regression Model</i>	14
3.5 BAYESIAN LOGISTIC REGRESSION MODEL	20
3.5.1 <i>Bayesian Logistic Regression Parameters</i>	20
3.5.2 <i>Likelihood Function</i>	20
3.5.3 <i>Prior distribution</i>	21
3.5.4 <i>Posterior Distribution</i>	22
3.5.5 <i>Markov Chain Monte Carlo Methods</i>	23
3.5.6 <i>Gibbs Sampling</i>	23
3.5.7 <i>Test of Convergence of the algorithm</i>	24
3.5.8 <i>Assessing Accuracy of the Bayesian Logistic Regression model</i>	25
4. RESULTS AND DISCUSSIONS	26
4.1 DESCRIPTIVE ANALYSIS	26

4.2. BINARY LOGISTIC REGRESSION ANALYSIS	28
4.2.1 <i>Model adequacy checking</i>	28
4.2.2 <i>Interpretation of the results from the Binary Logistic Regression model</i>	30
4.3 RESULTS OF MULTILEVEL BINARY LOGISTIC REGRESSION MODEL.....	32
4.3.1 <i>Null model with random intercept</i>	33
4.3.2 <i>Results of Random Intercept Multilevel Analysis</i>	34
4.3.3 <i>Results for Random intercept with random coefficients model</i>	36
4.4 BAYESIAN BINARY LOGISTIC REGRESSION ANALYSES.....	38
4.4.1 <i>Assessment of Model Convergence</i>	38
4.4.2 <i>Assessing Accuracy of the Bayesian Logistic Regression model</i>	41
5. DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS	45
5.1 DISCUSSION.....	45
5.2 CONCLUSIONS	47
5.3 RECOMMENDATIONS	48
REFERENCE.....	49
APPENDICES.....	52

Acknowledgements

First of all, I would like to acknowledge my thesis advisor Dr. Dejen Tesfaw for spending his precious time to direct me. He always gave me a clear direction to follow and also taught me with his generous patience. Under his guidance, I may feel a lot easier to finish this thesis. In addition, I am grateful to thank for giving me the opportunity to work on his 3rd round thematic research project of Addis Ababa University, Office of Vice President for Research and Technology Transfer (VPRTT), on utilization of youth friendly reproductive health services.

Thirdly, my deepest thanks go to all members at Addis Ababa University especially, my instructors and all individuals in the Department of Statistics for their moral support and positive attitude when I needed them during my course work, and research period.

Furthermore, I would like to appreciate all the thematic research team members; Muluken Derebew, Teklehamanot Mezgebe, Tadesse Kassahun, Anteneh Tesema, Abrham Keralem and G/Mariam B/Mariam. And also, I would like to appreciate my family and my friends who had ever given me a hand.

I would also like to acknowledge that my work here at Addis Ababa University was made possible by Debre Markos University.

Thanks to God for his uncountable . . . !!

Abstract

Utilization of Reproductive Health Services among Undergraduate Graduating Class Students in Selected Public Universities of Ethiopia: Classical and Bayesian approaches

Young people and youth usually refer to those between the ages of 10 and 24 (Bearinger et al 2007). Reproductive health service is health facility-based sexual and reproductive health services provided to the adolescents. Young people face greater reproductive health (RH) risks than adults, yet they are less willing and able to access RH services. The objective of this study was to identify the factors that affect utilization of students' for reproductive health service methods among public university students in Ethiopia. A total of 2,384 students' utilization was considered in this study. Descriptive analysis, single level and multilevel logistic regression, and Bayesian logistic regression were used for data analysis using socio-economic, demographic, and health related variables and utilization of reproductive health service as the dependent variable. The results of the study showed that, out of a total of 2,384 sampled students 1,227 (representing 51.5%) used the reproductive health services while 1,157(representing 48.5%) did not. The single level and multilevel and Bayesian logistic regression analyses revealed that the variables that affect the students' utilization of RHS in public universities of Ethiopia were age, awareness of students towards RHS, convenient time to go to health service centers, students monthly income, parents' monthly income, service fee, parents' occupation, place of residence, and gender. The multilevel logistic regression analysis revealed that there was significant variation with regard to students' utilization of RHS across the departments under investigation. The results obtained by applying Bayesian logistic regression analysis show that the standard errors for the variables incorporated in the model were smaller than the classical binary logistic regression analysis. This implies that the Bayesian logistic regression model may give a better accurate estimation than the classical approach.

Keywords: Bayesian analysis, multilevel analysis, reproductive health services, Utilization of university students'

List of tables

Table 3.1: Socio-economic, demographic and health related variables.....	8
Table 4.1: Frequency distribution of utilization of RHS.....	26
Table 4.2: Cross tabulation of RHS with predictor variables.....	27
Table 4.3: Hosmer-Lemeshow Test.....	28
Table 4.4: Summary Statistics of the Likelihood Ratio Test and Akaike and Schwarz Information Criteria.....	29
Table 4.5: Results of the binary logistic regression model	30
Table 4.6: Summary results of model selection criteria	33
Table 4.7: Results of the null model	33
Table 4.8: Results of the random intercept multilevel logistic regression model	35
Table 4.9: Results of random coefficient multilevel binary logistic regression model.....	36
Table 4.10: Results of comparison of MC error with 5% of s.d.....	42
Table 4.11: Posterior summaries of parameters in Bayesian Logistic Regression Model.....	43

List of figures

Figure 4.1: Time series plots for convergence of coefficients for age, awareness, convenient time, service fees.....	39
Figure 4.2: Gelman-Rubin statistics for convergence of coefficients for age, awareness, family income and place of residence.....	40
Figure 4.3: Density plots for Convergence of Coefficients of age, awareness, convenient time and service fees.....	40
Figure 4.4: Convergence of autocorrelation plots for coefficients awareness, convenient time, gender and family income.....	41

Acronyms and Abbreviations

AIDS	Acquired Immune Deficiency Syndrome
FHI	Family Health International
FMOH	Federal ministry of health
HIV	Human Immune-Deficiency Virus
IPPF	International Planned Parenthood Federation
MCMC	Markov Chain Monte Carlo
OLS	Ordinary least square
RH	Reproductive health
RHS	Reproductive health service
SRHS	Sexual and reproductive health service
STD	Sexual transmitted diseases
STI	Sexual transmitted infection
YFRHS	Youth friendly reproductive health service
WHO	World health organization
WinBUGS	Windows for Bayesian inference Using Gibbs Sampling

1. Introduction

1.1 Background of the study

Young people and youth usually refer to those between the ages of 10 and 24 (Bearinger et al 2007) with adolescents comprising the 10 to 19 age group and young adults the 20 to 24 range. Reproductive health service (RHS) is health facility-based sexual and reproductive health services provided to adolescents or the youths in a youth-friendly manner. That means it is accessible, acceptable and appropriate for the youth, in the right place at the right price and delivered in the right style to be acceptable to young people and is effective, safe and affordable.

The friendliness of reproductive health services for the youth are because of the specific biological and psychological needs of the youth, the high risks of STIs, HIV, pregnancy, and excessively high risk of sexual abuse (Senderowitz, 2003). These services include provision of educational services, contraceptives, STI treatment, post-abortion and antenatal services (IPPF, 2007).

Ethiopian government accepts youth friendly health services and health extension program (HEP) to strengthen the delivery of preventive, primitive and basic health care in the rural area to reach adolescents and youth at the community level. However, youth couldn't properly utilize the available health service (MoH, 2006).

Several reproductive health challenges face youth globally and more distributive in developing countries where services and facilities are absent. Young people face greater reproductive health risks than adults, yet they are less willing and able to access RH services. Lack of awareness, inadequate information, and significant barriers posed by the current state of most RH services are perceived as unwelcoming to young clients. Given that the consequences of poor reproductive health in adolescence, such as those stemming from a too-early pregnancy or acquiring a STI or HIV, have serious implications for the future as well as the present, clinical programs need to find practical ways to assess their current operations (Senderowitz, 2003).

1.2 Statement of the Problem

Youth and young people are extremely vulnerable to STD/HIV/AIDS, unwanted pregnancy and abortion (FMOH, 2006-2015) because youths often lack access to health information and health care services. Ethiopia is a nation of young people (10-24 years old) constitute about 35% of the total population (world's youth 2013) and a nation whose youths have profound reproductive health needs. Available SRH services are underutilized, despite being offered free to all individuals regardless of age. Underutilization of SRH services by youths leads to low contraceptive prevalence among youths, high teenage pregnancy as well as high prevalence of HIV and sexually transmitted infections among the youth (MOH, 2006). Hence, the SRH of youth should receive special attention because, young people or youths usually have a tendency to experience for physiological vulnerability, high susceptibility to peer pressure, tendency to engage in risk-taking behavior, less ability to negotiate safer sex practices and difficulty accessing reproductive health information and services (Baloyi 2006). In addition, early pregnancy and STIs (including HIV) threaten the health of youth more than at any other age group (Bearinger et al 2007). Investing in the SRH of youth is that changing the behavior of young people provides the greatest opportunity for prevailing against STIs and AIDS. In addition, investing in the health of youth not only improves the health of youth today, but also ensures that the next generation of youth is healthier.

Since most of university students are youths, this study will try to assess factors affecting RH service utilization of students in the university which is very crucial and needs priority for future quality improvement in education, medical, social and developmental perspective.

Furthermore, several studies in Ethiopia on utilization of youth friendly reproductive health services were not using advanced statistical models. Therefore, this study mainly concentrated on classical (single level and multilevel) logistic regression and Bayesian logistic regression models (especially hierarchical data type) to identify significant factors associated utilizing youth friendly reproductive health services.

1.3 Objective of the Study

General objective

The main objective of the study was to identify factors that affect utilization of reproductive health services among selected public university students in Ethiopia by using both classical (single level and multilevel logistic regression model) and Bayesian logistic regression analyses.

Specific objectives:

- To identify socio-economic and demographic determinants of utilization of reproductive health service among selected public university students.
- To select the variables that is related to the risk of not utilizing RHS among selected public university students.
- To examine within and between students and department level differences in determining the students' utilization of reproductive health service.

1.4 Significance of the Study

The result of this study will provide information on determinants of utilization of youth friendly reproductive health services among students on the student's utilization of reproductive health services. A better understanding of the factors influencing RH service utilization is an important way of keeping youth healthy. Therefore, the finding of this study expected:

- To give some knowledge or awareness of factors or risk of factors on the utilization of RH services to university students.
- Provide some information for decision makers to fill gaps of the actual practice on students' utilization of RH services in Ethiopian public universities.
- To generate ideas for reducing students' negative perceptions and attitudes towards use of RH services
- It will also serve as a reference for giving intervention accordingly to conduct further researches and to make strategic recommendations.

2. Literature Review

This chapter discusses the literature available on the reproductive health service of university students. It basically reviews summaries of these literatures with respect to relevant studies on the sexual and reproductive health of youth students. Young people and Youth refer to anyone between the ages of 10 to 24 years of age and “adolescents” refers to people between the ages of 10 and 19 years of age (Bearinger et al 2007) and Youth friendly health services are “those that attract young people, respond to their needs and retain young clients for continuing care” (IPPF 2007).

Youth-Friendly Health Services (YFHS) offered should be based on an understanding of what youths want and need, and must have respect for the realities of young people’s different backgrounds (IPPF 2008). Although YFHS consider all aspects of the health and well-being of young people, particularly concern on SRH, which includes sexual development, reproductive health, interpersonal relationships, friendliness, and gender roles (Baloyi 2006).

Several policy-related barriers exist for young person which places them at a much greater sexual risk. Globally, existing barriers to access and utilization include poor access, availability and acceptability of the services (WHO, 2004). Lack of clear directions and services on offer, crowding, lack of privacy, appointment times that do not accommodate young people’s work and school schedules, little or no accommodation for walk-in patients, and limited services and contraceptive supplies and options calling for referral are also impediments (WHO, 2004). Youths’ utilization of reproductive health services not only differs from one part of the world to other part, but also it varies within a single country. It also noted that utilization of RHS in many countries is low and indeed lags far behind what is expected to be even after decades of investments (Malarcher, 2010).

As WHO’s report, the Russian government identified young people’s reproductive health needs as a priority, health care and education systems are not yet properly equipped to address the youth’s specific reproductive health systems. The youth aged 15-18 year olds are served by pediatricians but health reports show that these young people who had a long relationship with pediatricians are often embarrassed to discuss difficult issues such as

contraception or sexually transmitted infections (STIs) and may also worry about breaches of confidentiality (WHO, 2010).

Most African countries as a follow up to ICPD (1994) have put up youth friendly health services with the combined partnership between the United Nations Population Fund (UNFPA), (Senderowitz, 2003). Despite these efforts, most countries in sub-Saharan Africa, youth still encounter significant obstacles to receive sexual and reproductive health services to obtaining effective, modern contraception and condoms to protect against sexually transmitted infections (STIs), including HIV. A study conducted in Malawi, Ghana, Burkina Faso and Uganda, showed that financial cost, socio-demographic, shame (fear of seen by others) were the most common barriers to youths access reproductive health services (Biddlecom et al 2007).

A study conducted to evaluate factors that discouraged the youth from using youth friendly reproductive services in South Africa found that inconvenient hours or locations, unfriendly staff and lack of privacy were among the reasons for not using RHS (FHI, 2006).

Obonyo (2009) used logistic regression model to identify determinants of utilization of RHS among school and college youth in Kenya. This study showed that age, sex, level of education, knowledge towards RHS had significant effect on utilization of RHS such as family planning, counseling, voluntary counseling and testing for HIV and treatment of sexually transmitted infections.

Tolessa (2014) employed logistic regression model in order to assess the utilization of RHS among college youth in Asela, Ethiopia and found that demographic, economic and facility barriers are the factors of reproductive health service utilization. The study revealed that older youths did not utilize the services more than the younger youths. It was found that the college youth utilization of RHS is low. Therefore, creating awareness about these services is important to enable increase their knowledge and understanding and in turn to scale up their utilization.

A study by Meskerem and Worku (2014) found that among 818 high school students in Bahir Dar town 32% of youths had reported utilization of youth friendly reproductive health services within the past one year of the study period. The major reported factors in utilizing reproductive health by youths were inconvenience service hours (31.8%), feel fear to be seen by parents or other adults (28.5%) and too long waiting hours (28.4%).

A study conducted by Ayalew and Yeshigeta (2009) in Jimma on reproductive health accessibility and utilization by youths indicated that out of 1,082 adolescents, 445(41.1%), and 375(34.7%) of them were ever and current users of RH services, respectively. Thirty four percent (370) of adolescents ever used health services for Information, Education and Communication (IEC) followed by family planning 190 (17.6%). Forty nine (5%) and 34(3.1%) of them used health services for STI treatment and abortion care, respectively. Another study by Amanuel and Assefa (2013) in Machakel district, North West Ethiopia, indicated that 31 (21.5%) of the adolescents ever utilized RH services and 6 (18.8%) have visited an RH services providing center in the last six months.

The economic costs of health care seeking include not only payment for treatment but also loss of productive or school time for the pupil/student, and the travelling expenses. This means that persons of low socio-economic status can have difficulty to afford the costs associated with utilization of healthcare making utilization unlikely unless they are provided with subsidized costs. Poverty has led some school youth to engage in pre-marital sex in exchange for gift or economic support further exposing them to RH risks (CORHA, 2005). A study by Senderowitz et al. (2003) on rapid assessment reproductive health services concluded that youth are unwilling to seek care due to the national laws and policies restricting care based on age and/or marital status, poor understanding of their changing bodies and insufficient awareness of risks associated with early sexual debut, STI/HIV and pregnancy.

A study conducted in Bahir Dar (Meskerem and Worku, 2014) and East Gojam (Amanuel and Assefa, 2013) using logistic regression model and showed that sex, age, school attendance and educational status were found to be statistically significant effect with utilization of RH services. However, marital status, religion, ethnicity, families' educational background, family size, family income and means of communication had no

statistically significant association with RH services utilization. Also, a study conducted in Harar by Motuma et al., (2016) found that using friends, health care providers, and schools as source of information for RHS and having knowledge about RHS were significantly associated with the utilization of YFS.

3. Data and Methodology

3.1 Source of data

The source of data used for the study is secondary which is obtained from the 3rd round thematic research project of Addis Ababa University, Office of Vice President for Research and Technology Transfer (VPRTT), under the supervision of Dr. Dejen Tesfaw on the utilization of youth friendly reproductive health service among public university students in Ethiopia. The data for this study is collected in 2017 academic year.

Sampling Design and Selection Scheme

Ethiopian public universities are classified into three generations based on their establishment time (Dugassa and Alemfrie, 2015). Therefore, the whole Ethiopian public universities were stratified into three non-overlapping subgroup based on year of establishment in order to improve efficiency. Then, considering costs and logistical feasibilities seven targeted universities (i.e. Addis Ababa University, Gondar University, Wolaita Sodo University, Wollo University, Axum University, Ambo University and Wolkite University) were selected by using stratified systematic sampling with probability proportional to size sampling technique; size being number of students in a university. And then, colleges and departments were selected from each of the above public universities, by using power allocation sampling technique. Based on cost and logistic feasibility, the research team suggested four colleges/faculties and three departments in each college/faculty. Finally, at most two sections were taken from each department depending on the size of the class.

3.2 Variable description

The dependent variable

The dependent variable of the study is utilization of youth friendly reproductive health services, which is dichotomous. Thus, the binary response variable for the i^{th} student is represented by a random variable Y with two possible values measuring the utilization of RHS which is coded by 0 and 1.

$$y_i = \begin{cases} 1, & \text{if the } i^{th} \text{ student utilizes RHS} \\ 0, & \text{otherwise} \end{cases}$$

Independent variables

The independent variables considered in the study were grouped into socio-economic, demographic, health and socio-cultural factors and they are believed to influence utilization of youth friendly reproductive health services among undergraduate graduating class students in Ethiopian public universities. The variables with description and their categories are given below.

Table 3.1: Socio-economic, demographic and health related variables

Variables	Descriptions	Values
Age	Age	0 = 18-20 1 = 21-23 2 = >23
Awareness	Awareness of RHS	1 = yes 0 = no
Ctime	Convenient time to visit RHS centers	0 = In the usual health institute working hours 1 = In the hours when other users are not around
Poccup	Parents' occupation	0 = formal employee 1 = farmer 2 = casual laborer 3 = self-employee
Gender	Gender	0 = male 1 = female
Live	Where the student live	0 = on the campus 1 = out of the campus
Income	Students' monthly average income	0 = <250 ETB 1 = 251-500 ETB 2 = 501-1000 ETB 3 = 1001-1500 ETB 4 = >1500 ETB

Residence	Place of residence	0 = urban 1 = rural
Free	Preference of service fees for RHS	0 = At usual rate 1 = With discount 2 = Free of charge
Famincome	Parents' monthly average income	0 = <1000 ETB 1 = 251-500 ETB 2 = 501-1000 ETB 3 = 1001-1500 ETB 4 = >1500 ETB
Religion	Religion	0 = Orthodox 1 = Muslim 2 = Protestant 3 = Catholic 4 = Other

3.3 Methods of Analysis

The single and multilevel classical logistic regression and Bayesian logistic regression model were used to predict a binary dependent variable from a set of independent variables.

3.3.1 Introduction to Logistic Regression model

Regression methods have an integral component of any data analysis concerned with describing the relationship between a response variable and one or more explanatory variables. It is often the case that the outcome variable is categorical, taking on two or more values. When the outcome variable is binary or dichotomous many distribution functions have been proposed for use. Logistic regression model can be used mainly for two reasons. The first is from a mathematical point of view, it is an extremely flexible and easily used function, and the second it leads itself to meaningful interpretation (Hosmer and Lemeshow, 2011).

Odds and Odds Ratio

Odds are the ratio of probability of an event will occur divided by the probability of it will not occur. In this study, the event is that the student i utilizes the youth friendly reproductive health service, $y_i=1$ and given by:

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

where, $P(E)$ is the probability of success.

Odds always have values greater than zero and if odds value is larger than one it means that success will occur more likely than failure. Odds ratio, as the name indicates, is the ratio of two Odds and given as follows:

$$\text{Odds ratio} = \left(\frac{p_1(E)}{1-p_1(E)} \right) / \left(\frac{p_2(E)}{1-p_2(E)} \right)$$

Here, p_1 and p_2 refer to the probability of success in group 1 and group 2 respectively. If the odds ratio value is greater than one indicates that the odds of the outcome in group 1 is larger than in group 2. Thus, subjects in group 1 are more likely to have success than subjects in group 2. In binary logistic regression analysis, odds ratio is the exponential of the estimated coefficient $\hat{\beta}$, ($\exp(\hat{\beta})$).

3.3.2 The Multiple Logistic Regression Model

Logistic regression model is a special type of generalized linear model with many interesting properties. Loosely speaking logistic regression analysis does not require strict assumptions about the distribution of the response variable, although, it is clear that the response has a binary outcomes. This means implicitly that the Bernoulli/Binomial distributions are the natural choices. Therefore, the assumption on distribution of the response is quite evident. Thus, logistic regression model is appropriate to predict the binary dependent variable.

In logistic regression, a single outcome variable y follows a Bernoulli probability function that takes on the value 1 with probability P_i and 0 with probability $1 - P_i$. Then P_i is varies over the observations as an inverse logistic function of a vector x , which includes a constant and k explanatory variable (Efron, B. (1975).

The specific form of the logistic regression model with unknown parameters $\beta_0, \beta_1, \dots, \beta_k$ is

$$P_i = P(y_i = 1|x_i) = \frac{e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}}$$

At times, it is convenient to change the notation slightly by writing $x_0 = 1$, thus the above model becomes

$$P_i = \frac{e^{X_i' \beta}}{1 + e^{X_i' \beta}} = \frac{1}{1 + e^{-X_i' \beta}} \quad (3.1)$$

where, $X_i = (x_0 = 1, x_1, \dots, x_p)'$ and $\beta = \beta_0, \beta_1, \dots, \beta_k$.

The logit link function of P_i and X in equation (3.1) are nonlinear. However, it is possible to form linear relationship between the response and explanatory variables by applying the logit transformation, and is given by

$$\text{Logit}(P_i) = \log\left(\frac{P_i}{1 - P_i}\right). \quad (3.2)$$

Under the above transformation, we can write the regression model (3.1) as

$$\text{logit}(P_i) = X_i' \beta. \quad (3.3)$$

3.3.3 Maximum Likelihood (ML) Estimation of the Parameters

The most commonly used methods of estimating the parameter of a logistic regression model is the method of maximum likelihood (ML) estimation. The method of maximum likelihood estimation yields to estimate values for the unknown parameters which maximize the probability of obtaining the observed set of data.

Suppose we have a sample of n independent observations $(y_i, x_i), i = 1, 2, \dots, n$. where y_i denotes the value of a dichotomous outcome variable and x_i is the value of the explanatory variables for the i^{th} subject and assume $Y_i \sim \text{Bernoulli}(1, P_i)$. To find the ML of $\beta = \beta_0, \beta_1, \dots, \beta_k$ in (3.1), we define the likelihood function as follows

$$\begin{aligned} \mathcal{L} = L(\beta) &= \prod_{i=1}^n [P_i^{y_i} (1 - P_i)^{1-y_i}] \\ &= \prod_{i=1}^n \left[\frac{P_i}{1 - P_i} \right]^{y_i} (1 - P_i) \\ &= \prod_{i=1}^n \frac{e^{y_i x_i' \beta}}{1 + e^{x_i' \beta}}. \end{aligned} \quad (3.4)$$

Taking the natural logarithm of both sides yields the following expression for log likelihood function:

$$\ell = \text{Log}L(\beta) = \sum_{i=1}^n e^{y_i x_i' \beta} - \sum_{i=1}^n \log(1 + e^{x_i' \beta}). \quad (3.5)$$

It can be verified that the first two partial derivatives of the log likelihood function exists and are given as follows:

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^n (y_i - \mu_i) x_{ij}, \text{ where } \mu_i = E(y_i) = P_i.$$

$$\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_k} = - \sum_{i=1}^n P_i (1 - P_i) x_{ij} x_{ik}.$$

Hence, through maximization of equation (3.5) or (3.4) we can theoretically estimate the parameter vector β . But the equation is nonlinear in β and the estimates do not have a closed form expression. Therefore, β will be obtained by maximizing (3.5) using a numerical iterative method (Agresti, 1996). Newton Raphson method is used to obtain the MLE.

3.3.4 Goodness-of-fit of the Model

Measures of goodness of fit are statistical tools used to explore the extent to which the fitted response obtained from the postulated model compares with the observed data. Clearly, the fit is good if there is a good agreement between the fitted and the observed data.

Likelihood-Ratio Test

The likelihood ratio test statistic (LRT) is the most common test for assessment of overall goodness of fit of logistic regression model. The likelihood ratio test is used to test the significance of a number of explanatory variables. This is appropriate for a variety of types of statistical models. The likelihood-ratio test is used to test the ratio of the maximized value of the likelihood function for the full model (L_{ful}) over the maximized value of the likelihood function for the reduced model (L_{red}).

The likelihood-ratio test statistic is given by:

$$\text{LRT} = -2(l_{\text{red}} - l_{\text{ful}}),$$

where, l_{red} and l_{ful} are the log likelihood function of the reduced and full model, respectively (Hosmer and Lemeshow, 2011).

Hosmer-Lemeshow Test

The Hosmer-Lemeshow test statistic evaluates the goodness-of-fit of the model by creating 10 equal groups of subjects and then compares the number actually in each group (observed) to the number predicted by the logistic regression model. The test is similar to a χ^2 test statistic and has the advantage of partitioning the observations into groups of approximately equal size, and therefore, there are less likely to be groups with very low observed and expected frequencies. In this case, better model fit is indicated by a smaller difference in the observed and predicted classification. The Hosmer-Lemeshow test statistic is given by:

$$\hat{C} = \sum_{k=1}^g \frac{(O_k - E_k)^2}{V_k},$$

where, $E_k = np_k$, $V_k = np_k(1 - p_k)$, g is the number of groups, O_k is observed number of events in the k^{th} group. This test statistic has approximately χ^2 distribution with $(g - 2)$ degrees of freedom (Agresti, 1996).

3.3.5 Statistical tests of individual parameters

Wald test

The Wald test is also an alternative test which is commonly used to test the significance of the individual logistic regression coefficients for each independent variable (that is, to test the null hypothesis in logistic regression analysis that a particular logit (effect) coefficient is zero i.e. $H_0: \beta_i = 0$ against $\beta_i \neq 0$). The Wald test statistic is:

$$W = \frac{\hat{\beta}_i^2}{\text{var}(\hat{\beta}_i)}$$

For large sample size this test statistic has an approximate chi-square distribution with one degree of freedom (Menard, 2002). Furthermore, likelihood ratio test and score test also used for a significance test of the null hypothesis $H_0: \beta_i = 0$. They all exploit the large sample normality of maximum likelihood estimators. For small to moderate sample sizes, the likelihood-ratio test is usually more reliable than the Wald test (Agresti, 1996).

3.4 Multilevel Logistic Regression Model

Multilevel logistic regression model is appropriate for research designs where data for respondents are organized more than one level (i.e., nested data). The units of analysis are usually individuals at a lower level (students) who are nested within contextual/aggregate units at a higher level (departments). A multilevel logistic regression model is also referred to as a hierarchical logistic regression model, or as a random effects (mixed effects) logistic regression model. The multilevel logistic regression extends from single level logistic regression model by including random effects from the model (Snijders and Bosker, 1999).

Multilevel logistic regression analysis can be employed in the simplest case without explanatory variables, (usually called the empty model) and also with explanatory variables by allowing only the intercept term or both the intercept and slopes (regression coefficients) to vary randomly. In this study, multilevel logistic regression model taking into account the data to be analyzed on the case of two-levels. We note that extensions to the case of three or higher levels is straight forward. In this study, students were considered as level 1 and departments were considered as level 2 (Snijders and Bosker, 1999).

3.4.1 A Two Level Logistic Regression Model

Multilevel analysis is a methodology for the analysis of data with complex patterns of variability, with a focus on nested sources of variability. The best way to analysis multilevel data is an approach that represents within-group as well as between group relations within a single analysis, where 'group' refers to the units at the higher levels of the nesting hierarchy. Very often it makes sense to use probability models to represent the variability within and between groups, in other words, to conceive of the unexplained

variation within groups and the unexplained variation between groups as random variability. For example, a study of students' within departments means that not only unexplained variation between students', but also unexplained variation between departments' is regarded as random variable. This can be expressed by statistical models called random coefficient model. Multilevel analysis is an approach to the analysis of such data including the statistical techniques as well as the methodology of how to use these (Snijders and Bosker, 1999).

3.4.1.1 Testing heterogeneous proportions

The most commonly used test statistic to check for heterogeneity of proportion between groups (departments) which is proper application of multilevel analysis is chi-square test statistic. To test whether there are indeed systematic differences between the groups (department), the chi-square test can be used and written as:

$$\chi^2 = \sum_{j=1}^g n_j \frac{(\bar{Y}_{.j} - \hat{P}_{.})^2}{\hat{P}_{.}(1 - \hat{P}_{.})} \quad (3.6)$$

where, $\bar{Y}_{.j}$ is group average, obtained as $\bar{Y}_{.j} = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij}$ is the proportion of successes in group j which is an estimate for the group-dependent probability p_j and $\hat{P}_{.}$ is the overall average, i.e. $\hat{P}_{.} = \bar{Y}_{..} = \frac{1}{n} \sum_{j=1}^g \sum_{i=1}^{n_j} Y_{ij}$ is the overall proportion of successes. The decision is based on chi-square distribution with $g-1$ degrees of freedom (Agresti, 1996).

3.4.1.2 Estimation of between and within-group variance

Consider a population having two-levels, the basic data structure of two-level logistic regression analysis is a collection of N groups (units at level-two (departments)) and within group j ($j= 1, 2, \dots, N$) a random sample of n_j level-one units (students). The outcome variable is dichotomous and denoted by Y_{ij} , ($i = 1, 2, \dots, n_j, j = 1, 2, \dots, N$) for level-one unit i in group j . The total sample size is $M = \sum_{j=1}^N n_j$.

Then, the theoretical variance between the groups (departments) dependent probabilities, i.e., the population value of $\text{Var}(P_j)$, can be estimated by:

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

where, $\tilde{n} = \frac{1}{N-1} \left(M - \frac{\sum_{j=1}^N (n_j^2)}{M} \right) = \bar{n} - \frac{S^2(n_j)}{N\bar{n}}$

For dichotomous dependent variable, the observed between- groups variance is closely related to the Chi-squared test statistic (Snijders and Bosker, 1999). They are given by the formula:

$$S^2_{\text{between}} = \frac{\hat{P}(1 - \hat{P})}{\tilde{n}(N - 1)} \chi^2,$$

where, χ^2 is as given by equation (3.6), and the within- group variance in the dichotomous case is a function of the group:

$$S^2_{\text{within}} = \frac{1}{M - N} \sum n_j P_j (1 - P_j)$$

i. The Empty Logistic Regression Model

The empty level-2 model for a dichotomous outcome variable refers to a population of groups (level-two units, i.e. departments) and specifies the probability distribution for group-dependent probabilities P_j without taking further explanatory variables into account. This model only contains random groups and random variation within groups. It can be expressed with logit link function as follows.

$$\text{logit}(P_j) = \beta_0 + u_{0j}, \tag{3.7}$$

where, β_0 is the average of the outcome variable (intercept) of the transformed probabilities and u_{0j} the random deviation from this average for group j . For the deviations u_{0j} is assumed to be independent random variables with a normal distribution with mean 0 and variance σ_0^2 i.e. $u_{0j} \sim N(0, \sigma_0^2)$.

ii. The Random Intercept Logistic Regression Model

In the random intercept logistic regression model the intercept is the only random effect meaning that the groups differ with respect to the average value of the response variable. It represents the heterogeneity between groups in the overall response.

The logistic random intercept model expresses the log-odds, i.e. the logit of P_{ij} , as a sum of a linear function of the explanatory variables and the random part of the model. That is,

$$\begin{aligned}\text{logit}(P_{ij}) &= \log\left(\frac{P_{ij}}{1 - P_{ij}}\right) = \beta_{0j} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \cdots + \beta_k x_{kij} \\ &= \beta_{0j} + \sum_{k=1}^K \beta_k x_{kij}\end{aligned}$$

where, the intercept term β_{0j} is assumed to vary randomly and is given by the sum of an average intercept β_0 and group- dependent deviations u_{0j} . That is:

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

As a result

$$\text{logit}(P_{ij}) = \log\left(\frac{P_{ij}}{1 - P_{ij}}\right) = \beta_0 + \sum_{k=1}^K \beta_k x_{kij} + u_{0j}, \quad (3.8)$$

where, β_0 is the log-odds that $y = 1$ when $x = 0$ and $u = 0$, β_k is effect on log-odds of one unit increase in X for individuals in same group (same value of u), $\exp(\beta_k)$ is an odds ratio, comparing odds for individuals spaced 1-unit apart on X but in the same group (department). u_{0j} is the effect of being in group j on the log-odds that $y = 1$ also known as a level 2 residual, σ_0^2 is the level 2 (residual) variance, or the between-group variance.

Note that the first part of the left-hand side of (3.8), incorporating the regression coefficients, $\beta_0 + \sum_{k=1}^K \beta_k x_{kij}$ 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 that the residual, u_{0j} are mutually independent and normally distributed with mean zero and

variance σ_0^2 . Equation (3.8) is considered as a mixed model because it has both fixed effects and random effects (Snijders and Bosker, 1999).

iii. The Random Coefficients Logistic Regression Model

In the random intercept logistic regression model the intercept is the only random effect meaning that the groups differ with respect to the average value of the dependent variable. But we have assumed that the effects of the explanatory variables are the same for each department. This assumption is considered by allowing the difference between explanatory variables within a department to vary across departments. To allow for this effect, we will need to use a random coefficient for those explanatory variables. So, random coefficient model represents heterogeneity in relationship between the response and explanatory variables.

As stated above the response variable in the study, RHS, is binary and the statistical model employed is the two-level random coefficient logistic regression model. The model, with k level-1 predictors and p level-2 predictors, can be expressed as:

$$\text{logit}(P_{ij}) = \log\left(\frac{P_{ij}}{1 - P_{ij}}\right) = \beta_{0j} + \sum_{k=1}^k \beta_k x_{kij} + \sum_{p=1}^p u_{pj} x_{pij},$$

where, $\beta_{0j} = \beta_0 + u_{0j}$, $u_{0j} \sim iid(0, \sigma_0^2)$ and $u_{pj} \sim iid(0, \sigma_p^2)$.

Now the above equation is written as

$$\text{logit}(P_{ij}) = \log\left(\frac{P_{ij}}{1 - P_{ij}}\right) = \beta_0 + \sum_{k=1}^K \beta_k x_{kij} + u_{0j} + \sum_{p=1}^p u_{pj} x_{pij} \quad (3.9)$$

The first part of equation (3.9), $\beta_0 + \sum_{k=1}^K \beta_k x_{kij}$, is called the fixed part of the model and the second part $u_{0j} + \sum_{p=1}^p u_{pj} x_{pij}$ is called the random part (Snijders and Bosker, 1999).

3.4.1.3 Intra-class Correlation Coefficient (ICC)

The intra-class correlation coefficient (ICC) measures the proportion of variance in the outcome explained by the grouping structure. ICC can be calculated using an intercept-only model given in Equation (3.7), which results in the following equation:

$$ICC = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2}$$

where, σ_e^2 is variance of individual (lower) level units.

Since the logistic distribution for the level one residual variance implies a variance of $\pi^2/3 \approx 3.29$ (Snijders and Bosker, 1999) and this formula can be reformulated as:

$$ICC = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + 3.29}$$

3.4.1.4 Parameter Estimation for Multilevel Logistic Regression Model

Parameter estimation for multilevel logistic regression model is not straightforward like the methods for classical logistic regression model. The most common methods to estimate the parameters in multilevel logistic regression model are based on the marginal maximum likelihood (MML) (Bock and Aitkin, 1981). This approach uses a fairly straightforward extension of the usual method of ML, but it is more complicated due to the need to perform numerical integration to obtain a marginal likelihood to maximize. So that approaches based on various approximations to MML are popular. These are referred to as quasi-likelihood approaches, the major variants being marginal quasi-likelihood (MQL) (Goldstein, 1991; Goldstein and Rasbash; 1996) and penalized quasi-likelihood (PQL) (Laird, 1978; Breslow and Clayton, 1993). Both MQL and PQL are based on Taylor series expansion to achieve the approximation of parameter estimation. Based on the first and second term of Taylor expansion, MQL and PQL are often known as first order MQL and second-order MQL, respectively. After applying these quasi-likelihood methods, the model is then estimated using iterative generalized least squares (IGLS) or reweighted IGLS (RIGLS) (Goldstein, 2003)).

3.5 Bayesian Logistic Regression Model

The use of Bayesian inference has become increasingly popular in modern statistical analysis, with applications in numerous scientific fields. Bayesian inferences differ from classical inference by considering parameters as random variable and using the data to update prior knowledge about parameters and functional of those parameters (Congdon, 2005). Thus, prior knowledge about a parameter is an important aspect of the inference process. Bayesian statistics provides much more complete picture of the uncertainty in the estimation of the unknown parameters (Lee, 2010 and Tanner 2011).

The idea of Bayesian statistics is based on Bayes' theorem. Assume that we observed a random variable Y and inferences about other random variable β , where β is drawn from some distribution $P(\beta)$. The purpose is to generate the posterior distribution of the unknown parameters given both the data and some prior density for the unknown parameters.

$$P(\beta/y) = \frac{P(\beta, y)}{P(y)} = \frac{P(y/\beta)P(\beta)}{P(y)} \quad (3.10)$$

where, $P(y/\beta)$ is the likelihood of y under a model and $P(\beta)$ is the prior density, or the density of β before y is observed. This density expresses accumulated knowledge about β , or the degree of uncertainty about β (Congdon, 2005).

3.5.1 Bayesian Logistic Regression Parameters

Bayesian logistic regression procedure is adopted to make inference about the parameters of a logistic regression model i.e. a response variable of interest has only two possible outcomes that can be represented by a binary indicator variable taking values 0 and 1. Bayesian inference for logistic regression model is derived by applying a Markov Chain Monte Carlo (MCMC) algorithm to simulate from the joint posterior distribution of the regression and the link parameters (Congdon, 2005).

3.5.2 Likelihood Function

Frequentist analysis focuses on the likelihood function $P(y/\beta)$ without introducing a prior, whereas Bayesian analysis updates the prior information about β with the

information contained in the data. The joint probability distribution of n independent Bernoulli trials is the product of each Bernoulli densities, where the sum of n independent and identically distributed Bernoulli trials has a binomial distribution. Specifically, let y_1, y_2, \dots, y_n be independent Bernoulli trials with success probabilities P_1, P_2, \dots, P_n that is $P(y_i = 1)$ with probability P_i or $P(y_i = 0)$ with probability $1 - P_i$, for $i = 1, 2, \dots, n$. Since, the trials are independent, the joint distribution of y_1, y_2, \dots, y_n is the product of n Bernoulli probabilities. The probability of success in logistic regression model varies from one subject to another, depending on their covariates (Congdon, 2005). Thus, the likelihood function is illustrated in Section 3.3.3 of Equation (3.4).

3.5.3 Prior distribution

Prior distribution plays an important role and the basis in Bayesian analysis i.e. one of the pre-condition in Bayesian analysis is the choice of a prior distribution. Logistic regression model using Bayesian statistics require the formulation of a set of prior distributions for any unknown parameters. The probability distribution express one's uncertainty about unknown parameters before the data is taken into account. Different types of prior distributions exist, namely informative and non-informative.

Non-informative prior distributions are distributions that have no population basis and play a minimal role in the posterior distribution (Clark et al., 2002 and Mila et al., 2003). It is used when we have very little knowledge or information about the prior distribution. That means the idea behind the use of non-informative prior distributions is to make inferences that are not greatly affected by external information or when external information is not available. On the other hand, informative priors have a stronger influence on the posterior distribution. The influence of the prior distribution on the posterior is related to the sample size of the data and the form of the prior. Since non-informative priors return results very close to classical statistics, for this study non-informative prior distributions were desired.

The most common Bayesian approach to logistic regression model is to impose a univariate Gaussian prior with mean 0 and variance $\sigma^2 > 0$ on each parameter β_j (i.e the

most common choice of priors in logistic regression parameters is normal distribution), and is given as follows

$$Pr(\beta_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{1}{2}\left(\frac{\beta_j - \mu_j}{\sigma_j}\right)^2} \quad (3.11)$$

3.5.4 Posterior Distribution

The posterior distribution is the conditional probability that is assigned after the relevant evidence is taken into account. The prior information is synthesized with the information in the data to produce the posterior distribution, which expresses what we know about the parameters after observing the data. Therefore, the inference of β should be characterized by the joint posterior density of logistic regression model parameters.

The posterior distribution is obtained as the product of the prior distribution of the parameters and the likelihood function. Thus, the posterior distribution is given as follows:

$$\begin{aligned} P(\beta/y) &= P(y/\beta) \times P(\beta) \\ &= \prod_{i=1}^n \left[\frac{P(x_i)}{1 - P(x_i)} \right]^{y_i} (1 - P(x_i)) \times \prod_{j=1}^p \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{1}{2}\left(\frac{\beta_j - \mu_j}{\sigma_j}\right)^2} \end{aligned} \quad (3.12)$$

where, $P(x_i) = \frac{e^{x_i'\beta}}{1 + e^{x_i'\beta}}$ and $P(\beta/y)$ are the posterior distribution which is the product of the likelihood function of the logistic regression model and the normal prior distribution for the parameter β_j . Conditioning upon the observed data, the posterior distribution is used to make statements about β , which is still a random variable. For instance, the mean of the posterior distribution can be used as a point estimate of β . Computing the estimate of β of the posterior distribution may be difficult; to overcome this situation, we need to use non- numerical integration method such as simulation techniques. The most popular and common method of simulation technique is the Markov chain Monte Carlo methods which was used in this study.

3.5.5 Markov Chain Monte Carlo Methods

Bayesian inference is solved by randomly drawing a very large sample from the posterior distribution. The idea of drawing a large sample from the posterior distribution is called Markov Chain Monte Carlo. Using MCMC techniques such as Gibbs sampling or the Metropolis–Hastings algorithm, we can directly sample sequences of values from the posterior distribution of interest, giving up the need for analytic solutions. MCMC methods have transformed Bayesian inference to a practical area of modern statistics. (Geman and Geman, 2009)

3.5.6 Gibbs Sampling

The Gibbs sampler (Geman and Geman, 2009) is widely used MCMC technique and is a special case of Metropolis-Hastings algorithm where the random value is always accepted. The goal of Gibbs sampling is to find estimates for the parameters of interest in order to determine how well the observable data fits the model of interest. To implement the Gibbs sampler one starts with initial guesses of the β_i values such as $\beta_1^{(0)}, \dots, \beta_p^{(0)}$ and then simulates one at a time simultaneously.

Once all of the parameters of interest have been sampled, the nuisance parameters are sampled given the parameters of interest and the observed data.

Gibbs sampling algorithm is especially useful in the binary response models in applications of Bayesian analysis that generates random variables indirectly from univariate distributions without having to calculate the density for which a wide variety of computational tools exist (Gilks et al., 2011). Usually, these conditional distributions have a known form and thus, random numbers is simulated using standard functions in statistical and computing software.

The Gibbs sampling algorithm is defined by sampling the set of full conditional posterior distributions (Gilks et al., 2011) and which is given as follows:

$$f_0(\beta_0|\beta_1, \dots, \beta_p); f_1(\beta_1|\beta_0, \beta_2, \dots, \beta_p); \dots; f_p(\beta_p|\beta_1, \dots, \beta_{p-1}).$$

Gibbs sampler algorithm will be stated as follows:

1. Specify an initial value: $\beta^{(0)} = (\beta_0^{(0)}, \beta_1^{(0)}, \dots, \beta_p^{(0)})$

2. Repeat for $j = 0, 1, \dots, M - 1$

Generate $\beta_0^{(j+1)}$ from, $f_0(\beta_0 | \beta_1^{(j)}, \beta_2^{(j)}, \dots, \beta_p^{(j)})$

Generate $\beta_1^{(j+1)}$ from, $f_1(\beta_1 | \beta_0^{(j+1)}, \beta_2^{(j)}, \dots, \beta_p^{(j)})$

⋮

Generate $\beta_p^{(j+1)}$ from, $f_p(\beta_p | \beta_0^{(j+1)}, \beta_1^{(j+1)}, \dots, \beta_{p-1}^{(j+1)})$

3. Return the values $(\beta^{(1)}, \beta^{(2)}, \dots, \beta^{(M)})$

3.5.7 Test of Convergence of the algorithm

The empirical results from a given MCMC analysis are not viewed as reliable until the chain has reached its stationary distribution. To account this, the term convergence of an MCMC algorithm refers to whether the algorithm has reached its equilibrium (target) distribution. If this is true, then the generated sample comes from the correct target distribution. Hence, monitoring the convergence of the algorithm is essential for producing results from the posterior distribution of interest. Among several convergence assessment methods, basically, the most popular approaches used to determine convergence for Markov chains are discussed below.

I. Autocorrelation: High autocorrelation between the parameters of a chain tends to give slow convergence, whereas high autocorrelation within a single parameter chain leads to slow mixing and possibly individual non-convergence to the limiting distribution because the chain will tend to explore less space with much time. In analyzing Markov chain autocorrelation, it is helpful to identify lags in the series in order to calculate the longer-run trends in correlation, and in particular whether they decrease with increasing lags (Merkle and Trisha (2011)).

II. Time series plots or trace plots: Iteration numbers on x -axis and parameter value on y -axis are commonly used to assess convergence (Merkle and Trisha (2011)). If the plot looks like a horizontal band, with no long upward or down ward trends, then we have evidence that the chain has converged. The posterior distribution is obtained by sampling toward the end of this longer iteration sequence when the posterior distribution is stationary, as determined by an examination of trace plots of the iteration history of selected model quantities.

III. Gelman-Rubin statistic: for a given parameter, this statistic assesses the variability within parallel chains as compared to variability between parallel chains (Merkle and Trisha (2011)). The model is judged to have converged if the ratio of between to within variability is close to one.

IV. Density plot: This is another technique for identifying convergence and a classic sign of non-convergence is multimodality of the density estimate (Merkle and Trisha (2011)).

3.5.8 Assessing Accuracy of the Bayesian Logistic Regression model

Once we are happy that convergence has been achieved, we need to run the simulation for a further number of iterations to obtain samples that can be used for posterior inference. One way to assess the accuracy of the posterior estimates is by calculating the Monte Carlo error for each parameter (Brooks and Gelman, 1998).

As a rule of thumb, the simulation should be run until the Monte Carlo error for each parameter of interest is less than about 5% of the parameter's standard deviation. Running multiple chains with different starting values is also as a way of assessing convergence.

Statistical software

SPSS, SAS, STATA and WinBUGS statistical packages will be used for the analysis part of the study.

4. Results and Discussions

The objective of this Chapter is to describe and make analysis about the effect of major socio-economic, demographic and health characteristics on utilization of youth friendly reproductive health services. In this study, the analysis consists of three sections. In the first section, descriptive analysis is presented using SPSS; in the second section, the results obtained from single level and multilevel binary logistic regression analysis are presented using SPSS, SAS and STATA software. In the third and final section, single level Bayesian logistic regression analysis will be employed to examine factors which influence utilization of youth friendly reproductive health services using WinBUGS software.

4.1 Descriptive Analysis

Descriptive analysis is a process of describing a given set of data by tables, graphs and summary calculations. In this study, the researchers employed cross tabulation to describe both dependent and independent variables.

A total of 2,384 sampled undergraduate graduating class students in selected Ethiopian public universities were included in the study. As shown in Table 4.1, out of a total 2,384 public university students who were included in the study, 1,227 (representing 51.5%) students have used the reproductive health service at least once in their life time, while 1,157 (representing 48.5%) students did not.

Table 4.1: Frequency distribution of utilization of RHS of students in selected Ethiopian public universities

		Count	Percent
Utilization of RHS	Yes	1227	51.5
	No	1157	48.5

Based on Table 4.2, the proportion of utilization of students' reproductive health services with age groups: 18-20 years, 21-23 years, and >23 years were 48.2%, 50.0% and 58.6%, respectively.

The proportions of utilization of youth friendly reproductive health services based on awareness of RHS categories: yes and no, were 58.6% and 34.3%, respectively. As can also be seen from the table, female students (representing 57.9%) used more RHS compared to male students (representing 47.1%). Utilization of RHS was higher for those students who had been living out of campus (54.8%) compared with those who were living on the campus (representing 51.1%).

Table 4.2: Cross tabulation of RHS with predictor variables

Variable	Category	Utilization of RHS		Total	Chi-sqr	Df	P-value
		Yes Count (%)	No Count (%)				
Age	18-20	136(48.2%)	146(51.8%)	282	12.300	2	0.002
	21-23	816(50.0%)	817(50.0%)	1633			
	>23	275(58.6%)	194(41.4%)	469			
Awareness	No	197(34.3%)	378(65.7%)	575	89.824	1	0.000
	Yes	1030(56.9%)	779(43.1%)	1809			
Convenient time to visit RHS centers	In the usual working hours	680(50.1%)	676(49.9%)	1356	2.196	1	0.138
	when other users not around	547(53.2%)	481(46.8%)	1028			
Parents' monthly average income	<1,000 ETB	241(42.5%)	326(57.5%)	567	35.791	4	0.000
	1,000-2,500	520(52.2%)	477(47.8%)	997			
	2,501-5,000	45(56.2%)	35(43.8%)	80			
	5,001-7,500	310(54.1%)	263(45.9%)	573			
	>7,500 ETB	111(66.5%)	56(33.5%)	167			
Preference of service fees for RHS	At usual rate	286(58.0%)	207(42.0%)	493	10.660	2	0.005
	With discount	290(49.7%)	294(50.3%)	584			
	Free of charge	651(49.8%)	656(50.2%)	1307			
Students' monthly average income	<250 ETB	359(47.1%)	403(52.9%)	762	41.099	4	0.000
	251-500 ETB	308(46.5%)	354(53.5%)	662			
	501-1000 ETB	250(53.1%)	221(46.9%)	471			
	1001-1500ETB	140(61.7%)	87(38.3%)	227			
	>1500 ETB	170(64.9%)	92(35.1%)	262			
Parent's occupation	Formal employees	424(50.4%)	418(49.6%)	842	48.018	3	0.000
	Farmer	410(61.3%)	259(38.7%)	669			
	Casual laborer	210(41.2%)	300(58.8%)	510			
	Self-employees	183(50.4%)	180(49.6%)	363			
Place of residence	Urban	591(48.8%)	621(51.2%)	1212	.226	1	0.007
	Rural	636(54.3%)	536(45.7%)	1172			
Religion	Orthodox	767(52.2%)	701(47.8%)	1468	3.588	4	0.465
	Muslim	764(52.1%)	703(47.9%)	1467			
	Protestant	162(48.2%)	174(51.8%)	336			
	Catholic	262(51.9%)	243(48.1%)	505			
	Others	10(40.0%)	15(60.0%)	25			
Gender	Female	558(57.9%)	406(42.1%)	964	26.670	1	0.000
	Male	669(47.1%)	751(52.9%)	1420			
Living style in university	In campus	1083(51.1%)	1038(48.9%)	2121	1.277	1	0.258
	Out of campus	144(54.8%)	119(45.2%)	263			

The chi-square test was carried out to determine the association between the dependent variable (utilization of RHS) and the independent variables (age, convenient time, service fee's, family's income, parent's occupation, student's monthly income, gender, awareness, living style in university, place of residence and religion). The result revealed that all independent variables except living style in university and religion had statistically significant association with youth friendly reproductive health service at 0.25 level of significant.

The chi-square test does not give any information about the strength of the relationship between the variables. Hence, we should identify statistically significant predictor variables and determine the direction of relationship with the dependent and independent variables using single level, multilevel and Bayesian logistic regression analyses.

4.2. Binary Logistic Regression Analysis

Binary logistic regression model was used to analyze the relationships between the students' utilization of RHS in public universities of Ethiopia and each of the independent variables which are incorporated in the model. The analysis was employed by using SAS and the output of the software is interpreted as follows.

4.2.1 Model adequacy checking

Hosmer - Lemeshow Test

Assessing the overall significance of a statistical model is essential in order to get valuable information from the data that we have collected for the research. To achieve these objectives, Hosmer-Lemeshow test and likelihood ratio test were considered.

Table 4.3: Hosmer-Lemeshow Test

Chi-square	Df	p-value
13.5697	8	0.0937

The Hosmer-Lemshow test is a test of assessing goodness of fit of the model. Well-fitting models show non-significance of the Hosmer-Lemshow goodness-of-fit test, indicating model

prediction is not significantly different from observed values. The Hosmer-Lemeshow statistic is used to test the hypothesis:

Ho: the model is a good fit

Ha: the model is not a good fit.

As displayed in Table 4.3, we do not reject the null hypothesis at 5% level of significance. This shows that there is no sufficient evidence to reject the null hypothesis. It indicates that the model is a good fit.

Likelihood Ratio Test

The difference between $-2\log$ -likelihood for model fitted with independent variables and $-2\log$ -likelihood for null model (at step 0, before any variables have been added to the model) is distributed χ^2 with degrees of freedom equal to the difference between the numbers of parameters in the two models.

Table 4.4: Summary Statistics of the Likelihood Ratio Test and Akaike and Schwarz Information Criteria

Mode fit	Model fitting criteria			LRT		
	-2log-likelihood	AIC	SIC	Chi-sqr	Df	p-value
Null model	3,302.870	3,304.870	3,310.647	353.353	19	<.0001
Full model	2,949.517	2,995.517	3,128.377			

As shown in Table 4.4, the $-2\log$ -likelihood value for the null (intercept only) model and final model were 3,302.870 and 2,949.517, respectively. To test the significance of the final model over the null model the likelihood ratio test provided a chi-square value of 353.353 ($p < 0.0001$) which would imply good fit for the final model. The model with the smallest values of AIC and SC is also considered as the best fit. Moreover, the value of AIC and SC for null model was 3,304.870 and 3,310.647, and for full model was 2,995.517 and 3,128.377, respectively. Hence, the final model fits the data well, indicating that the independent variables had significant effect with the dependent variable.

4.2.2 Interpretation of the results from the Binary Logistic Regression model

As can be seen in Table 4.5, all the independent variables incorporated in the binary logistic regression model had statistically significant effect on utilization of reproductive health services.

Table 4.5: Results of the binary logistic regression model

Variables	$\hat{\beta}$	S.E	Wald	p-value	AOR	95% CI for OR	
						Lower	Upper
Intercept	1.323	0.297	19.888	<.001	3.754		
Age							
18-20	-0.480	0.166	8.353	0.004	0.619	0.447	0.857
21-23	-0.410	0.115	12.595	0.004	0.664	0.529	0.832
>23(Ref)							
Aware							
No	-0.816	0.106	59.245	0.001	0.442	0.359	0.544
Yes(Ref)							
Convenient time for RHS							
In the usual working hours when other users not around (Ref)	-0.242	0.091	6.992	0.008	0.785	0.656	0.939
Parents' monthely average income							
<1,000ETB	-0.784	0.201	15.204	0.001	0.457	0.308	0.677
1,000-2,500ETB	-0.456	0.191	5.716	0.017	0.634	0.436	0.921
2,501-5,000ETB	-0.295	0.301	0.962	0.327	0.744	0.413	1.342
5,001-7,500ETB	-0.450	0.198	5.151	0.023	0.638	0.432	0.940
>7,500 ETB(Ref)							
Preference of service fees for RHS							
At usual rate	0.351	0.117	9.069	0.003	1.421	1.130	1.786
With discount	0.051	0.108	0.222	0.637	1.052	0.852	1.300
Free of charge (Ref)							
Gender							
Female	0.594	0.093	40.784	0.001	1.810	1.509	2.172
Male(Ref)							
Students monthly average income							
<250 ETB	-0.732	0.167	19.257	0.001	0.481	0.347	0.667
251-500 ETB	-0.767	0.165	21.537	0.001	0.465	0.336	0.642
501-1000 ETB	-0.350	0.173	4.124	0.042	0.704	0.502	0.988
1001-1500 ETB	0.101	0.202	0.248	0.619	1.106	0.744	1.643
>1500 ETB(Ref)							
Parent's Occupation							
Formal employees	0.134	0.152	0.785	0.376	1.144	0.850	1.539
Farmer	0.480	0.150	10.245	0.001	1.616	1.204	2.168
Casual laborer	-0.308	0.150	4.179	0.041	0.735	0.548	0.987
Self-employees (Ref.)							
Place of residence							
Urban	-0.196	0.089	4.855	0.028	0.822	0.691	0.979
Rural(Ref)							

Ref =reference category

Moreover, the table revealed that, at 5% level of significance, the likelihood of utilization of reproductive health services of students was lowest for age group 18-20 (aOR =0.619; CI: 0.447, 0.857) and for age group 21-23 (aOR= 0.664; CI: 0.529, 0.832) compared to the age group >23 while holding the other variables constant in the model. This means that the utilization of RHS for students whose age group 18-20 and 21-23 were about 38.1% and 33.6%, lower than those whose age group was >23, respectively.

The odds of utilizing the reproductive health service of students who preferred to go to health service centers in the usual working hours was 21.5% lower than those who preferred when other users were not around there (aOR=0.785; CI: 0.656, 0.939). Students who had not awareness on the reproductive health services were 55.8% less likely to utilize youth friendly reproductive health services than those who had (aOR=0.442; CI: 0.359, 0.544).

The adjusted odds ratio for students whose family's income was <1000, 1000-2500, 2501-5000, 5001-7500 ETB as compared to those whose family's income >7500 ETB were aOR= 0.457 (CI: 0.308, 0.677), and aOR= 0.634(CI: 0.436, 0.921), aOR= 0.744 (CI: 0.413, 1.342) and aOR= 0.638 (0.432, 0.940), respectively. This means that students' utilization of RHS increases as students' family' income increases.

Table 4.5 also shows that female students were about 81.0% more likely to use RHS compared to the males (aOR =1.810; CI: 1.509, 2.172).

The adjusted odds ratio for the students whose parent's occupation were formal employment, farmer and casual laborer was aOR= 1.144 (CI: 0.850, 1.539), aOR= 1.616 (CI: 1.204, 2.168) and aOR= 0.738 (CI: 0.548, 0.987), respectively. This implies that the odds of utilizing the RHS of students whose parents' were formal employees and farmers were 14.4% and 61.6% higher than those students whose parents were self-employed. Moreover, students whose parents were casual laborers were 26.2% less likely than those who had self-employed parents to utilize the RHS. The adjusted odds ratio of utilization of RHS for students whose parents' had been living in urban area compared to those whose parents' place of residence were rural was aOR=0.822, (CI: 0.691, 0.979).

4.3 Results of Multilevel Binary Logistic Regression Model

In this study, Multilevel Binary Logistic Regression model was employed in order to compare the existence of variation with regard to utilization of RHS among and within departments' of public universities in Ethiopia. There were 42 departments considered as the second-level units and 2,384 students who were considered as first level units. Three multilevel logistic regression models; namely, empty model with random intercept, random intercept with fixed effects model and random coefficient with random intercept model were used to describe the data at both levels.

Test of Heterogeneity

A Chi-square test was applied to assess heterogeneity in the proportion of students who were using RHS among the 42 departments. The test yield $\chi^2 = 150.676$, P-value= 0.000. Hence, there was evidence for heterogeneity with respect to students' utilization of RHS across departments. As we did in the single level binary logistic regression model, we should assess the overall significance of the three multilevel models before interpreting the coefficients.

Goodness of fit test

The deviance and AIC values were used to select the best fitting model among the three fitted two level logistic regression models. The deviance of the null model is 3,244.138 and random intercept with fixed coefficient model is 2,915.906. These indicate that the random intercept with fixed coefficient model is better than the null model. And also, the deviance of the random coefficient model (2,897.25) shows that the random coefficient model is better than the random intercept with fixed coefficient model.

The values of the Akaike Information Criterion (AIC) were used to make an overall comparison of the three models. The computed AIC value for the random coefficient model, AIC = 2,955.251 was less than that of the fixed slope model with random intercept (AIC = 2,963.906) and the empty model with random intercept (AIC = 2,955.251). This indicated that the random coefficient model was a better fit as compared to the empty model with random intercept and the random intercept and fixed effect model.

Table 4.6: Summary results of model selection criteria

Model selection criteria	Null model	Random intercept model	Random coefficient model
Deviance-based χ^2	58.73 (0.000)	33.61 (0.000)	52.27(0.000)
Deviance	3244.138	2915.906	2897.25
AIC	3248.138	2963.906	2955.251

Moreover, the significant deviance based chi-square value for the empty model, as shown in Table 4.6, implied that a null model with random effect was better than an empty model without random effect. Similarly, the deviance based chi-square test for significance of the random intercept model with the fixed coefficient and random coefficient model indicated that the random intercept model with the fixed coefficient were a better fit compared to the multiple logistic regression model.

4.3.1 Null model with random intercept

We first fitted an empty model with no explanatory variable (intercept-only model) that predicts the probability of students being utilized the youth friendly reproductive health services. The simplest non-trivial specification of the hierarchical linear model is a model in which only the intercept varies between level two units and no independent variables are entered in the model. The empty model contains no explanatory variables and it can be considered as a parametric version of assessing heterogeneity among departments with respect to utilization of RHS.

Table 4.7: Results of the null model

Fixed part	Estimate	S.E	Z-value	p-value	95% C.I. for est.	
					Lower	Upper
$\beta_0 = \text{Intercept}$	0.237	0.0969	2.45	0.014	0.047	0.427
Random effect	Estimate	S.E	Wald type approximation 95% C.I			
Between-department variance($\hat{\sigma}_u^2$)	0.254	0.0917	0.125	0.515		

As shown in Table 4.7, the log-odds of utilizing the reproductive health services given in all departments under investigation, on average, was estimated as $\hat{\beta}_0=0.237$. In addition, the

between-department intra class correlation coefficient (ICC) obtained from the empty model with random effect was $0.254/(0.254+3.29) = 0.072$. This value is called the intra class correlation coefficient (ICC) and interpreted as 7.2% of the variation in the utilization of reproductive health services can be explained by grouping the students with respect to their departments. The remaining value, 92.8%, is the variation in utilization of reproductive health services can be explained within individuals (lower level units).

Random effect tests examine hypothesis that whether or not the random intercept or between-department variance is needed for these data or statistically:

$$H_0: \sigma_{u0}^2 = 0$$

$$H_1: \sigma_{u0}^2 > 0.$$

Because the constrained variance component test lies on the boundary of the parameter space, the likelihood ratio test can break down asymptotically. It has been shown that tests for a single variance component can be carried out using mixtures of chi-square distributions. In this study, we show that the null distribution of this one sided LRT statistic converges to a 50:50 mixture of chi-square distributions with 0 and 1 degree of freedom given as $0.5\chi^2_0 + 0.5\chi^2_1$.

$$P\text{-value} = 0.5Pr(\chi^2_{0:1} > \text{LRT}) = 0.5Pr(\chi^2_0 > 58.73) + 0.5P(\chi^2_1 > 58.73).$$

The likelihood ratio test statistic is equal to twice the difference of the log likelihoods, or

$$\text{LRT} = -2\{l(y|H_0) - l(y|H_1)\},$$

In which $l(y|H_0)$ and $l(y|H_1)$ are the log likelihoods under the null and alternative hypotheses evaluated at their maximum likelihood estimates, respectively. The critical value for an $\alpha = 0:05$ test using this mixture distribution is 2.71, indicating we would reject H_0 for $\text{TLR} > 2.71$. This indicated that adding the random intercept or between-department variance in the model was necessary in order to detect the effect of utilization of the reproductive health services among students.

4.3.2 Results of Random Intercept Multilevel Analysis

To assess the effect of independent variables on utilization of youth friendly reproductive health services, we considered random intercept model. As shown in Table 4.9, the variance component has decreased by 0.032 (the difference of 0.254 and 0.222) as

compared to the null model. So, adding independent variables in the random intercept model had significant contribution in explaining the variation of students' utilization of youth friendly reproductive health services among departments.

Table 4.8: Results of the random intercept multilevel logistic regression model

Estimation of Fixed effect							
Variable	Categories	Estimate	S.E	Z-value	p-value	95% C.I. for est.	
						Lower	Upper
	Intercept	1.5191	0.3201	4.74	0.000	0.892	2.1467
Age	18-20	-0.5386	0.1766	-3.05	0.002	-0.885	-0.193
	21-23	-0.4671	0.1231	-3.79	0.000	-0.7084	-0.226
	>23(Ref)						
Awareness	No	-0.8243	0.110	-7.49	0.000	-1.040	-0.609
	Yes(Ref)						
Convenient time	In the usual working hours	-0.2245	0.0941	-2.39	0.017	-0.4089	-0.040
	when other users not around (Ref)						
Parents monthly average income	<1,000ETB	-0.8217	0.2065	-3.98	0.000	-1.227	-0.417
	1,000-2,500ETB	-0.4704	0.1959	-2.40	0.016	-0.854	-0.086
	2,501-5,000ETB	-0.2661	0.3069	-0.87	0.386	-0.868	0.335
	5,001-7,500 ETB	-0.4877	0.2035	-2.40	0.017	-0.887	-0.089
	>7,500 ETB (Ref)						
Service fee	At usual rate	0.3373	0.1198	2.81	0.005	0.102	0.572
	With discount	0.0197	0.1107	0.18	0.859	-0.197	0.237
	Free of charge(ref)						
Gender	Female	0.5792	0.0969	5.98	0.000	0.389	0.769
	Male (Ref)						
Students' monthly average income	<250 ETB	-0.6734	0.1721	-3.91	0.000	-1.011	-0.336
	251-500 ETB	-0.6946	0.1708	-4.07	0.000	-1.030	-0.360
	501-1000 ETB	-0.2946	0.1778	-1.66	0.097	-0.643	0.0538
	1001-1500 ETB	0.1599	0.2073	0.77	0.440	-0.246	0.567
	>1500 ETB(Ref)						
Parent's Occupation	Formal employees	0.1357	0.1566	0.87	0.386	-0.171	0.443
	Farmer	0.4740	0.1544	3.07	0.002	0.171	0.777
	Casual laborer	-0.2748	0.1543	-1.78	0.075	-0.577	0.028
	Self-employees (Ref)						
Place of residence	Urban	-0.1932	0.0910	-2.12	0.034	-0.372	-0.015
	Rural(Ref)						
Estimation of Random effect							
		Estimate	Standard error	Wald type approximate 95% c.l			
	Between-department variance($\hat{\sigma}_u^2$)	0.2220	0.0921	0.098		0.500	

The table also indicated that fixed parameter estimates of all independent variables were found to be statistically significant, although the values had been slightly changed when compared to the single level logistic regression analysis. Moreover, the estimated coefficients and odds ratio had similar interpretation with that of single level logistic regression analysis just mentioned in Section 4.2.2.

4.3.3 Results for Random intercept with random coefficients model

Multilevel logistic regression analysis allows the coefficients of level-one independent variables to vary across departments. This allows departments to have different slopes; implying that the coefficients of independent variables are random at level two (department-level).

Now we are going to see the effect of awareness, place of residence, convenient time and fee of the service independent variables by allowing them to vary randomly across departments. Estimates of this model show that the estimated variance of random slopes of all included variables is zero except for the variables awareness and convenient time. This indicates that only the effects of awareness and convenient time on utilizing of youth friendly reproductive health services varied across departments whereas the effect of other covariates for youth friendly reproductive health services remain fixed across departments. Results of the random coefficient model are presented in Table 4.9.

Table 4.9: Results of random coefficient multilevel binary logistic regression model

Estimation of Fixed effect							
Variable	Categories	Estimate	S.E	Z-value	P-value	95% C.I. for est.	
						Lower	Upper
	Intercept	1.6193	0.3356	4.83	0.000	0.962	2.277
Age	18-20	-0.5622	0.1795	-3.13	0.002	-0.914	-0.210
	21-23	-0.4604	0.1248	-3.69	0.000	-0.705	-0.216
	>23(Ref)						
Aware	No	-0.7385	0.1327	-5.57	0.000	-0.999	-0.478
	Yes(Ref)						
Convenient time	In the usual working hours	-0.4473	0.1556	-2.87	0.004	-0.752	-0.142
	when other users not around(Ref)						
Parents'	<1,000ETB	-0.8284	0.2094	-3.96	0.000	-1.239	-0.418

monthly average income	1,000-2,500ETB	-0.4897	0.1985	-2.47	0.014	-0.879	-0.101
	2,501-5,000ETB	-0.3175	0.3126	-1.02	0.310	-0.930	0.295
	5,001-7,500 ETB	-0.5105	0.2064	-2.47	0.013	-0.915	-0.106
	>7,500 ETB(Ref)						
Service fee	At usual rate	0.3597	0.1219	2.95	0.003	0.121	0.599
	With discount	0.0488	0.1128	0.43	0.665	-0.172	0.269
	Free of charge(ref)						
Gender	Female	0.5778	0.0983	5.88	0.000	0.385	0.770
	Male(Ref)						
Students monthly average Income	<250 ETB	-0.6700	0.1740	-3.85	0.000	-1.011	-0.329
	251-500 ETB	-0.6825	0.1731	-3.94	0.000	-1.022	-0.343
	501-1000 ETB	-0.2802	0.1803	-1.55	0.120	-0.634	0.073
	1001-1500 ETB	0.1772	0.2111	0.84	0.401	-0.237	0.591
	>1500 ETB(Ref)						
Parents' Occupation	Formal employees	0.1762	0.1603	1.10	0.272	-0.138	0.490
	Farmer	0.5078	0.1580	3.21	0.001	0.198	0.818
	Casual laborer	-0.2646	0.1572	-1.68	0.092	-0.573	0.043
	Self-employees(Ref)						
Place of residence	Urban	-0.1911	0.0921	-2.07	0.038	-0.372	-0.011
	Rural(Ref)						

Estimation of Random effect

	Estimate	Standard error	Wald type approximate 95% c.l	
Between-department variance, $\hat{\sigma}_{0j}^2$	0.1744	0.1044	0.034	0.884
$\hat{\sigma}_{1j}^2 = \text{var}(\text{ctime})$	0.4034	0.1995	0.153	1.063
$\hat{\sigma}_{2j}^2 = \text{var}(\text{aware})$	0.0559	0.0865	0.003	1.159
$\hat{\sigma}_{12j} = \text{Cov}(\text{ctime}, \text{aware})$	-0.1057	0.1092	-0.320	0.108
$\hat{\sigma}_{01j} = \text{cov}(\text{ctime}, \text{_cons})$	0.2223	0.1264	-0.226	0.270
$\hat{\sigma}_{02j} = \text{cov}(\text{aware}, \text{_cons})$	-0.0165	0.0854	-0.184	0.151
$\hat{\sigma}_e^2$	3.29			

In the random effect part, the value of 0.1744, 0.4034 and 0.0559 are the estimated variance of intercept (department), convenient time and awareness of RHS respectively.

The log odds of utilizing the reproductive health services in an average department (with $u_j=0$) is estimated as $\hat{\beta}_0=1.6193$ that is shared by all departments. The log-odds of the probability of utilizing the youth friendly reproductive health services for department j is given by $1.6193+\hat{u}_j$ where the variance of the intercepts across department is estimated as $\text{var}(u_j)=0.1744$, which is

referred to as the between-department variance (or simply the level 2 variance) adjusted for explanatory variables.

The intra class correlation coefficient (ICC) or between department variations in the random coefficient model is decreased by 2% as compared to the empty random intercept model. This value implied that only 5% ($0.1744/(0.1744+3.29) = 0.0503$) of the variation in the use of reproductive health services can be explained by choosing convenient time between departments.

4.4 Bayesian Binary Logistic Regression Analyses

In addition to the classical approach, the Bayesian logistic regression analysis was considered to make inference. Bayesian method gives estimates of parameters by sampling from their posterior distributions using the MCMC method. Hence, we used the Gibbs sampler algorithm to estimate the parameters by approximate the properties of the marginal posterior distributions for each parameter. WinBUGS software was used. We run a simulation with 20,000 iterations, discarding the first 5,000 iterations as burn in. In this study, three different initial values were implemented and we assumed that the regression parameters follow a normal distribution with mean = 0 and precision = $1.0e-3$.

The results of the Bayesian logistic regression revealed that all independent variables were found to have significant effect on students' utilization of reproductive health services. Before we proceed to the results of the estimated parameter or examine the model, we should make sure that the sample was truly representative of the stationary or posterior distribution. In order to do this, various schemes of diagnosis were applied to check the convergence of the Markov chains to the target distribution.

4.4.1 Assessment of Model Convergence

There are several methods to check for convergence. These are Time series plot, Autocorrelation Plot, Density plot, Gelman–Rubin Statistics and comparing the MC error to its posterior standard errors (Ioannis, 2009 and Gelman, 2005).

1. Time series plot

The time series plots in Bayesian simulation are commonly used to assess convergence. Satisfactory convergence is guaranteed when three independently generated chains will mix together (overlap) or if the plot looks like a horizontal band, with no long upward or downward trends. The WinBUGS package gives these plots with iteration number on the X axis and parameter value on the Y-axis. As shown in Figure 4.1, the three independently generated chains demonstrated good chain mixture. The plots showed that the chains with three different colors overlap one over the other which is an indicating of convergence in Bayesian simulation. Hence, the results show that convergence has been achieved.

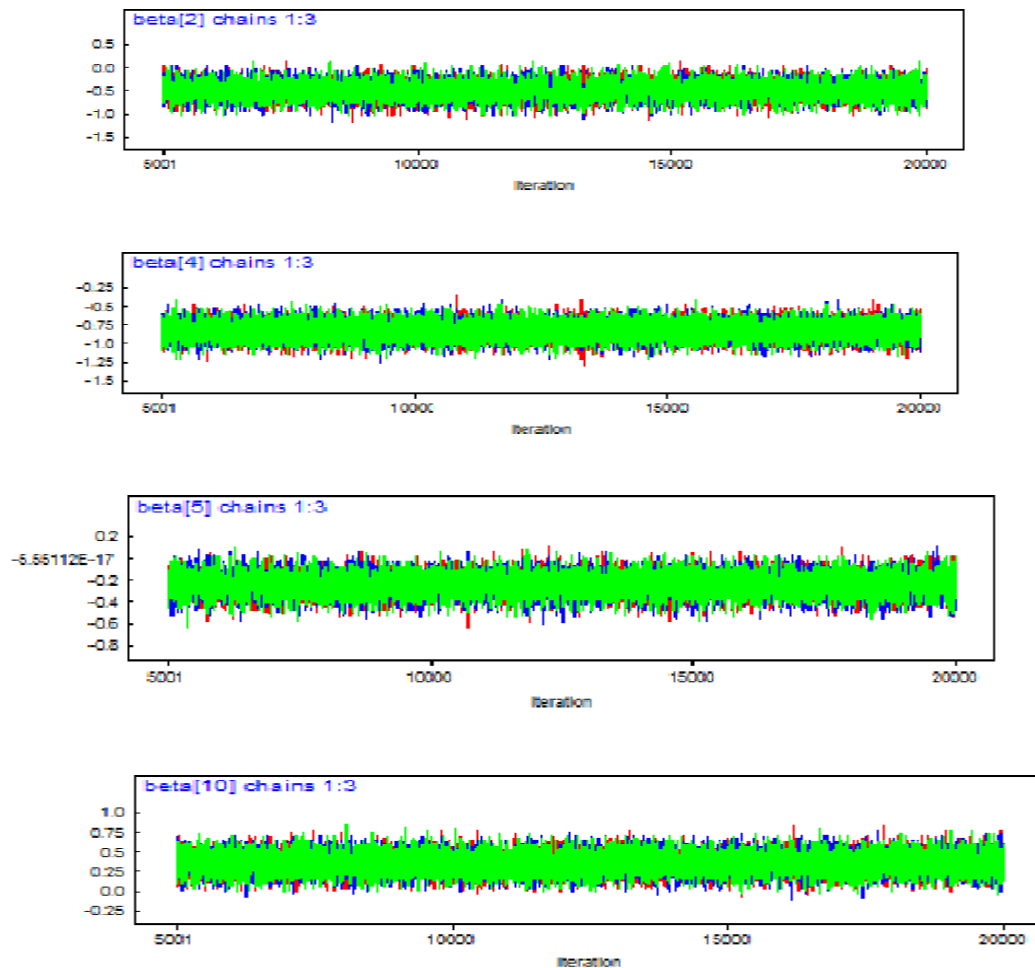


Figure 4.1: Time series plots for convergence of coefficients for age, awareness, convenient time, service fees

2. Gelman-Rubin Statistics

The Gelman-Rubin statistic is used to see the variability within parallel chains as compared to variability between parallel chains. The model is judged to have converged if the ratio of between and within variability is close to 1. The green line represents the between variability, the blue line represents the within variability, and the red line represents the ratio. Evidence for convergence comes from the red line being close to 1 on the y-axis and from the blue and green lines being stable across the width of the plot. Figure 4.2 shows that the red line is close to one, indicating that there is an evidence for convergence.

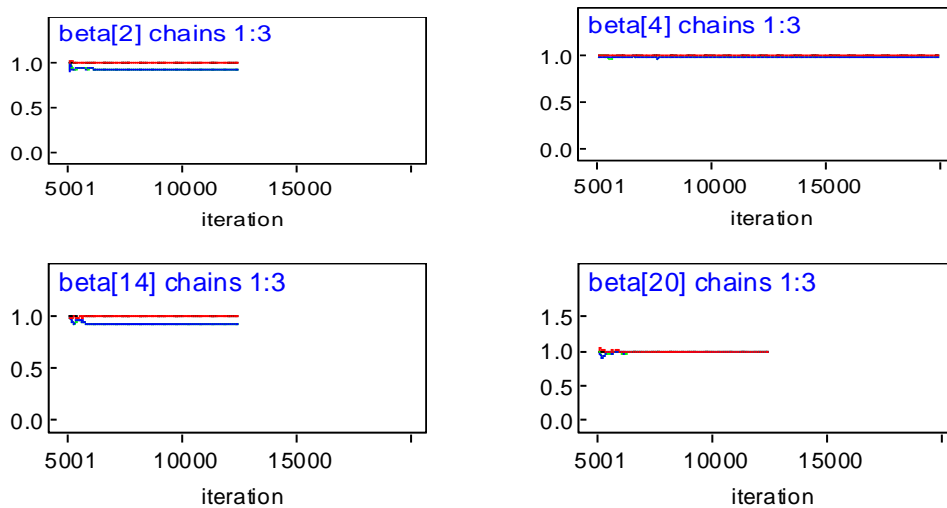
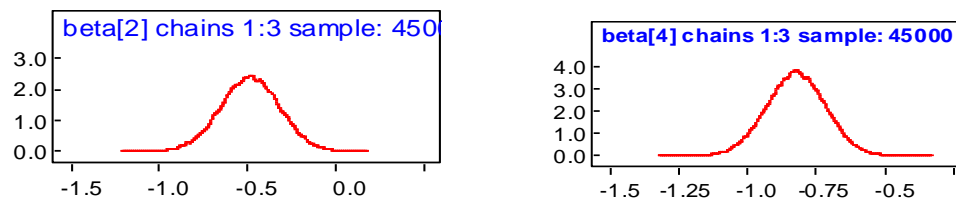


Figure 4.2: Gelman-Rubin statistics for convergence of coefficients for age, awareness, family income and place of residence

3. Density plot

The kernel density which is the graphical method is also used as an alternative method to identify model convergence by assessing whether the Markov chain has converged to its stationary distribution or not. Figure 4.3 reveals that the coefficients of the independent variables were normally distributed and hence the Markov chain has attained to its posterior distribution (the simulated parameter values were converged).



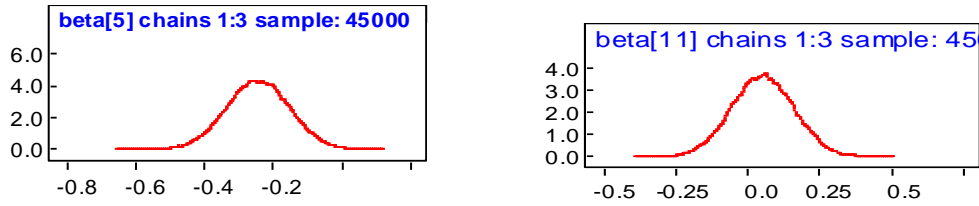


Figure 4.3: Density plots for Convergence of Coefficients of age, awareness, convenient time and service fees

4. Autocorrelation plot

Autocorrelation plot produces lag-autocorrelations for the monitored parameters within each chain. In Markov chain autocorrelation analysis, it is necessary to identify lags in the series in order to calculate the long-run trends in correlation, and in particular whether they decrease with increasing lags. In Figure 4.5, the autocorrelations for all parameters become small after a lag equal to 50 and the three independent chains overlap. So the figure has an evidence of convergence.

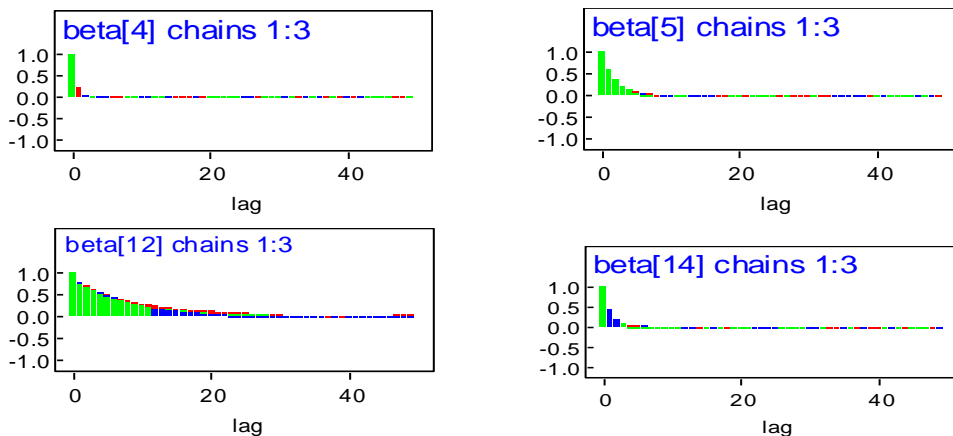


Figure 4.4: Convergence of autocorrelation plots for coefficients awareness, convenient time, gender and family income

4.4.2 Assessing Accuracy of the Bayesian Logistic Regression model

One way to assess the accuracy of the posterior estimates is by calculating the Monte Carlo error for each parameter (Brooks and Gelman, 1998). This is an estimate of the difference between the mean of the sampled values (which we are using as the estimate of the posterior mean for each parameter) and the true posterior mean. As a rule of thumb, to

have accurate posterior estimates the simulation should be run until the Monte Carlo error for each parameter of interest is less than 5% of the parameter's standard deviation. As shown in Table 4.10, MC error for each significant independent variable was less than 5% of its posterior standard error. This implied that convergence and accuracy of posterior estimates has been attained and the model was appropriate to estimate the posterior statistics.

Table 4.10: Results of comparison of MC error with 5% of s.d.

Variables	Categories	Node	MC error	5% of s.d.
	Intercept	beta[1]	0.009342	0.014685
Age	18-20	beta[2]	0.001701	0.00832
	21-23	beta[3]	0.001679	0.00578
	>23(Ref)			
Awareness	No	beta[4]	6.008E-4	0.005325
	Yes(Ref)			
Convenient time	In the usual working hours	beta[5]	7.948E-4	0.00457
	other not around(Ref)			
Parents monthly average income	<1,000ETB	beta[6]	0.004627	0.010125
	1,000-2,500ETB	beta[7]	0.004463	0.009585
	2,501-5,000ETB	beta[8]	0.004544	0.015125
	5,001-7,500 ETB	beta[9]	0.004503	0.009975
	>7,500 ETB(Ref)			
Service fee	At usual rate	beta[10]	7.27E-4	0.005845
	With discount	beta[11]	6.76E-4	0.00542
	Free of charge(ref)			
Gender	Female	beta[12]	6.336E-4	0.004671
	Male(Ref)			
Students monthly income	<250 ETB	beta[13]	0.002674	0.00826
	251-500 ETB	beta[14]	0.002727	0.00821
	501-1000 ETB	beta[15]	0.002812	0.008505
	1001-1500 ETB	beta[16]	0.002851	0.009985
	>1500 ETB(Ref)			
Parent's Occupation	Formal employee	beta[17]	0.002552	0.007625
	Farmer	beta[18]	0.002326	0.00756
	Casual laborer	beta[19]	0.002117	0.00754
	Self-employee(Ref)			
Place of residence	Urban	beta[20]	7.038E-4	0.004487
	Rural(Ref)			

Ref=reference category

As we did in the classical logistic regression, we also used the nine independent variables in Bayesian logistic regression model. The results of all the independent variables were found to have statistically significant association with students' utilization of reproductive health services on public universities in Ethiopia.

Table 4.11: Posterior summaries of parameters in Bayesian Logistic Regression Model

Variables	Estimate	S.d.	Mc error	Odds ratio	95% C.I. for est.	
					Lower	Upper
Intercept	1.351	0.2937	0.00934	3.861	0.7847	1.927
Age						
18-20	-0.4882	0.1664	0.001701	0.614	-0.818	-0.165
21-23	-0.4162	0.1156	0.00168	0.660	-0.644	-0.190
>23(Ref)						
Awareness						
No	-0.8251	0.1065	6.008E-4	0.438	-1.035	-0.618
Yes(Ref)						
Convenient time for RHS						
In the usual working hours when other not around (Ref)	-0.2453	0.0914	7.948E-4	0.782	-0.425	-0.067
Family monthly average Income						
<1,000ETB	-0.8005	0.2025	0.004627	0.449	-1.204	-0.412
1,000-2,500 ETB	-0.4703	0.1917	0.004463	0.625	-0.853	-0.099
2,501-5,000ETB	-0.3068	0.3025	0.004544	0.736	-0.902	0.286
5,001-7,500ETB	-0.4642	0.1995	0.004503	0.629	-0.863	-0.08
>7,500ETB(Ref)						
Service Fee						
At usual rate	0.3555	0.1169	7.27E-4	1.426	0.1285	0.585
With discount	0.0509	0.1084	6.76E-4	1.052	-0.161	0.263
Free of charge (Ref)						
Gender						
Female	0.5988	0.0934	6.336E-4	1.820	0.416	0.783
Male (Ref)						
Students' monthly average income						
<250 ETB	-0.738	0.1652	0.002674	0.478	-1.064	-0.414
251-500 ETB	-0.7738	0.1642	0.002727	0.461	-1.098	-0.455
501-1000 ETB	-0.353	0.1701	0.002812	0.703	-0.686	-0.019
1001-1500 ETB	0.1045	0.1997	0.002851	1.110	-0.287	0.497
>1500 ETB(Ref)						
Parents' Occupation						
Formal employee	0.1336	0.1525	0.002552	1.143	-0.169	0.429
Farmer	0.4825	0.1512	0.002326	1.620	0.186	0.778
Casual laborer	-0.3125	0.1508	0.002117	0.732	-0.609	-0.019
Self-employee (Ref)						
Place of residence						
Urban	-0.1986	0.0897	7.038E-4	0.820	-0.375	-0.022
Rural (Ref)						

Ref = reference category

Table 4.11 shows that the estimated posterior quantities of interest such as posterior means, MC error, together with the estimated certainty or precision of these parameters in terms of posterior standard deviations, credible intervals, or highest posterior density intervals using the samples from the posterior distribution obtained by MCMC.

Here, the 95 percent credible intervals determine which components of estimates are relevant to the model. All selected predictors are significant because their respective 95 percent credible intervals do not contain zero. Since we used Bayesian approach with non-informative priors, the inferences from Bayesian and frequentist are numerically similar. For example, 95% confidence intervals were very similar to the 95% credible intervals.

Bayesian credible intervals are directly interpreted as the probability that the parameter is in the credible interval, given the data and any prior information. Frequentist confidence intervals cannot be interpreted in this way, if the confidence interval procedure were to be used repeatedly, then 95% of all intervals will contain the true value.

5. Discussion, Conclusions and Recommendations

In this Chapter, we present the discussion, conclusions drawn and recommendations made.

5.1 Discussion

This study attempted to determine the socio-economic, demographic, and health related factors of utilization of reproductive health services among students in selected Ethiopian public universities. The results of the study showed that, out of a sample of 2,384 students from selected universities considered, 1,227 (51.5%) used the youth friendly RHS while 1,157 (48.5%) never used the RH services.

The chi-square test was carried out to determine the association between utilization of reproductive health service and individual independent variables. As a result, age, awareness of RHS, convenient time to get the services, family income, service fees, gender, student's income, place of residence (where parents' live) and parent's occupation were significantly associated with the utilization of reproductive health services. On the other hand, living style in university and religion were not significantly associated with students utilization of reproductive health services at 0.25 level of significance. This finding was related with the findings of Abajobir and Seme (2014) and Tolessa (2014).

Both classical (single level and multilevel logistic regression) and Bayesian logistic regression techniques were employed to analyze factors that affect the status of utilization of reproductive health services. The classical and Bayesian logistic regression analyses revealed that age of the student had statistically significant effect on utilization of youth friendly reproductive health services. The finding indicated that the utilization of youth friendly reproductive health services was higher among older students. This result was in line with the findings Obonyo (2009), and Meskerem and Worku (2014) who had conducted research works on utilization of youth friendly reproductive health services. However, our finding is different from the findings of Tolessa (2014) about utilization of RH services.

We found that awareness about RH services and the convenient time to get RH services had a significant effect on the utilization of the reproductive services. This result was in

line with the results obtained by Senderowitz (2003), FHI (2000) and Motuma et al. (2016). Two other determinant factors which affect students' utilization of reproductive health services were family's income and students' income. Students having low income were found to be less likely to utilize the reproductive health services than the students from higher income families. Thus, students of low socio-economic status can have difficulty in affording the costs associated with utilization of reproductive health services (CORHA, 2005). This finding does not agree with other studies (Meskerem and Worku, 2014; Amanuel and Assefa, 2013). Gender of student is also found to be significantly related to utilization of reproductive health services. Our finding agrees with other studies conducted by Katharine (2012), Obonyo (2009), (Meskerm and Worku, 2014) and Amanuel and Assefa, 2013).

Multilevel logistic regression model allows for comparison of variations between departments. Before the analysis of data using multilevel approach, heterogeneity of the status of utilizing of youth friendly reproductive health services with regard to departments was checked first using chi-square test and it was statistically significant. In multilevel logistic regression models with fixed effects of the explanatory variables had similar interpretation as that of the logistic regression model as discussed above whereas the random parts of the intercept and the coefficients provided additional information. Results obtained based on the empty model the overall variance of the constant term suggest that student's status of utilizing the reproductive health service differed across departments. In addition to null model, two other models, one with random intercept and fixed slope model and another with random coefficient model were used. The overall variance constant term in both models was found to be statistically significant implying that utilization of youth friendly reproductive health services differ across departments.

The random coefficient model showed that the random effects of awareness about RHS and convenient time to get the RH services vary across departments in explaining the utilization of youth friendly reproductive health services.

In addition to the multiple logistic regression and multilevel logistic regression analysis, Bayesian logistic regression was carried out to see the effect of predictor variables on the

utilization of youth friendly reproductive health services among university students. The Bayesian logistic regression analysis also revealed that all independent variables (those we have seen in logistic and multilevel logistic regression analysis) are statistically significant. Using Bayesian simulation MC error for each significant predictor was found to be less than 5% of its posterior standard error. This implies convergence and accuracy of posterior estimates of the Bayesian simulation were attained. The estimate of the parameters in Bayesian logistic regression is numerically similar to the estimates obtained in classical logistic regression.

5.2 Conclusions

- The study revealed that awareness about RHS encourages students to use RHS. It is also indicated that the higher income students and their family had, the better was their utilization of RHS.
- The single level binary logistic regression analysis revealed that the independent variables that affect the students' utilization of RHS in public universities of Ethiopia were "age", "awareness of students towards RHS", "convenient time to go to health service centers", "students' monthly income", "parents' monthly income", "service fee", "parents' occupational", "place of residence", and "gender".
- The multilevel logistic regression analysis revealed that there was significant variation with regard to students' utilization of RHS across the departments.
- The results obtained by applying Bayesian logistic regression analysis showed that all selected predictors were significant. Compared to frequentist approaches, in Bayesian logistic regression analysis there is considerable flexibility in which prior evidence about parameters can be incorporated in an analysis.

5.3 Recommendations

Based on the findings of the study, we forward the following recommendations:

- Universities and other concerned bodies in Ethiopia should focus on creating awareness towards students' utilization of RHS. They should also arrange the appropriate/convenient time for students so that they can easily access health services at the right time and with fair price.
- Further studies should be conducted by taking three or four level logistic regression into account to assess the effect of utilization of youth friendly reproductive health services across the college and university levels.
- And also we recommend that sensitivity analysis should be conducted in Bayesian logistic regression which is used for model checking to see whether changes occur within models which results changes in posterior inferences. This may do by comparing models with plausible but different priors or changing the prior variance of the parameter, different sampling distributions, or differences in other aspects of the model such as the inclusion or exclusion of explanatory variables. While an exhaustive check is impossible, at least some effort in this area should be made.
- Finally, further investigations should be conducted on the basis of Bayesian multilevel logistic regression model.

Reference

- Agresti, A. (1996). *An Introduction to Categorical data Analysis*. John Wiley and Sons, Inc., New York.
- Amanuel A. and Assefa S. (2013). Reproductive Health Knowledge and Services Utilization among Rural Adolescents in Machakal district, Northwest Ethiopia. *Asian Journal of Pharmacy, Nursing and Medical Sciences: 01(01); 25-37*
- Ayalew T. and Yeshigeta G. (2009). Accessibility and utilization of adolescent reproductive health services in Jimma city. *Ethiopian Journal of Health Science: 19(2); 91-102*.
- Baloyi, G O. 2006. The evaluation of the National Adolescent Friendly Clinic Initiative (NAFCI) Program in the Greater Tzaneen Sub-District, Limpopo Province: South Africa. University of South Africa thesis.
- Bearinger, L.H., Sieving, R., Ferguson, J. and Sharma, V. (2007) Global Perspectives on the Sexual and Reproductive Health of Adolescents: Patterns, Prevention, and Potential. *Lancet*, **369**, 1220-1231.
- Biddlecom, A, Munthali, A, Singh, S and Woog, V. 2007. Adolescents' views of and preferences for sexual and reproductive health services in Burkina Faso, Ghana, Malawi and Uganda. *African Journal of Reproductive Health* 11(3).
- Bock R. and Aitkin M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika*: 46, 443-459.
- Breslow N. and Clayton D. (1993). Approximate inference in generalized linear mixed models. *Journal of American Statistical Association*: 88, 9-25.
- Brooks S. and Gelman A.(1998). General Methods for Monitoring Convergence of Iterative Simulations. *American Statistical Association, Volume 7, Number 4, 434-455*.
- Clark, T., Hall, G. and Griffiths, R. (2002). *Bayesian Logistic Regression Using a Perfect Phylogeny*. Department of Epidemiology and Puplic Health Imperial College. London, UK.
- Congdon P. (2005). *Bayesian models for categorical data*/Peter Congdon. Wiley series in probability and statistics: ISBN 0-470-09237-8 (cloth: alk. paper).
- Consortium on Reproductive Health Association (CORHA), (2005) Assessment of the reproductive health situations/ problems of students in the Addis Ababa, Bahirdar, Jimma, and Mekele Universities.
- Dugassa and Alemfrie, (2015). Entrepreneurship Education in Ethiopian universities: Institutional assessment Synthesis Report <https://www.msm.nl/.../Entrepreneurship-Synthesis-Report-FINAL-10July2015.pdf>

- Efron, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis. *Journal of American Statistical Association*, 70, 892-898
- FHI, (2006) Meeting the health needs of young clients: A guide to providing Reproductive Health Services to adolescents
- FMOH, 2006 – 2015 national reproductive health strategy; Federal democratic republic Ethiopia ministry of health March 2006
- Gelman, A. (2005). Alternative Methods for Monitoring Convergence of Markov Chain Monte Carlo Iterative Simulations. *Journal of Computational and Graphical Statistics*. 7, 434-455.
- Geman, S. and Geman, D. (2009). Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721-741.
- Gilks, W., Richardson, S. and Spiegelhalter, J. (2011). *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London, UK.
- Goldstein, H. (1991). Nonlinear multilevel models with an application to discrete response data. *Biometrika*: 78, 45-51.
- Goldstein H. (2003). *Multilevel Statistical Models*. 3rd Edition London: Arnold; New York: Oxford University Press Inc.
- Goldstein H. and Rasbash J. (1996). Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society*: 159: 505-513.
- Hosmer, D. and Lemeshow, S. (2011). *Applied Logistic Regression* (5th Edition). New York: John Wiley and Sons.
- International Planned Parenthood Federation (IPPF) (2007). A guide for developing policies on the sexual and reproductive health and rights of young people in Europe. IPPF European Network; Safe Project. UK.
- International Planned Parenthood Federation (IPPF) (2008). Provide: Strengthening youth friendly services. *Inspire* pack. UK.
- Ioannis, N. (2009). *Bayesian Modeling Using Win BUGS*. Department of Statistics Athens University of Economics and Business Athens, Greece.
- Laird, N. M. (1978). Empirical Bayes methods for two-way contingency tables. *Biometrika*, 65, 581–590.
- Lee, P. (2010). *Bayesian Statistics: An Introduction*, 2nd edition, Arnold, London.

- Malarcher S. (2010). A review of sexual and reproductive health through the equity lens: In Malarcher, S. (ed), *Social Determinants of Sexual and Reproductive Health*, WHO Press, Switzerland.
- Menard, S. (2002). *Applied Logistic Regression*, 2nd Edition: Quantitative Applications in the Social Sciences, Sage publications.
- Merkle, E., Sheu, C. and Trisha, G. (2011). Simulation-Based Bayesian Inference using WinBUGS. WinBUGS Tutorial Outline: <http://www.bu.cam.ac.uk/winbugs/cont.shtml>.
- Meskerem A. and Worku A. (2014). Utilization of youth reproductive health services and associated factors among high school students in Bahir Dar, Amhara Regional State, Ethiopia. *Open Journal of Epidemiology*, 4, 69-75 Published Online May 2014 in *SciRes*.
- Mila, A.L., Yang, X.B. and Carriquiry, A.L. (2003). Bayesian Logistic Regression of Clinical Epidemiology for Uncertainty in Parameter Estimation. *Basic Science for Clinical Medicine*: Little, Brown and Company, Boston.
- Ministry of Health launched Ethiopia, the National Reproductive Health Strategy in 2006 (2006a.).
- Motuma et al., (2016) Utilization of youth friendly services and associated factors among youth in Harar town, east Ethiopia: a mixed method study. *BMC Health Services Research*. 16(272):1-10.
- Obonyo Perez Akinyi, (2009). Determinants of Utilization of Youth Friendly Reproductive Health Services Among School and College Youth in Thika West District, Kiambu County, KENYA; REG. NO: P57/PT/12366 / 2009
- Senderowitz, J., Hainsworth, G. and Solter, C (2003). A Rapid Assessment of Youth Friendly Reproductive Health Service, Pathfinder International Technical Guidance Series Number 4, page 2.
- Snijders, T. and Bosker, R. (1999). *Multilevel Analysis: an Introduction to Basic and Advanced Multilevel Modeling*. London/Thousand Oaks/ New Delhi: Sage Publications *Statistical Association*, B 50 (2): 225-265.
- Tanner, M.A. (2011). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, New York: Springer-Verlag.
- The World's Youth 2013 Data Sheet—Population Reference Bureau: www.prb.org/pdf13/youth-data-sheet-2013.pdf
- Tolessa kebede (2014). Assessment of utilization of youth friendly reproductive health services among collage youth in Asela town, Oromia regional state, Ethiopia. Addis Ababa. Addis Ababa University thesis.
- United Nations (1994a) Report of the International Conference on Population and Development Cairo, 5 – 13 September 1994 New York United Nations
- WHO, (2004). Adolescent friendly health services in South East Asia region: Report of 9 Regional consultation, 9-14 February 2004, Bali, Indonesia.
- WHO, (2010). *Quality assessment guidebook a guide to assessing health services for adolescent clients*, Geneva

Appendices

Appendix A: SAS codes for binary logistic regression analysis

Sas codes:

* SAS outputs for binary logistic regression model;

ods graphics on;

title 'Occurrence of constriction';

proc logistic data=final DESCENDING;

class age aware Ctime Famincome Fee Gender income poccupation Residence /
param=ref;

model RHS= age aware Ctime Famincome Fee Gender income poccupation Residence
/selection=forward outroc= roc expB LACKFIT influence iplots CTABLE;

run;

ods graphics off;

Appendix B: SAS outputs for binary logistic regression model

A) Model Fit Statistics

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	3304.870	2995.517
SC	3310.647	3128.377
-2 Log L	3302.870	2949.517

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	353.3533	19	<.0001
Score	329.9459	19	<.0001
Wald	289.1958	19	<.0001

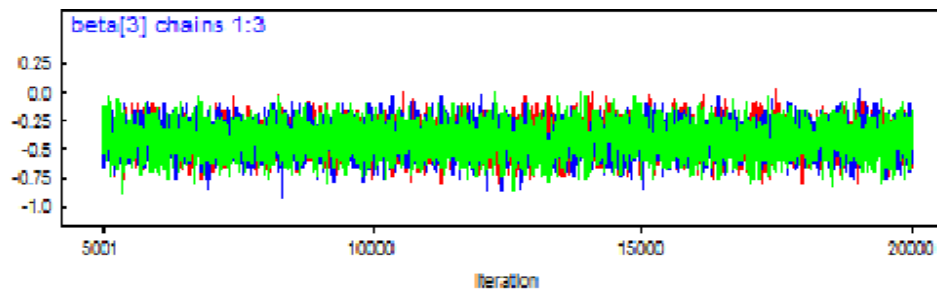
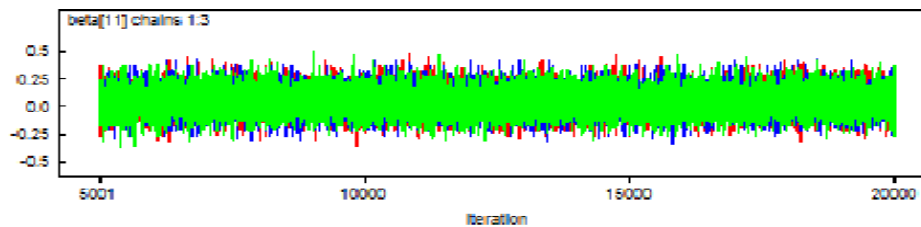
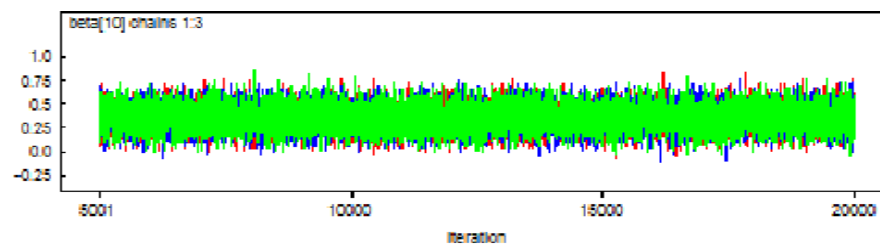
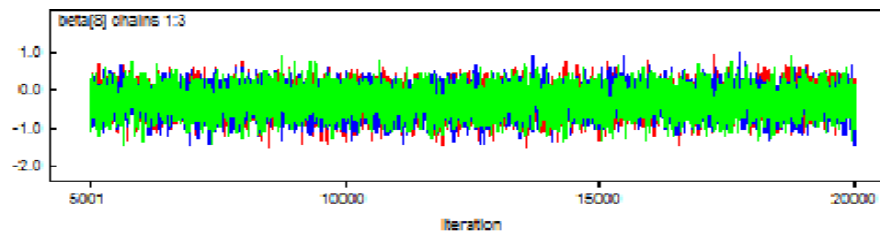
B) Type 3 analysis of effects

Type 3 Analysis of Effects

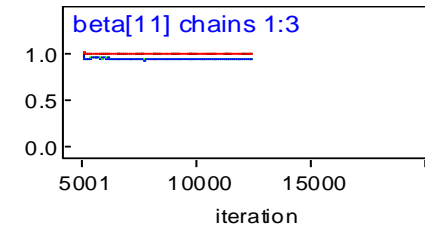
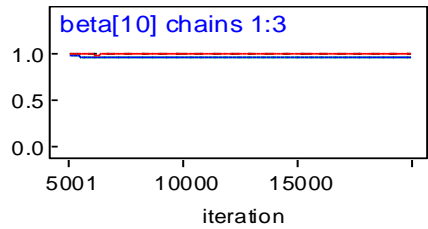
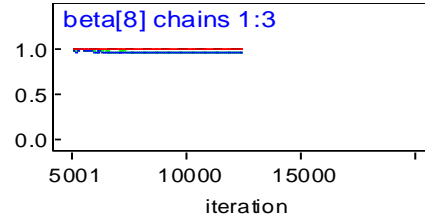
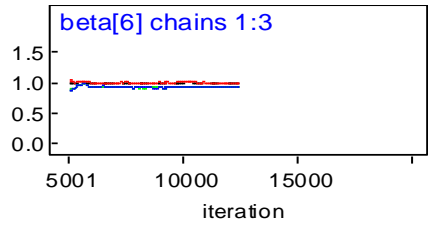
Effect	DF	Wald Chi-Square	Pr > ChiSq
age	2	13.8290	0.0010
aware	1	59.2453	<.0001
Ctime	1	6.9920	0.0082
Famincome	4	18.1773	0.0011
Fee	2	9.2361	0.0099
Gender	1	40.7842	<.0001
income	4	45.2924	<.0001
poccupation	3	35.8424	<.0001
Residence	1	4.8548	0.0276

Bayesian Diagnostics of Convergence for significant parameters

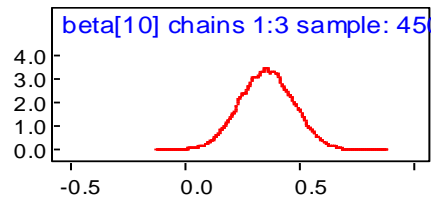
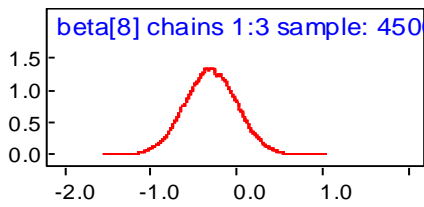
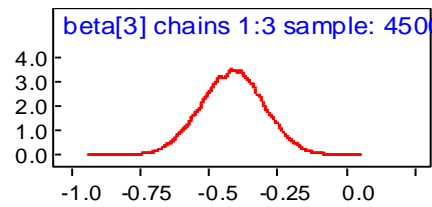
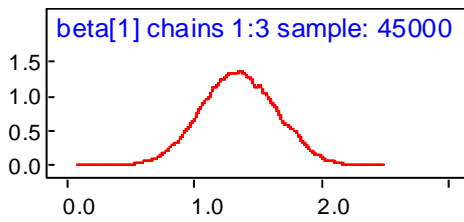
I. Time series



II. Gelman –Rubin Statistics



III. Kernel density



IV. Auto correlation plot

